Econometric Theory, **40**, 2024, 1120–1158.
doi:10.1017/S0266466623000026

CONSISTENT NON-GAUSSIAN PSEUDO MAXIMUM LIKELIHOOD ESTIMATORS OF SPATIAL AUTOREGRESSIVE MODELS

Fei Jin
Fudan University and Shanghai Institute of International Finance and Economics

Yuqin Wang
Shanghai University of Finance and Economics and Key Laboratory of Mathematical Economics (SUFE)

This paper studies the non-Gaussian pseudo maximum likelihood (PML) estimation of a spatial autoregressive (SAR) model with SAR disturbances. If the spatial weights matrix M_n for the SAR disturbances is normalized to have row sums equal to 1 or the model reduces to a SAR model with no SAR process of disturbances, the non-Gaussian PML estimator (NGPMLE) for model parameters except the intercept term and the variance σ_0^2 of independent and identically distributed (i.i.d.) innovations in the model is consistent. Without row normalization of M_n, the symmetry of i.i.d. innovations leads to consistent NGPMLE for model parameters except σ_0^2. With neither row normalization of M_n nor the symmetry of innovations, a location parameter can be added to the non-Gaussian pseudo likelihood function to achieve consistent estimation of model parameters except σ_0^2. The NGPMLE with no added parameter can have a significant efficiency improvement upon the Gaussian PML estimator and the generalized method of moments estimator based on linear and quadratic moments. We also propose a non-Gaussian score test for spatial dependence, which can be locally more powerful than the Gaussian score test. Monte Carlo results show that our NGPMLE with no added parameter and the score test based on it perform well in finite samples.

1. INTRODUCTION

The spatial autoregressive (SAR) model, originated in Cliff and Ord (1973, 1981), is a popular spatial econometric model. It has been applied in a range of fields

We are very grateful to the Editor (Peter Phillips), the Co-Editor (Dennis Kristensen), and two anonymous referees for helpful comments that led to substantial improvements of this paper. We also thank Zaichao Du, Zhonghao Fu, Gaosheng Ju, Lung-Fei Lee, Hongjun Li, Xi Qu, Ke Miao, Qiao Wang, Xiaohu Wang, Xingbai Xu, and seminar participants at Fudan University and the 2021 Symposium on Econometrics and Big Data at Xiamen University for helpful comments. Fei Jin gratefully acknowledges the financial support from the National Natural Science Foundation of China (NNSFC) (Nos. 71973030 and 71833004). Yuqin Wang thanks the NNSFC financial support under No. 72103122. Address correspondence to Yuqin Wang, Institute for Advanced Research, Shanghai University of Finance and Economics, Shanghai 200433, China; e-mail: wang.yuqin@sufe.edu.cn

in economics to capture spatial dependence.[1] In this paper, we consider the non-Gaussian pseudo maximum likelihood (PML) estimation of the SAR model with SAR disturbances (SARAR model), with no need to correctly specify the distribution of independent and identically distributed (i.i.d.) innovations in the model. We provide conditions for the consistency of the non-Gaussian PML estimator (NGPMLE) and prove its asymptotic distribution. Our applications to several popular datasets in the spatial econometric literature show some evidence of nonnormal and leptokurtic innovations for these datasets.[2] In such situations, our NGPMLE on the basis of leptokurtic distributions can have significant efficiency improvements over existing estimators including the Gaussian PML estimator (GPMLE) (Lee, 2004), and lead to different but more reliable empirical results.

We consider the following SARAR model:

$$Y_n = \lambda_0 W_n Y_n + X_n \beta_0 + U_n, \;\; U_n = \rho_0 M_n U_n + \sigma_0 V_n, \tag{1}$$

where n is the sample size, $Y_n = [y_{n1}, \ldots, y_{nn}]'$ is an $n \times 1$ vector of observations on the dependent variable, X_n is an $n \times k_x$ matrix of exogenous variables, $W_n = [w_{n,ij}]$ and $M_n = [m_{n,ij}]$ are spatial weights matrices with zero diagonals,[3] the innovations v_i's in $V_n = [v_1, \ldots, v_n]'$ are i.i.d. with mean zero and unit variance, λ_0 and ρ_0 are scalar spatial dependence parameters, β_0 is a $k_x \times 1$ parameter vector, and σ_0 is a standard deviation (SD) parameter. We formulate an NGPMLE using a chosen density function for v_i that can differ from its true density function. Our results on the consistency of the NGPMLE for the SARAR model extend those in Newey and Steigerwald (1997) for conditional heteroskedasticity models, by properly taking into account spatial dependence.[4] We show that, when the spatial weights matrix M_n in the SAR process of disturbances is normalized to have row sums equal to 1,[5] the NGPMLE for model parameters except the intercept term and the variance σ_0^2 of i.i.d. innovations is consistent under regularity conditions; without row normalization of M_n, if the innovations are symmetric, the NGPMLE for model parameters except σ_0^2 is consistent; and with neither row normalization of M_n nor the symmetry of innovations, a location parameter can be added to the pseudo

[1] Reviews on studies about the class of SAR models can be found in, e.g., Anselin and Bera (1998), Anselin (2010), and Arbia (2016).

[2] See Section 5 and the Supplementary Material.

[3] The zero diagonals of the spatial weights matrices exclude self-influence. It is a normalization condition usually maintained in the literature (see, e.g., Kelejian and Prucha, 1998; Lee, 2004). Indeed, it is not used in our theoretical analysis.

[4] Other studies on the NGPMLE include, among others, Gouriéroux, Monfort, and Trognon (1984), Francq, Lepage, and Zakoïan (2011), Fan, Qi, and Xiu (2014), and Fiorentini and Sentana (2019). The results in Gouriéroux et al. (1984) are on the basis of a density function $f(x,m)$ or $f(x,m,\Sigma)$, where m is the mean and Σ is the variance of the distribution. They focus on the exponential family, for which all moments exist. Our analysis does not restrict the density function to be of the form $f(x,m)$ or $f(x,m,\Sigma)$, and we can use a density function which does not have a finite moment with an order higher than 3. Francq et al. (2011) and Fan et al. (2014) propose modifications of NGPMLEs for GARCH models with zero conditional mean. Fiorentini and Sentana (2019) propose consistent NGPMLEs for GARCH models with nonzero conditional mean and for some other location-scale models such as multivariate regressions.

[5] We refer to a matrix with all row sums equal to 1 as a row-normalized matrix hereafter.

likelihood function to obtain consistent estimators of model parameters except σ_0^2. An important special case of the SARAR model is the SAR model with exogenous variables but with no SAR process of disturbances. Consistent non-Gaussian PML estimation of model parameters except σ_0^2 only requires an intercept term in the model. Furthermore, although we only consider SAR models in this paper, consistent NGPMLEs can also be extended to other spatial econometric models.[6] We expect that NGPMLEs for those models can be more efficient than existing estimation methods.

We prove the $\sqrt{n}$-consistency and asymptotic normality of our NGPMLE under the condition that the innovations have a finite third moment, which can allow for innovations with relatively heavy tails. By contrast, the $\sqrt{n}$-consistency of the GPMLE is established under the existence of a moment of innovations with an order higher than 4 (Lee, 2004). Furthermore, using numerical integration and Student's t distribution to formulate a likelihood function, we show that the NGPMLE with no added parameter can have a uniform efficiency improvement upon the GPMLE, and can also have a significantly larger efficiency improvement than the best generalized method of moments (GMM) estimator on the basis of linear-quadratic moments (Liu, Lee, and Bollinger, 2010), but the NGPMLE with an added parameter can be less efficient than the GPMLE. An intuitive explanation from the non-Gaussian score is that, unlike GPMLE and the best GMM estimator (BGMME), the NGPMLE with no added parameter does not restrict the moments to be linear and quadratic in innovations. The NGPMLE with an added parameter loses some efficiency since one more parameter has to be estimated. Our Monte Carlo experiments further corroborate the efficiency improvement of the NGPMLE with no added parameter upon the GPMLE and BGMME.

We also propose a non-Gaussian score test for spatial dependence in SAR models, which only requires the restricted NGPMLE. The test statistic generalizes the Moran I test statistic that is quadratic in estimated innovations (Moran, 1950). If the NGPMLE is asymptotically more efficient than the GPMLE, then the non-Gaussian score test is locally more powerful than the Gaussian score test.

Estimation methods for SAR models include maximum likelihood (ML) (Ord, 1975), generalized spatial two-stage least squares (GS2SLS) (Kelejian and Prucha, 1998), Gaussian PML,[7] GMM (Lee, 2007),[8] best GMM, and adaptive estimation (Robinson, 2010; Lee and Robinson, 2020), among others. GS2SLS is computationally simpler than ML, Gaussian PML, GMM, and best GMM, but is less efficient. Like our NGPMLE, the GPMLE does not need the distribution of

[6]For example, the matrix exponential spatial specification (LeSage and Pace, 2007), spatial moving average models (e.g., Haining, 1978; Cliff and Ord, 1981; Fingleton, 2008; Doğan and Taşpınar, 2013), and high-order versions of those models (e.g., Blommestein, 1983, 1985). See the Supplementary Material for some consistency analysis.

[7]Exact and high-order properties of the GPMLE are studied in Bao (2013) and Hillier and Martellosio (2018). Gupta and Robinson (2018) study the GPMLE of SAR models with increasingly many parameters.

[8]A related estimation method is the generalized empirical likelihood (Jin and Lee, 2019), which is asymptotically as efficient as the GMM with the same moments, but can have smaller higher-order bias.

innovations to be correctly specified, and it is relatively efficient.[9] In addition, whether the NGPMLE or the ML estimator based on nonnormal distributions is consistent or not is not clear according to the existing literature. Thus, the GPMLE is popular in practice (see, e.g., Robinson, 2010). However, it can have a significant efficiency loss compared with the ML estimator, when the innovations are far from normally distributed (Fan et al., 2014). The adaptive estimation in Robinson (2010) requires that each unit is influenced aggregately by a significant portion of units in the population, which is a very stringent condition that may not be reasonable in some practical circumstances.[10]

This paper is organized as follows: In Section 2, we prove the convergence and asymptotic distribution of the NGPMLE for the SARAR model, and compare its efficiency with those of the GPMLE and BGMME. In Section 3, the non-Gaussian score test is investigated. Monte Carlo and application results are reported in Sections 4 and 5, respectively. Section 6 concludes. Proofs and other materials are collected in the Appendix and in the Supplementary Material.

2. NGPMLE

Let $\theta_0 = [\lambda_0, \rho_0, \beta_0', \sigma_0^2]'$ be the true parameter vector in model (1), and let $\theta = [\lambda, \rho, \beta', \sigma^2]'$ be a general parameter vector. We consider a density function $f(x, \eta)$ of a random variable with mean zero and unit variance, where η is a $k_\eta \times 1$ parameter vector. For example, $f(x, \eta)$ can be the density function of a standardized Student's t distribution with η degrees of freedom. The pseudo log-likelihood function of the SARAR model (1), as if v_i had the density function $f(v_i, \eta)$, is

$$\ln L_n(\gamma) = \sum_{i=1}^{n} \ln f\big(v_i(\theta), \eta\big) - \frac{n}{2}\ln(\sigma^2) + \ln|S_n(\lambda)| + \ln|R_n(\rho)|, \tag{2}$$

where $\gamma = [\theta', \eta']'$, $S_n(\lambda) = I_n - \lambda W_n$ with I_n being the n-dimensional identity matrix, $R_n(\rho) = I_n - \rho M_n$, and $v_i(\theta) = \frac{1}{\sigma} e_{ni}' R_n(\rho)[S_n(\lambda)Y_n - X_n\beta]$, with e_{ni} being the ith column of I_n. We may fix η at some particular value or estimate it jointly with θ. We focus on the case where η is estimated jointly with θ, as in Fiorentini and Sentana (2019). An NGPMLE of γ is derived by maximizing $\ln L_n(\gamma)$ in (2).

We first introduce some regularity conditions for later analysis on model (1).

Assumption 1 (Topological space). Let $\mathbb{D} \subset \mathbb{R}^{c_d}$, $c_d \geq 1$, be a lattice of (possibly) unevenly placed locations in $\mathbb{R}^{c_d}$. $\mathbb{D}$ is infinitely countable and the distance $d(i,j)$ between any two elements i and j in $\mathbb{D}$ is larger than or equal to a specific positive constant, say 1 without loss of generality. n individual units in an economy for model (1) are located or living in a region $\mathbb{D}_n \subset \mathbb{D}$, where the cardinality of $\mathbb{D}_n$ is n.

[9] It is asymptotically equivalent to a GMM estimator with linear and quadratic moments, where the linear moments correspond to the instrumental variables estimation of the parameters in the equation on the dependent variable in a GS2SLS approach.

[10] This condition is the same as that for the consistency of the ordinary least-squares estimator (Lee, 2002).

Since the general density function $f(x,\eta)$ can introduce nonlinearity into the pseudo log-likelihood function, we require a proper law of large numbers (LLN) for analysis. We use the LLN for near-epoch-dependent (NED) spatial processes, developed in Jenish and Prucha (2012). Assumption 1 maintains some conditions required for such an LLN. The assumption provides basic settings on individual units. The minimum distance assumption on individual units corresponds to increasing domain asymptotics in the spatial literature.[11]

Let $\|\cdot\|_\infty$ and $\|\cdot\|_1$ be, respectively, the row sum and column sum matrix norms.

Assumption 2 (Basic conditions on model elements). (i) v_i's are i.i.d. with mean zero and unit variance. (ii) W_n and M_n are nonstochastic matrices such that $\sup_n \|W_n\|_\infty < \infty$ and $\sup_n \|M_n\|_\infty < \infty$. (iii) $c_0 \equiv \max\{|\lambda_0| \sup_n \|W_n\|_\infty, |\rho_0| \sup_n \|M_n\|_\infty\} < 1$. (iv) The elements of X_n are uniformly bounded constants.

We consider i.i.d. innovations as in many papers on spatial econometric models. The uniform boundedness condition on the spatial weights matrices in Assumption 2(ii), originated in Kelejian and Prucha (1998, 1999, 2001), limits the degree of spatial dependence to be manageable.[12] The elements of spatial weights matrices are often nonnegative in practice, but our theoretical analysis does not require such an assumption. Assumption 2(iii) implies the nonsingularity of $R_n \equiv R_n(\rho_0)$ and $S_n \equiv S_n(\lambda_0)$ for any n. In Assumption 2(iv), the elements of X_n are assumed to be constants for simplicity, as in Lee (2004).[13]

2.1. Consistency

Model (1) can be written as

$$R_n S_n Y_n = R_n X_n \beta_0 + \sigma_0 V_n. \tag{3}$$

Thus, for given λ_0 and ρ_0, (3) is a linear regression model with $R_n S_n Y_n$ being a vector of observations on the dependent variable and $R_n X_n$ being the explanatory variable matrix. Newey and Steigerwald (1997) establish a set of results on the consistency of the NGPMLE for coefficients in a conditional heteroskedasticity model, which nests the linear regression model as a special case. These results depend on whether the model has an intercept term or whether model innovations are symmetric.[14] The regression (3) may not have an intercept term, but if M_n is

[11] Another commonly used asymptotic method is called infill asymptotics, for which the sample region is fixed and the growth of the sample size is achieved by sampling points arbitrarily dense in the given region. See Cressie (1993) and Conley (1999) for more explanations and examples. If $f(x,\eta)$ is the density function of normal distributions, then $\ln f(x,\eta)$ is a quadratic function of x. In this special case, asymptotic analysis can be based on the LLN for linear-quadratic forms (Kelejian and Prucha, 2001); therefore, Assumption 1 is not needed.

[12] In the spatial econometric literature, a spatial weights matrix is often assumed to be bounded in both the row- and column-sum norms. Later we introduce conditions that imply $\sup_n \|W_n\|_1 < \infty$ and $\sup_n \|M_n\|_1 < \infty$; therefore, Assumption 2(ii) only involves the row-sum norms of W_n and M_n.

[13] Alternatively, X_n can be allowed to be stochastic with the existence of certain moments.

[14] To gain some intuition on the results, consider the case that the assumed density f is symmetric and non-Gaussian. As f is not a Gaussian density, the mean of the dependent variable in a linear regression model is generally not a

row-normalized and X_n contains an intercept term such that $X_n = [1_n, X_{2n}]$, where 1_n is an $n \times 1$ vector of ones, then $R_n X_n = [(1-\rho_0)1_n, R_n X_{2n}]$ contains an intercept term. Hence, for given λ_0 and ρ_0, we expect the consistency of the NGPMLE of some parameters in (3) under some regularity conditions. However, we have to properly take into account that the spatial dependence parameters λ and ρ are also estimated. In the following, we provide sufficient conditions for the consistent NGPMLE of some parameters in (3).

Under regularity conditions, $\frac{1}{n}\ln L_n(\gamma) - \frac{1}{n}\mathrm{E}[\ln L_n(\gamma)]$ converges to zero uniformly on a compact parameter space of γ. Suppose that $\lim_{n\to\infty}\frac{1}{n}\mathrm{E}[\ln L_n(\gamma)]$ is uniquely maximized at some pseudo-true value of γ, then the NGPMLE of γ converges to the pseudo-true value in probability under regularity conditions. The following Assumptions 3 and 4 guarantee that $\mathrm{E}[\ln L_n(\gamma)]$ is uniquely maximized at the pseudo-true value, where some components of the pseudo-true value will be equal to their true values. Denote $\beta = [\beta_1, \beta_2']'$ in the case that X_n contains an intercept term, where β_1 is the parameter for 1_n. Accordingly, let $\beta_0 = [\beta_{10}, \beta_{20}']'$. For a square matrix A, let $\mathrm{vec}_{\mathrm{D}}(A)$ be a column vector formed by the diagonal elements of A. Denote $A_{1n} = M_n R_n^{-1}$, $A_{2n} = R_n W_n S_n^{-1} R_n^{-1}$, $A_{3n} = M_n W_n S_n^{-1} R_n^{-1}$, and $T_n(\tau) = R_n(\rho) S_n(\lambda) S_n^{-1} R_n^{-1} = [t_{n,ij}(\tau)]$ with $\tau = [\lambda, \rho]'$.

Assumption 3 (Identification A). (*i*) $f(x,\eta) > 0$, for any x and η, and $\mathrm{E}[\ln f(v_i(\theta),\eta)] < \infty$ for all γ in its parameter space. (*ii*) $X_n' R_n' R_n X_n$ is nonsingular. (*iii*) For any (α_1, α_2), every element of $1_n + \alpha_1 \mathrm{vec}_{\mathrm{D}}(A_{1n}) + \alpha_2 \mathrm{vec}_{\mathrm{D}}(A_{2n}) + \alpha_1\alpha_2 \mathrm{vec}_{\mathrm{D}}(A_{3n})$ is nonzero. (*iv*) $g_n(\tau) > 0$, for $\tau \neq \tau_0$, where $g_n(\tau) = \sum_{i=1}^n \ln|t_{n,ii}(\tau)| - \ln|T_n(\tau)|$.

Assumption 3(i) is a usual regularity condition. The nonsingularity of $X_n' R_n' R_n X_n$ in Assumption 3(ii) is for the identification of β_0. Assumption 3(iii) implies that $t_{n,ii}(\tau) \neq 0$ for any i and any τ. Note that $T_n(\tau_0) = I_n$, whose diagonal elements are all equal to 1. Then the assumption is satisfied at least for τ close to τ_0.

Assumption 3(iv) is for the identification of τ_0. It is a generalized version of Hadamard's inequality for positive semidefinite matrices. Lin and Sinnamon (2020) provide sufficient conditions for Assumption 3(iv), which require all principal minors of $T_n(\tau)$ to be nonnegative and to satisfy a Fischer-type inequality. Alternatively, we could investigate conditions for Assumption 3(iv) in a neighborhood of τ_0. Since $g_n(\tau_0) = 0$ and $\frac{\partial g_n(\tau_0)}{\partial \tau} = 0$, we have $g_n(\tau) > 0$ for $\tau \neq \tau_0$ in a neighborhood of τ_0 if $\frac{\partial^2 g_n(\tau_0)}{\partial\tau\partial\tau'}$ is positive-definite. Let $T_{1n} = A_{1n} - \mathrm{diag}(A_{1n})$ and $T_{2n} = A_{2n} - \mathrm{diag}(A_{2n})$, where $\mathrm{diag}(A)$ for a square matrix A denotes a diagonal

natural location parameter of the assumed density. Thus, if f differs from the true density, the consistency of the NGPMLE of the parameters for the mean is not guaranteed. When the true density is symmetric, the mean, median, and mode of the dependent variable are equal; thus, the mean and the natural location parameter are the same for f. It follows that the parameters for the mean can be consistently estimated by the non-Gaussian PML under regularity conditions. In the case that the true density is asymmetric, if there is no intercept term, the difference between the mean and the natural location parameter for f leads to the inconsistency of the NGPMLE of the parameters for the mean. The existence of an intercept in a linear regression model accounts for the difference, so other parameters for the mean can still be consistently estimated by the non-Gaussian PML.

matrix formed by the diagonal elements of A. Then $\frac{\partial^2 g_n(\tau_0)}{\partial\tau\partial\tau'}$ is positive-definite when W_n and M_n are equal, T_{1n} and T_{2n} are linearly independent, and either W_n is symmetric or it is row-normalized from a symmetric matrix (see Lemma B.1 in Appendix B).

Assumption 4 (Identification B). Either the following (*i*) or (*ii*) holds:

(i) (a) M_n is row-normalized. (b) X_n contains an intercept term. (c) $\mathrm{E}[\ln f(\frac{\sigma_0 v_i-\alpha}{\sigma},\eta)]-\ln(\sigma)$ has a unique maximum at $[\sigma_\infty,\alpha_\infty,\eta'_\infty]'$.

(ii) (*a*) v_i is symmetrically distributed around zero with unimodal density $k(v)$, which satisfies that $k(v_1)\le k(v_2)$ for $|v_1|\ge|v_2|$. (b) For each η, $f(v,\eta)=f(-v,\eta)$ and $f(v_1,\eta)<f(v_2,\eta)$ for $|v_1|>|v_2|$. (c) $\mathrm{E}[\ln f(\frac{\sigma_0 v_i}{\sigma},\eta)]-\ln(\sigma)$ has a unique maximum at $[\sigma_\infty,\eta'_\infty]'$.

The spatial weights matrix M_n can be either row-normalized or not row-normalized, but a row-normalized M_n facilitates the interpretation of the spatial dependence parameter ρ, since it indicates that each element of M_nU_n is a weighted average of U_n for a nonnegative M_n. Thus, spatial weights matrices are often row-normalized in practice.[15] An intercept term is usually included in the SARAR model in empirical research.[16] Assumption 4(i)(c) and (ii)(c) is the same as Assumptions 2.4 and 2.6 in Newey and Steigerwald (1997), respectively. Assumption 4(i)(c) strengthens Assumption 4(ii)(c). With a row-normalized M_n and an intercept term in X_n, the term $\frac{1}{\sigma}(\sigma_0 v_i-\alpha)$ in Assumption 4(i)(c) is equal to $v_i(\theta)$ evaluated at $\theta=[\lambda_0,\rho_0,\frac{\alpha}{1-\rho_0}+\beta_{10},\beta'_{20},\sigma^2]'$. Newey and Steigerwald (1997) provide some insights on Assumption 4(ii)(c). A necessary condition for it is that $\mathrm{E}[\ln f(\frac{\sigma_0 v_i}{\sigma},\eta_\infty)]-\ln\sigma$ is uniquely maximized at $\sigma=\sigma_\infty$. Therefore, $f(x,\eta)$ should be chosen such that σ_∞ minimizes the Kullback–Leibler distance between the true innovation density and the pseudo density $\frac{\sigma_0}{\sigma}f(\frac{\sigma_0 x}{\sigma},\eta_\infty)$. Such an assumption holds for the Gaussian likelihood, the likelihood for a standardized Student's t distribution with more than two degrees of freedom, and a generalized Gaussian likelihood with $\ln f(x,\eta)=-|x|^\eta[\Gamma(3/\eta)/\Gamma(1/\eta)]^{\eta/2}+c$, where c is a constant and $\Gamma(\cdot)$ denotes the gamma function (Fan et al., 2014). The assumption also implies that σ_∞ is generally different from σ_0, although it is straightforward to show that $\sigma_\infty=\sigma_0$ if $f(\cdot)$ is a Gaussian density.[17] For the case with symmetric innovations, Assumption 4(ii)(a) and (b) is the same as Assumption 2.3 in Newey

[15] Another reason is that it implies a simple interval of ρ for the nonsingularity of $I_n-\rho M_n$. See the discussions in, e.g., Kelejian and Prucha (2010). Some authors prefer not to row-normalize a spatial weights matrix (e.g., Baltagi, Egger, and Pfaffermayr, 2008).

[16] In some rare cases, an intercept term is not included, e.g., when Y_n and X_n are normalized to have mean zero. An example can be found in LeSage (1999, p. 72).

[17] Furthermore, σ_∞/σ_0 and η_∞ only depend on the true disturbance distribution and the chosen density function $f(v,\eta)$, but do not depend on model characteristics such as spatial weights matrices, exogenous variables, and parameter values. The σ_∞/σ_0 differs from 1 even when the true innovation distribution and the chosen density function $f(v,\eta)$ are spherically symmetric. We report the values of σ_∞/σ_0 for some chosen disturbance distributions and a density function $f(v,\eta)$ in the Supplementary Material.

and Steigerwald (1997). Both the true density function of v_i and the assumed density function $f(v,\eta)$ are required to be unimodal.

PROPOSITION 1. *(i) If Assumptions 1–3 and 4(i) are satisfied, then* $\mathrm{E}[\ln L_n(\gamma)]$ *is uniquely maximized at* $\gamma_* = [\lambda_0, \rho_0, \beta_{1\infty}, \beta_{20}', \sigma_\infty^2, \eta_\infty']'$*, where* $\beta_{1\infty} = \beta_{10} + \frac{\alpha_\infty}{1-\rho_0}$*. (ii) If Assumptions 1–3 and 4(ii) are satisfied, then* $\mathrm{E}[\ln L_n(\gamma)]$ *is uniquely maximized at* $\gamma_\# = [\lambda_0, \rho_0, \beta_0', \sigma_\infty^2, \eta_\infty']'$.

In the case with a row-normalized M_n, the intercept term and the variance parameter are generally not consistently estimated, whereas other model parameters can be consistently estimated; in the case with symmetric innovations, only the variance parameter is inconsistently estimated.

Remark 1. For a SAR model with no SAR process of disturbances, i.e., $Y_n = \lambda_0 W_n Y_n + X_n\beta_0 + \sigma_0 V_n$, a result similar to Proposition 1(i) holds, where Assumption 4(i) reduces to that X_n contains an intercept term and $\mathrm{E}[\ln f(\frac{\sigma_0 v_i - \alpha}{\sigma}, \eta)] - \ln(\sigma)$ has a unique maximum at $[\sigma_\infty, \alpha_\infty, \eta_\infty']'$. As M_n does not appear in the model, the condition of a row-normalized M_n is irrelevant. For a given λ_0, the SAR model is a linear regression model with the dependent variable S_nY_n and the exogenous variable matrix X_n. It can also be seen as a special case of the SARAR model with a row-normalized M_n and $\rho_0 = 0$; therefore, it is not considered separately.[18]

In the case with neither row normalization of M_n nor the symmetry of innovations, we could add a location parameter α to the non-Gaussian pseudo log-likelihood function to obtain the modified function[19]

$$\ln L_n(\delta) = \sum_{i=1}^{n} \ln f\Big(v_i(\theta) - \frac{1}{\sigma}\alpha, \eta\Big) - \frac{n}{2}\ln(\sigma^2) + \ln|S_n(\lambda)| + \ln|R_n(\rho)|, \qquad \textbf{(4)}$$

where $\delta = [\lambda, \rho, \beta', \sigma^2, \alpha, \eta']'$. This function is formed as if we had the model $Y_n = \lambda_0 W_n Y_n + X_n\beta_0 + U_n$, where $U_n = \alpha_0 1_n + \rho_0 M_n U_n + \sigma_0 V_n$. This model can be rewritten as $R_nS_nY_n = R_nX_n\beta_0 + \alpha_0 1_n + \sigma_0 V_n$, which has an intercept term. Thus, as the above analysis under Assumption 4(i), we could show that $\mathrm{E}[\ln L_n(\delta)]$ is uniquely maximized at $\delta_\# = [\lambda_0, \rho_0, \beta_0', \sigma_\infty^2, \alpha_\infty, \eta_\infty']'$ under regularity conditions.

PROPOSITION 2. *If Assumptions 1–3 and* $4(i)(c)$ *are satisfied and* R_nX_n *does not contain an intercept term, then* $\mathrm{E}[\ln L_n(\delta)]$ *is uniquely maximized at* $\delta = \delta_\#$.

[18] See the Supplementary Material for formal analysis.

[19] When M_n is row-normalized and X_n contains an intercept term, since $v_i(\theta) - \frac{\alpha}{\sigma} = \frac{1}{\sigma} e_{ni}' R_n(\rho)[S_n(\lambda)Y_n - X_{2n}\beta_2] - \frac{(1-\rho)\beta_1 + \alpha}{\sigma}$, $\ln L_n(\delta)$ is not uniquely maximized and thus should not be used. When v_i is symmetric, $\ln L_n(\delta)$ can still be used to derive an NGPMLE, but there might be efficiency loss. Newey and Steigerwald (1997) study such efficiency loss for conditional heteroskedasticity models. We do not examine the issue theoretically for SAR models in this study, but we investigate it by Monte Carlo experiments.

The identification results in Propositions 1 and 2 are for a finite n. To prove the convergence of the NGPMLE, we need to strengthen the identification inequalities to the limit.[20]

Assumption 5 (Identification for large samples). For the log-likelihood function $\ln L_n(\gamma)$ in (2), assume that $\limsup_{n\to\infty} \frac{1}{n}\{\mathrm{E}[\ln L_n(\gamma)] - \mathrm{E}[\ln L_n(\gamma_*)]\} < 0$, for any $\gamma \neq \gamma_*$, if Assumption 4(i) holds, and assume that $\limsup_{n\to\infty} \frac{1}{n}\{\mathrm{E}[\ln L_n(\gamma)] - \mathrm{E}[\ln L_n(\gamma_\#)]\} < 0$, for any $\gamma \neq \gamma_\#$, if Assumption 4(ii) holds. For $\ln L_n(\delta)$ in (4), assume that $\limsup_{n\to\infty} \frac{1}{n}\{\mathrm{E}[\ln L_n(\delta)] - \mathrm{E}[\ln L_n(\delta_\#)]\} < 0$, for any $\delta \neq \delta_\#$.

We introduce more regularity conditions for the analysis on the consistency of NGPMLEs.

Assumption 6 (Consistency A). (i) $S_n(\lambda)$ is invertible for any λ in its parameter space Λ and $\{S_n^{-1}(\lambda)\}$ is bounded in either the row sum or column sum matrix norm uniformly on Λ. Similar conditions hold for $R_n(\rho)$. (ii) The parameter space Γ of γ is a compact subset of $\mathbb{R}^{k_\gamma}$, where k_γ is the length of γ. Similar conditions hold for δ and κ.

Assumption 6(i) is required due to the nonlinearity involved in the log Jacobians $\ln|S_n(\lambda)|$ and $\ln|R_n(\rho)|$ in the pseudo log-likelihood functions. The compactness of parameter spaces in Assumption 6(ii) is a familiar assumption on extremum estimators.

Assumption 7 (Consistency B). At least one of the following two conditions (i) and (ii) is satisfied:

(i) Only individuals whose distances are less than or equal to some specific constant $\bar{d}_0$ may affect each other directly, i.e., $w_{n,jk}$ and $m_{n,jk}$ can be nonzero only if $d(j,k) \leq \bar{d}_0$ for any j, k, and n.

(ii) (a) For every n, the number of columns $w_{n,\cdot j}$ of W_n with $|\lambda_0|\sum_{i=1}^n |w_{n,ij}| > c_0$ is less than or equal to some fixed nonnegative integer that does not depend on n, denoted as N.[21] A similar condition holds for M_n. (b) There are constants π_1 and π_2 with $\pi_2 > c_d$ such that $|w_{n,jk}| \leq \pi_1 d(j,k)^{-\pi_2}$ and $|m_{n,jk}| \leq \pi_1 d(j,k)^{-\pi_2}$, where c_d is in Assumption 1.

Assumption 8 (Consistency C). (i) $f(x,\eta)$ is differentiable with respect to x and η such that $|\frac{\partial \ln f(x,\eta)}{\partial x}| \leq c_f(|x|^{c_t}+1)$ and $\|\frac{\partial \ln f(x,\eta)}{\partial \eta}\| \leq c_f(|x|^{1+c_t}+1)$ for some constant c_f and $c_t = 0$ or 1. (ii) For the c_t in (i), $\mathrm{E}(|v_i|^{2+2c_t+\iota}) < \infty$, for some $\iota > 0$.

Assumptions 7 and 8 are maintained to show the NED properties of some relevant terms. Assumption 7 on the spatial weights matrices is the same as Assumption 3 in Xu and Lee (2015) for a SAR Tobit model. Assumption 7(i) does not allow direct interactions between individuals far from each other. While

[20] It is common to assume separate identification conditions for a finite n and for large samples in the spatial econometric literature. See, e.g., Assumption 8 in Xu and Lee (2015).

[21] The c_0 here is some positive number smaller than 1, which can be different from that in Assumption 2(iii). We use c_0 for simplicity as in Xu and Lee (2015).

Assumption 7(ii)(b) allows any off-diagonal element of spatial weights matrices to be nonzero, the interaction needs to decay fast enough. Assumption 7(ii)(a) corresponds to the existence of a limited number of spatial units that can have large aggregated effects on other spatial units.

Assumption 8(i) covers the case with a bounded $\frac{\partial \ln f(x,\eta)}{\partial x}$ and the case where $|\frac{\partial \ln f(x,\eta)}{\partial x}| \leq c_f(|x|+1)$ for some constant c_f. The derivative $\frac{\partial \ln f(x,\eta)}{\partial x}$ is bounded for a smooth enough $f(x,\eta)$ whose tail behavior is proportional to $|x|^{-a}$, for $a \geq 1$, or $e^{-b|x|^a}$, for $0 < a \leq 1$ and $b > 0$. Examples include Student's t and the logistic distributions. On the other hand, $|\frac{\partial \ln f(x,\eta)}{\partial x}| \leq c_f(|x|+1)$ for some constant c_f for a smooth enough $f(x,\eta)$ whose tail behavior is proportional to $e^{-b|x|^a}$, for $0 < a \leq 2$ and $b > 0$. An example is the normal distribution. The condition on $\frac{\partial \ln f(x,\eta)}{\partial \eta}$ is also satisfied for Student's t, logistic, and normal distributions. Depending on whether $\frac{\partial \ln f(x,\eta)}{\partial x}$ is bounded or $|\frac{\partial \ln f(x,\eta)}{\partial x}| \leq c_f(|x|+1)$, Assumption 8(ii) requires different moment conditions on v_i. With a bounded $\frac{\partial \ln f(x,\eta)}{\partial x}$, we only need v_i to have a finite moment with the order $2+\iota$ for some $\iota > 0$.

Denote the NGPMLEs that maximize $\ln L_n(\gamma)$ and $\ln L_n(\delta)$ by, respectively, $\hat{\gamma}$ and $\hat{\delta}$. The convergence of the NGPMLEs is summarized in the following theorem.

THEOREM 1. *Suppose that Assumptions 1–3 and 5–8 are satisfied.*

(i) *For the case with a row-normalized M_n, if Assumption 4(i) is also satisfied, then $\hat{\gamma} = \gamma_* + o_p(1)$.*

(ii) *For the case with symmetric v_i, if Assumption 4(ii) is also satisfied, then $\hat{\gamma} = \gamma_\# + o_p(1)$.*

(iii) *For the case with neither row-normalization of M_n nor the symmetry of v_i, if Assumption 4(i)(c) is also satisfied and R_nX_n does not contain an intercept term, then $\hat{\delta} = \delta_\# + o_p(1)$.*

2.2. Asymptotic Distributions

The asymptotic distributions of the NGPMLEs can be derived by mean value theorem expansions of their first-order conditions at the pseudo-true values, and applying a proper central limit theorem (CLT).

As an example, consider the case with symmetric v_i. With the reduced form $Y_n = S_n^{-1}(X_n\beta_0 + \sigma_0 R_n^{-1} V_n)$, each element of $\frac{\partial \ln L_n(\gamma_\#)}{\partial \gamma}$ is a special case of the general form

$$\omega_n = \varepsilon_n' A_n V_n + b_n' \varepsilon_n + 1_n' \Psi_n - \mathrm{E}(\varepsilon_n' A_n V_n), \tag{5}$$

where $\varepsilon_n = \left[\frac{\partial f(\frac{\sigma_0}{\sigma_\infty} v_1, \eta_\infty)}{\partial v}, \ldots, \frac{\partial f(\frac{\sigma_0}{\sigma_\infty} v_n, \eta_\infty)}{\partial v}\right]' \equiv [\epsilon_i]$, $\Psi_n = \left[\frac{\partial f(\frac{\sigma_0}{\sigma_\infty} v_1, \eta_\infty)}{\partial \eta}, \ldots, \frac{\partial f(\frac{\sigma_0}{\sigma_\infty} v_n, \eta_\infty)}{\partial \eta}\right]' c_\eta \equiv [\psi_i]$ with c_η being a $k_\eta \times 1$ vector of constants, $A_n = [a_{n,ij}]$ is an $n \times n$ nonstochastic matrix, $b_n = [b_{ni}]$ is an $n \times 1$ vector of constants, and ε_n, V_n, and Ψ_n have zero means (see the proof of Theorem 2). The ω_n can be shown to be asymptotically normal by a CLT for martingale difference arrays, as the proof for the asymptotic normality of linear-quadratic forms of innovations in Kelejian and Prucha (2001). Such a result is provided in Lemma 6 of Yang and Lee (2017).

We maintain the following assumption for the analysis on the asymptotic distributions.

Assumption 9 (Asymptotic distributions). (i) γ_*, $\gamma_\#$, and $\delta_\#$ are in the interior of their respective parameter spaces. (ii) $f(x,\eta)$ is thrice differentiable with respect to $z = [x,\eta']'$, such that $\|\frac{\partial^2 \ln f(x,\eta)}{\partial z \partial z'}\| \le c_f(|x|^{2c_t}+1)$ and $\|\frac{\partial^3 \ln f(x,\eta)}{\partial z \partial z' \partial z_i}\| \le c_f(|x|^{3c_t}+1)$ for each element z_i of z, where $c_t = 0$ for the case with bounded $\frac{\partial \ln f(x,\eta)}{\partial x}$, and $c_t = 1$ for the case with $|\frac{\partial \ln f(x,\eta)}{\partial x}| \le c_f(|x|+1)$, as stated in Assumption 8(i). (*iii*) $\mathrm{E}(|v_i|^{3c_t+3}) < \infty$. (*iv*) If Assumption 7(i) holds, assume that $\sup_n \|S_n^{-1}\|_1 < \infty$ and $\sup_n \|R_n^{-1}\|_1 < \infty$.

Assumption 9(i) is a familiar condition required for the $\sqrt{n}$-convergence of extremum estimators. Assumption 9(ii) contains further smoothness conditions on $f(x,\eta)$. It is similar to Assumption 10 in Xu and Lee (2018), and it is satisfied with $c_t = 0$ for Student's t, logistic, and normal distributions. With Assumption 9(ii), only a finite third moment of innovations is needed in Assumption 9(iii) for the case with bounded $\frac{\partial \ln f(x,\eta)}{\partial x}$.[22] As the GPMLE is shown to be $\sqrt{n}$-consistent only under the existence of moments of innovations with an order higher than 4, it is possible that it has a rate of convergence slower than $\sqrt{n}$ when innovations only have a finite third moment. In such a situation, the NGPMLE is certainly more efficient than the GPMLE by Theorem 2. Assumption 9(ii) and (iii) is maintained to show the convergence of the Hessian matrices $\frac{1}{n}\frac{\partial^2 \ln L_n(\hat{\gamma})}{\partial \gamma \partial \gamma'}$ and $\frac{1}{n}\frac{\partial^2 \ln L_n(\hat{\delta})}{\partial \delta \partial \delta'}$. Assumption 9(iv) of boundedness in the column-sum norm of S_n^{-1} and R_n^{-1} is required for asymptotic distributions as in Kelejian and Prucha (1998) and Lee (2004). It is not required in the situation of Assumption 7(ii) since it can be directly proved (see Lemma B.6).

THEOREM 2. *Suppose that Assumptions 1–3 and 5–9 are satisfied.*

(i) *For the case with a row-normalized M_n, if Assumption 4(i) is also satisfied, then $\sqrt{n}(\hat{\gamma} - \gamma_*) \xrightarrow{d} N(0, \lim_{n\to\infty} \mathcal{A}^{-1}\mathcal{B}\mathcal{A}^{-1})$, where $\mathcal{A} = -\frac{1}{n}\mathrm{E}(\frac{\partial^2 \ln L_n(\gamma_*)}{\partial \gamma \partial \gamma'})$ and $\mathcal{B} = \frac{1}{n}\mathrm{E}(\frac{\partial \ln L_n(\gamma_*)}{\partial \gamma}\frac{\partial \ln L_n(\gamma_*)}{\partial \gamma'})$.*

(ii) *For the case with symmetric v_i, if Assumption 4(ii) is also satisfied, then $\sqrt{n}(\hat{\gamma} - \gamma_\#) \xrightarrow{d} N(0, \lim_{n\to\infty} \mathcal{A}^{-1}\mathcal{B}\mathcal{A}^{-1})$, where $\mathcal{A} = -\frac{1}{n}\mathrm{E}(\frac{\partial^2 \ln L_n(\gamma_\#)}{\partial \gamma \partial \gamma'})$ and $\mathcal{B} = \frac{1}{n}\mathrm{E}(\frac{\partial \ln L_n(\gamma_\#)}{\partial \gamma}\frac{\partial \ln L_n(\gamma_\#)}{\partial \gamma'})$.*

(iii) *For the case with neither row-normalization of M_n nor the symmetry of v_i, if Assumption 4(i)(c) is also satisfied and R_nX_n does not contain an intercept term, then $\sqrt{n}(\hat{\delta} - \delta_\#) \xrightarrow{d} N(0, \lim_{n\to\infty} \mathcal{A}^{-1}\mathcal{B}\mathcal{A}^{-1})$, where $\mathcal{A} = -\frac{1}{n}\mathrm{E}(\frac{\partial^2 \ln L_n(\delta_\#)}{\partial \delta \partial \delta'})$ and $\mathcal{B} = \frac{1}{n}\mathrm{E}(\frac{\partial \ln L_n(\delta_\#)}{\partial \delta}\frac{\partial \ln L_n(\delta_\#)}{\partial \delta'})$.*

[22] It is possible to develop formal tests for finiteness of moments of innovations in the SARAR model, which is beyond the scope of this paper.

The specific expressions of $\mathcal{A}$ and $\mathcal{B}$ are in Appendix A.[23] For easy reference, denote the NGPMLE without added parameter by NGPMLE$_\text{o}$, and that with an added parameter by NGPMLE$_\text{a}$. For the special case of a spatial error model with symmetric innovations, i.e., model (1) with $\lambda_0 W_n Y_n$ omitted and symmetric v_i, we could show that $\mathcal{A}$ and $\mathcal{B}$ for NGPMLE$_\text{o}$ are block diagonal and the NGPMLE$_\text{o}$ of β has a more explicit expression, as presented in the following corollary.

COROLLARY 1. *For the spatial error model with symmetric v_i, the* NGPMLE$_\text{o}$ *of β has the asymptotic variance* $\lim_{n\to\infty} \frac{\sigma_0^2 \mathrm{E}(\xi_{1i}^2)}{[\mathrm{E}(\xi_{2i})]^2}(\frac{1}{n}X_n'R_n'R_nX_n)^{-1}$, *where* $\xi_{1i} = \frac{\sigma_0}{\sigma_\infty}\frac{\partial \ln f(\frac{\sigma_0}{\sigma_\infty}v_i,\eta_\infty)}{\partial v}$ *and* $\xi_{2i} = -\frac{\sigma_0^2}{\sigma_\infty^2}\frac{\partial^2 \ln f(\frac{\sigma_0}{\sigma_\infty}v_i,\eta_\infty)}{\partial v^2}$, *and the GPMLE and BGMME of β have the asymptotic variance* $\lim_{n\to\infty}\sigma_0^2(\frac{1}{n}X_n'R_n'R_nX_n)^{-1}$.

By the above corollary, for the spatial error model with symmetric v_i, the BGMME of β has no efficiency improvement over the GPMLE, and the efficiency of NGPMLE$_\text{o}$ relative to the GPMLE is determined by the scalar $\frac{\mathrm{E}(\xi_{1i}^2)}{[\mathrm{E}(\xi_{2i})]^2}$. For the general SARAR model, $\mathcal{A}$ and $\mathcal{B}$ are not block diagonal and the estimation of η may affect the asymptotic efficiency of the NGPMLE of model parameters. Thus, it is not easy to compare analytically the efficiencies of the NGPMLE and other estimators.

2.3. Efficiency Comparisons

In this subsection, we compare the estimation efficiency of our NGPMLE with those of the GPMLE and BGMME.[24] For the asymptotic variance of the NGPMLE, as the closed form is not available, we compute the asymptotic variance in Theorem 2 for a given sample size with numerical integration. Student's t distribution with unknown degrees of freedom is used in deriving the NGPMLE.[25]

[23] One may estimate $\mathcal{A}$ and $\mathcal{B}$ using the expressions in Appendix A for inference purposes. Alternatively, $\mathcal{A}$ can be estimated using $-\frac{1}{n}\frac{\partial^2 \ln L_n(\hat{\gamma})}{\partial\gamma\partial\gamma'}$ or $-\frac{1}{n}\frac{\partial^2 \ln L_n(\hat{\delta})}{\partial\delta\partial\delta'}$, and $\mathcal{B}$ can be estimated according to the martingale structure of the non-Gaussian score.

[24] Various impacts arising from a change in an exogenous explanatory variable, as defined in, e.g., LeSage and Pace (2009), are functions of the spatial lag parameter λ_0 and the coefficient on the variable. Then by the delta method, if the NGPMLE is asymptotically more efficient than other estimators, so are the impact estimators computed with the NGPMLE than those computed with other estimators. Some efficiency comparisons for impact estimators based on numerical integration and Monte Carlo experiments are provided in the Supplementary Material. The patterns are the same as those for estimators. We thank an anonymous referee for the suggestion of considering impact estimators.

[25] In this study, we have not theoretically considered the choice of distributions used to derive the NGPMLEs. As suggested in Fan et al. (2014), the distributions can be chosen to minimize the asymptotic variance of the NGPMLE in Theorem 2. In addition, the NGPMLE and the GPMLE can be aggregated to derive an estimator that is more efficient than both. A more practical method can be based on diagnostic tests. In the Supplementary Material, we derive some diagnostic tests such as the normality and excess kurtosis tests of innovations in the SARAR model. Nonnormal innovations imply that a proper NGPMLE can be more efficient than the GPMLE. If the excess kurtosis test suggests a positive excess kurtosis, then we can use a leptokurtic distribution such as Student's t distribution; otherwise, a platykurtic distribution such as the raised cosine distribution can be used. Our applications imply leptokurtic distributions of innovations; therefore, we use Student's t distribution with one parameter, which is relatively simple

TABLE 1. Models considered for efficiency comparisons.

	Row-normalized M_n	Non-row-normalized M_n
Spatial error model:	Symmetric and asymmetric v_i	—
SARAR model:	Asymmetric v_i	Symmetric and asymmetric v_i

The considered models are listed in Table 1. For the SARAR model, the spatial weights matrix M_n is block-diagonal and each diagonal block is based on the matrix for the study of crimes across 49 districts in Columbus, Ohio, in Anselin (1988); M_n is either row-normalized or normalized by its spectral radius; W_n is set to be equal to M_n; the exogenous variable matrix X_n contains an intercept term and a standard normal random variable; the spatial dependence parameters λ_0 and ρ_0 are equal to 0.4 and 0.2, respectively; the coefficients for X_n are set to 1; the true variance parameter σ_0^2 is 0.25; and the sample size is 147. For the case with symmetric innovations, v_i is set to be a mixture of two normal distributions with mean zero, and for the case with asymmetric innovations, v_i is an admissible fourth-order Gram–Charlier expansion of the standard normal distribution as a function of the skewness and kurtosis coefficients.[26] The settings for the spatial error model are the same as for the SARAR model, except for the omission of $\lambda_0 W_n Y_n$.

2.3.1. *Spatial Error Model with a Row-Normalized M_n.* We first consider the spatial error model with a row-normalized M_n. Figure 1 reports the results for both symmetric and asymmetric innovations. We observe that NGPMLE$_o$ improves upon GPMLE in all cases with a nonnormal true disturbance distribution, and the efficiency improvement can be up to about 50%. In the case with symmetric innovations, BGMME shows almost no efficiency improvement over GPMLE; in the case with asymmetric innovations, BGMME shows some efficiency improvement over GPMLE but usually much less than NGPMLE$_o$. Only in the case with asymmetric innovations and for the parameter β_2, BGMME can be slightly more efficient than NGPMLE$_o$, which occurs when the skewness coefficient is relatively large and the kurtosis coefficient is small. For the case with asymmetric innovations, the efficiency of NGPMLE$_o$ relative to GPMLE increases with kurtosis, whereas it is almost not affected by skewness.

2.3.2. *SARAR Model with a Row-Normalized M_n and Asymmetric v_i.* Figure 2 reports the efficiency comparison results for the SARAR model with a row-normalized M_n and asymmetric innovations. Similar to the results for the spatial

and can have various degrees of excess kurtosis. As pointed out by the Co-Editor and an anonymous referee, using a sufficiently general family of distributions can lead to efficiency loss because many more parameters are estimated alongside other model parameters, whereas using diagnostic tests to choose distributions can suffer from the pretesting issue (e.g., Giles and Giles, 1993). We leave those issues to future study.

[26]The admissible combinations of the skewness and kurtosis coefficients can be seen from, e.g., Spiring (2011).

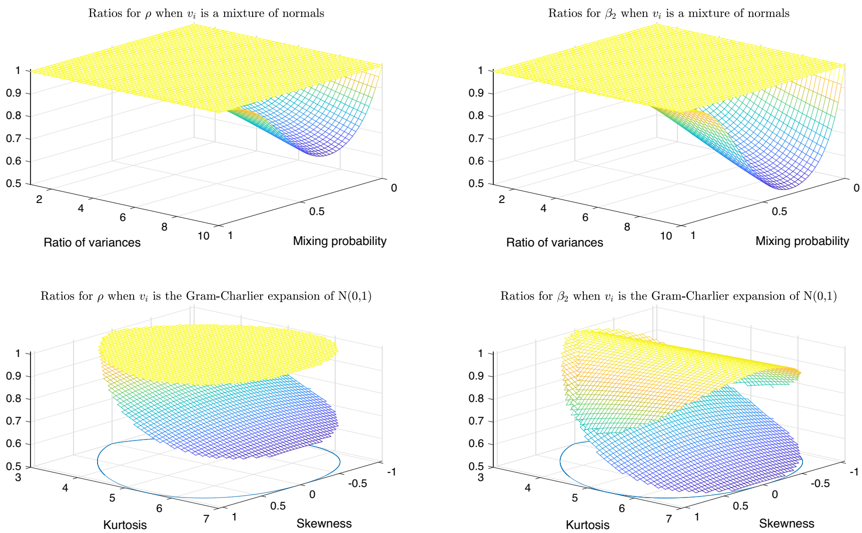

FIGURE 1. Efficiency comparisons of different estimators for the spatial error model with a row-normalized M_n. The lower mesh in each subfigure shows the ratios of the asymptotic variance of NGPMLE$_o$ to that of GPMLE, whereas the upper mesh shows the ratios of the asymptotic variance of BGMME to that of GPMLE.

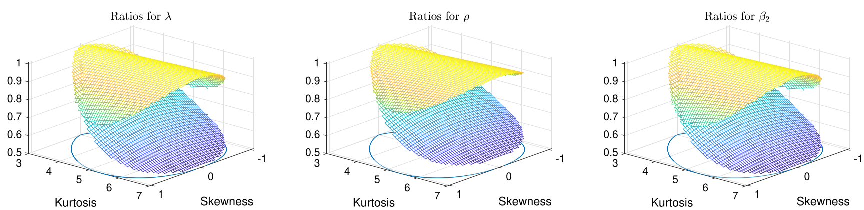

FIGURE 2. Efficiency comparisons of different estimators for the SARAR model with a row-normalized M_n and asymmetric innovations. The v_i is an admissible fourth-order Gram–Charlier expansion of the standard normal distribution as a function of the skewness and kurtosis coefficients. The lower mesh in each subfigure shows the ratios of the asymptotic variance of NGPMLE$_o$ to that of GPMLE, whereas the upper mesh shows the ratios of the asymptotic variance of BGMME to that of GPMLE.

error model, NGPMLE$_o$ shows a significant efficiency improvement over GPMLE, and the improvement is much larger than that of BGMME in most cases.

2.3.3. *SARAR Model with a Non-Row-Normalized M_n.* We next consider the SARAR model with a non-row-normalized M_n. When the innovations are symmetric, we consider NGPMLE$_o$ as well as NGPMLE$_a$ since both estimators of λ, ρ, and β_2 are consistent. Figure 3 shows the results. NGPMLE$_o$ is still observed

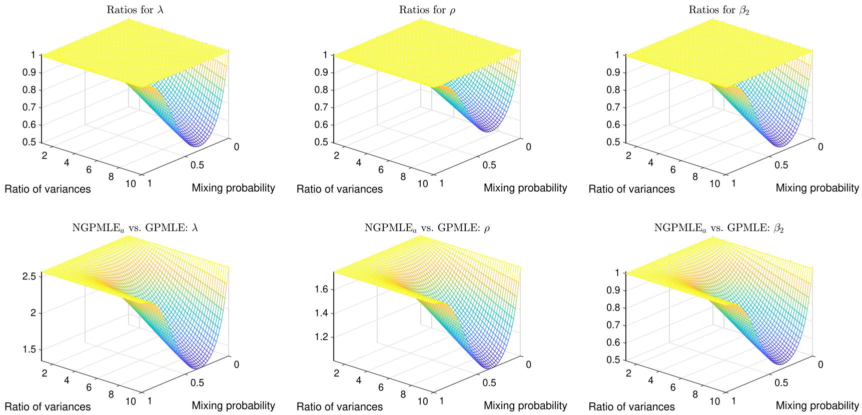

FIGURE 3. Efficiency comparisons of different estimators for the SARAR model with a non-row-normalized M_n and symmetric innovations. The v_i is a mixture of two normal distributions with mean zero. For the first three subfigures, the lower mesh in each subfigure shows the ratios of the asymptotic variance of NGPMLE$_\mathrm{o}$ to that of GPMLE, whereas the upper mesh shows the ratios of the asymptotic variance of BGMME to that of GPMLE. For the fourth to sixth subfigures, the mesh in each subfigure shows the ratios of the asymptotic variance of NGPMLE$_\mathrm{a}$ to that of GPMLE.

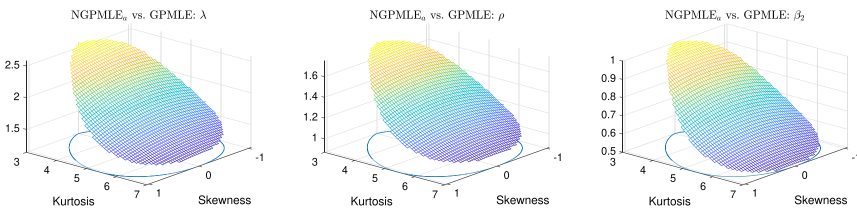

FIGURE 4. Efficiency comparisons of different estimators for the SARAR model with a non-row-normalized M_n and asymmetric innovations. The v_i is an admissible fourth-order Gram–Charlier expansion of the standard normal distribution as a function of the skewness and kurtosis coefficients. The mesh in each subfigure shows the ratios of the asymptotic variance of NGPMLE$_\mathrm{a}$ to that of GPMLE.

to have a significant efficiency improvement over GPMLE, but NGPMLE$_\mathrm{a}$ only has smaller variance than that of GPMLE for β_2, and its variances for the spatial dependence parameters λ and ρ are typically much larger than those of GPMLE. Figure 4 further demonstrates the efficiency loss of NGPMLE$_\mathrm{a}$ due to an added parameter, for the case with asymmetric innovations.

To summarize, our experiments based on Student's t distribution in Sections 2.3.1–2.3.3 show that NGPMLE$_\mathrm{o}$ has a uniform efficiency improvement upon GPMLE, which is usually much larger than the efficiency improvement of BGMME, but NGPMLE$_\mathrm{a}$, the NGPMLE with an added parameter, can be less efficient than GPMLE.

3. NON-GAUSSIAN SCORE TEST FOR SPATIAL DEPENDENCE

In this section, we propose a score test for spatial dependence based on the non-Gaussian pseudo log-likelihood function $\ln L_n(\gamma)$ in (2).[27]

Consider a test of the null hypothesis that $\tau_0 = 0$. Let $\check{\gamma} = [0,0,\check{\beta}',\check{\sigma}^2,\check{\eta}']'$ be the restricted NGPMLE of γ, which is derived by maximizing $\ln L_n(\gamma)$ in (2) with the restriction $\tau = 0$ imposed. The non-Gaussian score test is based on the asymptotic distribution of $\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\check{\gamma})}{\partial \tau}$. Note that

$$\frac{\partial \ln L_n(\check{\gamma})}{\partial \tau} = \left[-\frac{1}{\check{\sigma}}\sum_{i=1}^{n}\frac{\partial \ln f(v_i(\check{\theta}),\check{\eta})}{\partial v}e_{ni}'W_nY_n,\ -\sum_{i=1}^{n}\frac{\partial \ln f(v_i(\check{\theta}),\check{\eta})}{\partial v}e_{ni}'M_nV_n(\check{\theta})\right]',$$

where $\check{\theta} = [0,0,\check{\beta}',\check{\sigma}^2]'$, $v_i(\check{\theta}) = \frac{1}{\check{\sigma}}e_{ni}'(Y_n - X_n\check{\beta})$, and $V_n(\check{\theta}) = [v_1(\check{\theta}),\ldots,v_n(\check{\theta})]'$. A special case of interest is the test for spatial dependence in the spatial error model. In this case, the test is based on the asymptotic distribution of $-\sum_{i=1}^{n}\frac{\partial \ln L_n(v_i(\check{\theta}),\check{\eta})}{\partial v}e_{ni}'M_nV_n(\check{\theta})$ for a spatial weights matrix M_n. The statistic $-\sum_{i=1}^{n}\frac{\partial \ln L_n(v_i(\check{\theta}),\check{\eta})}{\partial v}e_{ni}'M_nV_n(\check{\theta})$ generalizes the quadratic form $V_n'(\check{\theta})M_nV_n(\check{\theta})$ for Moran's I test for spatial dependence, where $V_n'(\check{\theta})M_nV_n(\check{\theta})$ can also be derived from the Gaussian score (Burridge, 1980).

We may apply the mean value theorem to derive the asymptotic distribution of $\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\check{\gamma})}{\partial \tau}$ under the null hypothesis. Let $\mathcal{A} = -\frac{1}{n}\mathrm{E}(\frac{\partial^2 \ln L_n(\gamma_\infty)}{\partial\gamma\partial\gamma'})$ and $\mathcal{B} = \frac{1}{n}\mathrm{E}(\frac{\partial \ln L_n(\gamma_\infty)}{\partial\gamma}\frac{\partial \ln L_n(\gamma_\infty)}{\partial\gamma'})$.[28] For any two subvectors γ_1 and γ_2 of γ, denote $\mathcal{A}_{\gamma_1\gamma_2} = -\frac{1}{n}\mathrm{E}(\frac{\partial^2 \ln L_n(\gamma_\infty)}{\partial\gamma_1\partial\gamma_2'})$ and $\mathcal{B}_{\gamma_1\gamma_2} = \frac{1}{n}\mathrm{E}(\frac{\partial \ln L_n(\gamma_\infty)}{\partial\gamma_1}\frac{\partial \ln L_n(\gamma_\infty)}{\partial\gamma_2'})$. Under the null hypothesis and regularity conditions,

$$\begin{aligned}\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\check{\gamma})}{\partial \tau} &= \frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\gamma_\infty)}{\partial \tau} + \frac{1}{n}\mathrm{E}\Big(\frac{\partial^2 \ln L_n(\gamma_\infty)}{\partial\tau\partial\gamma_u'}\Big)\sqrt{n}(\check{\gamma}_u - \gamma_{u\infty}) + o_p(1)\\ &= \Delta\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\gamma_\infty)}{\partial\gamma} + o_p(1) \xrightarrow{d} N\Big(0, \lim_{n\to\infty}\Delta\mathcal{B}\Delta'\Big),\end{aligned}$$

where $\gamma_u = [\beta',\sigma^2,\eta']'$, $\gamma_{u\infty}$ is the pseudo-true value of γ_u, and $\Delta = [I_2, -\mathcal{A}_{\tau\gamma_u}\mathcal{A}_{\gamma_u\gamma_u}^{-1}]$. Let $\hat{\Delta}$ and $\hat{\mathcal{B}}$ be estimators of, respectively, Δ and $\mathcal{B}$, such that $\hat{\Delta} = \Delta + o_p(1)$ and $\hat{\mathcal{B}} = \mathcal{B} + o_p(1)$ under the null hypothesis. The test statistic has the form

$$\boldsymbol{t}_n = \frac{1}{n}\frac{\partial \ln L_n(\check{\gamma})}{\partial \tau'}(\hat{\Delta}\hat{\mathcal{B}}\hat{\Delta}')^{-1}\frac{\partial \ln L_n(\check{\gamma})}{\partial \tau}, \tag{6}$$

which is asymptotically chi-square distributed with two degrees of freedom under the null hypothesis.

---

[27] A test based on $\ln L_n(\delta)$ in (4) is omitted since the last section shows that the resulting NGPMLE can be less efficient than the GPMLE, and the efficiency of an estimator relates to the power of related tests, as shown in Theorem 3.

[28] We use γ_∞ here for simplicity. By $\tau_0 = 0$ and Theorem 1, $\gamma_\infty = [0,0,\beta_{1\infty},\beta_{20}',\sigma_\infty^2,\eta_\infty']'$ in the case with a row-normalized M_n, and $\gamma_\infty = [0,0,\beta_0',\sigma_\infty^2,\eta_\infty']'$ in the case with symmetric v_i.

For the asymptotic analysis on $\boldsymbol{t}_n$, we assume that the true τ in the data generating process follows the Pitman drift in the following assumption.

Assumption 10 (Pitman drift). $\tau_n = \frac{1}{\sqrt{n}} c_\tau$, where c_τ is a 2×1 vector of constants.

THEOREM 3. *If Assumptions 1–4 and 5–10 are satisfied, then* $\boldsymbol{t}_n \xrightarrow{d} \chi_2^2(\lim_{n\to\infty} c_\tau' \Lambda(\Delta\mathcal{B}\Delta')^{-1}\Lambda c_\tau)$, *where* $\Lambda = \mathcal{A}_{\tau\tau} - \mathcal{A}_{\tau\gamma_u}\mathcal{A}_{\gamma_u\gamma_u}^{-1}\mathcal{A}_{\gamma_u\tau}$ *and* $\chi_{a_1}^2(a_2)$ *denotes a noncentral chi-square distribution with* a_1 *degrees of freedom and noncentrality parameter* a_2.

By Theorem 2 and the partitioned matrix inverse formula, the asymptotic variance of the NGPMLE $\hat{\tau}$ has the form $\Upsilon = \lim_{n\to\infty} \Lambda^{-1}\Delta\mathcal{B}\Delta'\Lambda^{-1}$. The noncentrality parameter for the asymptotic noncentral chi-square distribution of $\boldsymbol{t}_n$ is equal to $c_\tau' \Upsilon^{-1} c_\tau$. Thus, if the NGPMLE of τ is asymptotically more efficient than the GPMLE, then the non-Gaussian score test is locally more powerful than the Gaussian score test.

4. MONTE CARLO

In this section, we implement some Monte Carlo experiments to investigate the finite-sample performance of the NGPMLE and non-Gaussian score test. As in Section 2.3, the NGPMLE is derived by assuming Student's t distribution with unknown degrees of freedom.

4.1. Estimators

We consider three cases: the SARAR model with a row-normalized M_n and asymmetric v_i, the SARAR model with a non-row-normalized M_n and symmetric v_i, and the SAR model with symmetric v_i.[29] For the SAR model, we also consider the adaptive estimators proposed in Robinson (2010).[30] Parameters for the innovations correspond to cases where the NGPMLE and BGMME show different levels of efficiency improvements in Section 2.3. The number of Monte Carlo repetitions is 5,000. Other settings are the same as those in Section 2.3.

Table 2 reports the biases, SDs, and root-mean-squared errors (RMSE) of various estimators for the SARAR model with a row-normalized M_n and asymmetric innovations. The biases of GPMLE, BGMME, and NGPMLE$_o$ are similar in magnitude. Since the biases are small compared with the SDs, the RMSEs

[29]For the three cases considered, the identification conditions in Assumption 3 are satisfied. In the Supplementary Material, we report some Monte Carlo results for the case when Assumption 3 fails. We observe that the NGPMLE for the spatial dependence parameters and the coefficients on nonintercept exogenous variables still has similar bias as the GPMLE. Thus, it is possible that the NGPMLE for some model parameters is consistent even when Assumption 3 fails. We leave this question to future research.

[30]The adaptive estimators do not apply to the SARAR model (Remark 3 on page 9 of Robinson, 2010).

Table 2. Performance of various estimators for the SARAR model with a row-normalized M_n and asymmetric v_i.

			λ			ρ			β_2		
Kurtosis	Skewness		Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
		Panel A: $n = 147$									
6	0.8	GPMLE	−0.008	0.076	0.077	−0.025	0.142	0.144	−0.001	0.042	0.042
		BGMME	−0.006	0.073	0.073	−0.013	0.143	0.144	0.000	0.040	0.040
		NGPMLE$_o$	−0.006	0.059	0.060	−0.018	0.119	0.120	−0.001	0.032	0.032
6	0.05	GPMLE	−0.006	0.076	0.076	−0.032	0.141	0.145	−0.002	0.041	0.041
		BGMME	−0.005	0.078	0.079	−0.017	0.146	0.147	−0.002	0.042	0.042
		NGPMLE$_o$	−0.004	0.060	0.060	−0.026	0.120	0.123	−0.001	0.033	0.033
4	0.4	GPMLE	−0.006	0.076	0.076	−0.029	0.143	0.146	−0.002	0.042	0.042
		BGMME	−0.005	0.077	0.077	−0.016	0.148	0.149	−0.002	0.042	0.042
		NGPMLE$_o$	−0.005	0.073	0.074	−0.028	0.140	0.143	−0.002	0.040	0.041
4	0.05	GPMLE	−0.008	0.077	0.078	−0.027	0.143	0.145	−0.001	0.042	0.042
		BGMME	−0.007	0.080	0.081	−0.012	0.148	0.149	−0.001	0.043	0.043
		NGPMLE$_o$	−0.008	0.075	0.075	−0.026	0.141	0.143	−0.001	0.041	0.041
3.05	0.05	GPMLE	−0.008	0.076	0.077	−0.027	0.142	0.144	−0.002	0.042	0.042
		BGMME	−0.007	0.080	0.080	−0.011	0.147	0.147	−0.002	0.043	0.043
		NGPMLE$_o$	−0.009	0.079	0.080	−0.026	0.146	0.148	−0.002	0.043	0.043

(continued)

TABLE 2. (continued)

Kurtosis	Skewness		λ Bias	λ SD	λ RMSE	ρ Bias	ρ SD	ρ RMSE	β_2 Bias	β_2 SD	β_2 RMSE
		Panel B: $n = 294$									
6	0.8	GPMLE	−0.003	0.052	0.052	−0.012	0.098	0.099	−0.001	0.030	0.030
		BGMME	−0.003	0.049	0.049	−0.007	0.098	0.098	−0.001	0.028	0.028
		NGPMLE$_o$	−0.002	0.040	0.040	−0.009	0.082	0.083	−0.001	0.023	0.023
6	0.05	GPMLE	−0.003	0.052	0.053	−0.014	0.099	0.100	−0.001	0.029	0.029
		BGMME	−0.002	0.053	0.053	−0.006	0.101	0.101	−0.001	0.030	0.030
		NGPMLE$_o$	−0.001	0.041	0.041	−0.012	0.082	0.083	−0.001	0.023	0.023
4	0.4	GPMLE	−0.003	0.052	0.052	−0.013	0.098	0.099	−0.001	0.029	0.029
		BGMME	−0.003	0.052	0.052	−0.007	0.099	0.099	0.000	0.029	0.029
		NGPMLE$_o$	−0.003	0.050	0.050	−0.013	0.096	0.097	−0.001	0.028	0.028
4	0.05	GPMLE	−0.004	0.051	0.052	−0.013	0.099	0.100	−0.001	0.029	0.029
		BGMME	−0.003	0.052	0.052	−0.006	0.101	0.101	−0.001	0.029	0.029
		NGPMLE$_o$	−0.003	0.050	0.050	−0.012	0.097	0.098	−0.001	0.028	0.028
3.05	0.05	GPMLE	−0.004	0.052	0.052	−0.014	0.098	0.099	0.000	0.029	0.029
		BGMME	−0.003	0.053	0.053	−0.005	0.099	0.099	−0.001	0.030	0.030
		NGPMLE$_o$	−0.004	0.055	0.055	−0.013	0.101	0.102	−0.001	0.030	0.030

Notes: The true disturbance distribution is a fourth-order Gram–Charlier expansion of the standard normal distribution as a function of the skewness and kurtosis coefficients. β_2 is the coefficient on the nonintercept variable in X_n. $\lambda_0 = 0.4$, $\rho_0 = 0.2$, $\beta_{10} = 1$, $\beta_{20} = 1$, and $\sigma_0^2 = 0.25$.

are similar to the SDs. NGPMLE has a smaller SD than GPMLE when the kurtosis coefficient of innovations is equal to 4 or 6, whereas BGMME only has a smaller SD for λ and β_2 when the kurtosis and skewness coefficients are both the largest, i.e., the kurtosis coefficient is 6 and the skewness coefficient is 0.8. When the kurtosis coefficient is 3.05 and the skewness coefficient is 0.05 so that the distribution of innovations is close to the normal distribution, NGPMLE and BGMME have slightly larger SDs than GPMLE. For NGPMLE, a larger kurtosis leads to a smaller SD, whereas skewness does not have much impact on the SD.

Table 3 reports the estimation results for the SARAR model with a non-row-normalized M_n and symmetric innovations. In addition to $\mathrm{NGPMLE_o}$, GPMLE, and BGMME, we also consider $\mathrm{NGPMLE_a}$ to investigate its efficiency loss due to an added parameter. The patterns for the relative efficiencies of GPMLE, $\mathrm{NGPMLE_o}$ and BGMME are similar to those in Table 2. When the disturbance distribution is a mixture of two normal distributions with mean zero and the ratio of the variances for the two distributions being close to 1, or when the innovations follow the normal distribution, NGPMLE and BGMME have slightly larger SDs than that of GPMLE. While the $\mathrm{NGPMLE_a}$ of β_2 has a smaller SD than that of GPMLE in some cases, the $\mathrm{NGPMLE_a}$ of λ and ρ has a significantly larger SD than that of GPMLE in most cases, which is consistent with the efficiency comparisons based on numerical integration in Section 2.3.

Estimation results for the SAR model with symmetric innovations are presented in Table 4. The disturbance distribution is a mixture of two normal distributions with mean zero. The ratio of variances for the two normal distributions is 10, and the mixing probability is 0.3.[31] We consider two adaptive estimators (AE) proposed in Robinson (2010): AE_a and AE_b, where AE_b is a bias-corrected version of AE_a. As in Robinson (2010), we use the polynomial functions $(x,\ldots,x^L)$ or the bounded functions $(\frac{x}{(1+x^2)^{1/2}},\ldots,\frac{x^L}{(1+x^2)^{L/2}})$ to estimate the score function for the AEs. An AE_a with $(x,\ldots,x^L)$ is denoted by $\mathrm{AE}_a(p,L)$, and that with $(\frac{x}{(1+x^2)^{1/2}},\ldots,\frac{x^L}{(1+x^2)^{L/2}})$ is denoted by $\mathrm{AE}_a(b,L)$. AE_b is similarly denoted. We set L to 1, 2, or 4, as in Robinson (2010). The initial estimate for the AEs is either the NGPMLE or ordinary least-squares estimate (OLSE). Table 4 shows that, while the biases of GPMLE, BGMME, and NGPMLE are relatively small, those of AEs can be large. Some versions of AEs can have smaller SDs than GPMLE, but all AEs have uniformly larger SDs and RMSEs than NGPMLE.

4.2. Non-Gaussian and Gaussian Score Tests

Tables 5 and 6 report, respectively, the empirical sizes and powers of score tests for spatial dependence in the SARAR model with a row-normalized M_n and asymmetric innovations. With a nominal size of 5%, the size distortions of the non-Gaussian and Gaussian score tests are all within 0.5 percentage point. Neither the

[31] Results for some other parameter settings are reported in the Supplementary Material. The patterns are similar.

Table 3. Performance of various estimators for the SARAR model with a non-row-normalized M_n and symmetric v_i.

RV		λ			ρ			β_2		
		Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
	Panel A: $n = 147$									
9	GPMLE	−0.004	0.064	0.064	−0.046	0.172	0.178	−0.002	0.042	0.042
	BGMME	−0.003	0.067	0.067	−0.023	0.186	0.188	−0.002	0.042	0.042
	NGPMLE$_o$	−0.003	0.049	0.049	−0.035	0.146	0.150	−0.001	0.032	0.032
	NGPMLE$_a$	−0.007	0.080	0.080	−0.038	0.157	0.161	−0.002	0.032	0.032
6	GPMLE	−0.004	0.064	0.064	−0.044	0.171	0.176	−0.001	0.042	0.042
	BGMME	−0.004	0.069	0.069	−0.022	0.187	0.188	−0.001	0.043	0.043
	NGPMLE$_o$	−0.003	0.055	0.055	−0.038	0.158	0.162	−0.001	0.036	0.036
	NGPMLE$_a$	−0.010	0.092	0.093	−0.040	0.175	0.179	−0.002	0.037	0.037
3	GPMLE	−0.002	0.064	0.064	−0.044	0.168	0.173	−0.002	0.042	0.042
	BGMME	−0.002	0.068	0.068	−0.020	0.181	0.182	−0.002	0.043	0.043
	NGPMLE$_o$	−0.003	0.062	0.062	−0.041	0.166	0.171	−0.002	0.041	0.041
	NGPMLE$_a$	−0.010	0.104	0.105	−0.048	0.184	0.191	−0.003	0.041	0.042
1.1	GPMLE	−0.005	0.064	0.064	−0.047	0.171	0.178	−0.001	0.042	0.042
	BGMME	−0.004	0.068	0.068	−0.021	0.184	0.185	−0.001	0.043	0.043
	NGPMLE$_o$	−0.006	0.068	0.069	−0.045	0.175	0.181	−0.001	0.043	0.043
	NGPMLE$_a$	−0.009	0.115	0.116	−0.069	0.194	0.206	−0.002	0.047	0.047
	Panel B: $n = 294$									
9	GPMLE	−0.002	0.044	0.044	−0.024	0.115	0.118	0.000	0.030	0.030
	BGMME	−0.001	0.045	0.045	−0.013	0.119	0.119	0.000	0.030	0.030
	NGPMLE$_o$	−0.001	0.034	0.034	−0.018	0.097	0.099	0.000	0.023	0.023
	NGPMLE$_a$	−0.004	0.055	0.055	−0.018	0.109	0.110	−0.001	0.023	0.023
6	GPMLE	−0.002	0.044	0.044	−0.021	0.117	0.119	0.000	0.030	0.030
	BGMME	−0.002	0.045	0.045	−0.009	0.120	0.121	0.000	0.030	0.030
	NGPMLE$_o$	−0.002	0.038	0.038	−0.017	0.108	0.109	0.000	0.025	0.025
	NGPMLE$_a$	−0.005	0.061	0.061	−0.018	0.119	0.121	−0.001	0.026	0.026
3	GPMLE	−0.002	0.044	0.044	−0.021	0.118	0.119	0.000	0.029	0.029
	BGMME	−0.002	0.045	0.045	−0.009	0.121	0.121	0.000	0.030	0.030
	NGPMLE$_o$	−0.002	0.042	0.042	−0.021	0.117	0.119	0.000	0.028	0.028
	NGPMLE$_a$	−0.006	0.071	0.071	−0.022	0.132	0.133	−0.001	0.029	0.029
1.1	GPMLE	−0.001	0.044	0.044	−0.022	0.115	0.117	0.000	0.029	0.029
	BGMME	−0.001	0.045	0.045	−0.009	0.118	0.118	0.000	0.029	0.029
	NGPMLE$_o$	−0.002	0.044	0.044	−0.022	0.117	0.119	0.000	0.029	0.029
	NGPMLE$_a$	0.000	0.084	0.084	−0.043	0.152	0.158	−0.001	0.032	0.032

(continued)

TABLE 3. (continued)

RV	λ Bias	λ SD	λ RMSE	ρ Bias	ρ SD	ρ RMSE	β_2 Bias	β_2 SD	β_2 RMSE
Panel C: Normal innovations, $n = 147$									
GPMLE	−0.004	0.063	0.064	−0.046	0.172	0.179	−0.001	0.042	0.042
BGMME	−0.004	0.068	0.068	−0.021	0.186	0.187	−0.001	0.044	0.044
NGPMLE$_o$	−0.005	0.071	0.071	−0.045	0.179	0.184	−0.001	0.043	0.043
NGPMLE$_a$	−0.008	0.113	0.113	−0.069	0.197	0.208	−0.002	0.051	0.051
Panel D: Normal innovations, $n = 294$									
GPMLE	−0.001	0.044	0.044	−0.024	0.117	0.119	0.000	0.029	0.029
BGMME	−0.001	0.045	0.045	−0.010	0.120	0.120	0.000	0.030	0.030
NGPMLE$_o$	−0.002	0.046	0.046	−0.023	0.118	0.121	0.000	0.030	0.030
NGPMLE$_a$	−0.003	0.088	0.088	−0.041	0.151	0.157	−0.001	0.033	0.033

Notes: For Panels A and B, the true disturbance distribution is a mixture of two normal distributions with mean zero. The mixing probability of the two normal distributions is set to 0.3. “RV” denotes the ratio of variances of the two distributions. β_2 is the coefficient on the nonintercept variable in X_n. $\lambda_0 = 0.4$, $\rho_0 = 0.2$, $\beta_{10} = 1$, $\beta_{20} = 1$, and $\sigma_0^2 = 0.25$.

Gaussian score test nor the non-Gaussian score test dominates each other in terms of size distortions. For the empirical powers, we observe that the non-Gaussian score test is uniformly more powerful than the Gaussian score test, except for the case when the innovations are very close to be normally distributed. The power of each test increases as λ_0 or ρ_0 increases.

5. EMPIRICAL APPLICATION

In this section, we apply our NGPMLE to the well-known Harrison and Rubinfeld (1978) hedonic pricing data from the Boston Standard Metropolitan Statistical Area with 506 observations.[32] This dataset is popular in the spatial econometric literature. It has been used in textbooks such as LeSage (1999), LeSage and Pace (2009), and Arbia (2014).

Following LeSage (1999, p. 78), we estimate an SARAR model, where the dependent variable is the log median value of owner-occupied homes in \$1,000’s, and the explanatory variables include crime rate (CRIM), proportion of area zoned with large lots (ZN), proportion of nonretail business areas (INDUS), location contiguous to the Charles River (CHAS), squared levels of nitrogen oxides (NOX^2), squared average number of rooms (RM^2), proportion of structures built before 1940 (AGE), weighted distances to the employment centers (DIS), an index

[32] Available at http://lib.stat.cmu.edu/datasets/. Gilley and Pace (1996) corrected several miscoded observations and Pace and Gilley (1997) added the location of each tract in latitude and longitude. In the Supplementary Material, we also apply our NGPMLE to the crime dataset with 49 observations in Anselin (1988) and to the presidential election dataset with 3,107 observations in Pace and Barry (1997).

Table 4. Performance of various estimators for the SAR model with symmetric v_i.

	λ			β_2		
	Bias	SD	RMSE	Bias	SD	RMSE
GPMLE	−0.010	0.055	0.055	0.000	0.042	0.042
BGMME	−0.008	0.055	0.056	−0.002	0.042	0.042
NGPMLE$_o$	−0.007	0.043	0.044	0.000	0.032	0.032
AEs with GPMLE as the initial estimate						
$\text{AE}_a(p,1)$	0.100	0.066	0.120	−0.010	0.043	0.044
$\text{AE}_a(b,1)$	0.078	0.053	0.094	−0.008	0.034	0.035
$\text{AE}_b(p,1)$	0.327	0.105	0.343	−0.031	0.050	0.059
$\text{AE}_b(b,1)$	0.229	0.080	0.243	−0.023	0.038	0.044
$\text{AE}_a(p,2)$	0.095	0.067	0.116	−0.010	0.044	0.045
$\text{AE}_a(b,2)$	0.075	0.054	0.092	−0.008	0.035	0.036
$\text{AE}_b(p,2)$	0.308	0.103	0.325	−0.030	0.050	0.059
$\text{AE}_b(b,2)$	0.222	0.080	0.236	−0.022	0.039	0.045
$\text{AE}_a(p,4)$	0.074	0.062	0.096	−0.008	0.042	0.042
$\text{AE}_a(b,4)$	0.060	0.063	0.087	−0.006	0.042	0.042
$\text{AE}_b(p,4)$	0.235	0.088	0.251	−0.023	0.045	0.051
$\text{AE}_b(b,4)$	0.189	0.086	0.207	−0.018	0.044	0.048
AEs with OLSE as the initial estimate						
$\text{AE}_a(p,1)$	0.044	0.060	0.074	−0.005	0.042	0.042
$\text{AE}_a(b,1)$	0.022	0.048	0.053	−0.003	0.033	0.034
$\text{AE}_b(p,1)$	0.318	0.109	0.336	−0.030	0.050	0.058
$\text{AE}_b(b,1)$	0.206	0.081	0.222	−0.020	0.037	0.043
$\text{AE}_a(p,2)$	0.041	0.061	0.074	−0.005	0.043	0.044
$\text{AE}_a(b,2)$	0.021	0.050	0.054	−0.003	0.035	0.035
$\text{AE}_b(p,2)$	0.298	0.108	0.317	−0.029	0.050	0.058
$\text{AE}_b(b,2)$	0.199	0.081	0.215	−0.020	0.039	0.043
$\text{AE}_a(p,4)$	0.018	0.059	0.061	−0.003	0.041	0.041
$\text{AE}_a(b,4)$	0.013	0.060	0.061	−0.002	0.042	0.042
$\text{AE}_b(p,4)$	0.214	0.089	0.232	−0.021	0.044	0.049
$\text{AE}_b(b,4)$	0.170	0.085	0.190	−0.017	0.044	0.047

Notes: The true disturbance distribution is a mixture of two normal distributions with mean zero. The ratio of variances for the two normal distributions is 10, and the mixing probability is 0.3. β_2 is the coefficient on the nonintercept variable in X_n. $\lambda_0 = 0.4$, $\rho_0 = 0.2$, $\beta_{10} = 1$, $\beta_{20} = 1$, and $\sigma_0^2 = 0.25$. The sample size is 147.

Table 5. Empirical sizes of score tests for spatial dependence in the SARAR model with a row-normalized M_n and asymmetric v_i.

		$n = 147$		$n = 294$	
Kurtosis	Skewness	GPMLE	NGPMLE$_o$	GPMLE	NGPMLE$_o$
6	0.8	0.051	0.051	0.048	0.049
6	0.05	0.047	0.044	0.052	0.051
4	0.4	0.048	0.047	0.045	0.046
4	0.05	0.056	0.051	0.049	0.047
3.05	0.05	0.046	0.049	0.045	0.046

Notes: The nominal size is 5%. The true disturbance distribution is a fourth-order Gram–Charlier expansion of the standard normal distribution as a function of the skewness and kurtosis coefficients. $\beta_{10} = 1$, $\beta_{20} = 1$, and $\sigma_0^2 = 0.25$.

of accessibility (RAD), property tax rate (TAX), pupil–teacher ratio (PTRATIO), black population proportion (B), and lower status population proportion (LSTAT). All variables are normalized to have mean zero and unit variance as in LeSage (1999). The spatial weights matrix W_n is a first-order continuity matrix and row-normalized. The M_n is set to equal W_n.

Table 7 reports the empirical results. We carry out several diagnostic tests. First, a normality test of innovations rejects the null of normal innovations at the 1% level. With nonnormal innovations, the GPMLE will lose efficiency compared to a true ML estimator. We further test skewness and excess kurtosis of innovations.[33] While the null hypothesis of zero skewness is not rejected at any usual significance level, the null hypothesis of zero excess kurtosis is rejected at the 1% level. The estimated kurtosis coefficient is 5.751. These results show some evidence of symmetric and leptokurtic innovations for this dataset. GPMLE, BGMME, and NGPMLE have the same sign for each model parameter except the coefficient on INDUS, but their differences in magnitude can be relatively large.[34] For example, for the variable AGE, BGMME is about 60% larger in magnitude than GPMLE, whereas NGPMLE is more than three times that of GPMLE. The standard errors (SEs) of BGMME are very close to those of GPMLE, whereas the SEs of NGPMLE are about 30% smaller than those of GPMLE. Due to the differences in the estimates and SEs, for the variables NOX^2 and AGE, we observe different results on coefficient significance from different estimation methods. For the coefficient on NOX^2, GPMLE and BGMME are significant at the 1% level, whereas NGPMLE is significant only at the 10% level; for the coefficient on AGE, GPMLE is not significant at any usual significance level, BGMME is significant

[33] All the test statistics are derived in the Supplementary Material. The normality test is a special case of that for the SARAR model with parametric heteroskedasticity in Jin, Lee, and Yang (2022), which follows the Lagrange multiplier principle as in Jarque and Bera (1980). We present it in the Supplementary Material for completeness. The skewness and excess-kurtosis tests are on the basis of the delta method, as in Godfrey and Orme (1991).

[34] We only consider the NGPMLE with no added parameter, since the NGPMLE with an added parameter does not perform well in Monte Carlo experiments.

Table 6. Empirical powers of score tests for spatial dependence in the SARAR model with a row-normalized M_n and asymmetric v_i.

Kurtosis	Skewness		$\lambda_0 = 0$			$\rho_0 = 0$		
			$\rho_0 = 0.1$	$\rho_0 = 0.2$	$\rho_0 = 0.3$	$\lambda_0 = 0.1$	$\lambda_0 = 0.2$	$\lambda_0 = 0.3$
		Panel A: $n = 147$						
6	0.8	GPMLE	0.114	0.330	0.663	0.259	0.784	0.982
		NGPMLE$_o$	0.144	0.436	0.794	0.369	0.918	0.998
6	0.05	GPMLE	0.106	0.330	0.673	0.255	0.777	0.983
		NGPMLE$_o$	0.128	0.420	0.789	0.355	0.909	0.996
4	0.4	GPMLE	0.111	0.321	0.656	0.254	0.779	0.985
		NGPMLE$_o$	0.119	0.340	0.679	0.272	0.800	0.990
4	0.05	GPMLE	0.112	0.340	0.666	0.261	0.789	0.982
		NGPMLE$_o$	0.119	0.352	0.684	0.275	0.798	0.985
3.05	0.05	GPMLE	0.103	0.328	0.653	0.259	0.772	0.984
		NGPMLE$_o$	0.104	0.328	0.652	0.260	0.772	0.983
		Panel B: $n = 294$						
6	0.8	GPMLE	0.171	0.598	0.929	0.473	0.974	1.000
		NGPMLE$_o$	0.237	0.738	0.978	0.661	0.998	1.000
6	0.05	GPMLE	0.182	0.589	0.928	0.478	0.980	1.000
		NGPMLE$_o$	0.237	0.714	0.977	0.660	0.998	1.000
4	0.4	GPMLE	0.182	0.601	0.933	0.470	0.972	1.000
		NGPMLE$_o$	0.190	0.623	0.941	0.498	0.977	1.000
4	0.05	GPMLE	0.176	0.587	0.931	0.463	0.975	1.000
		NGPMLE$_o$	0.188	0.608	0.940	0.491	0.982	1.000
3.05	0.05	GPMLE	0.174	0.591	0.927	0.479	0.977	1.000
		NGPMLE$_o$	0.175	0.589	0.925	0.478	0.976	1.000

Notes: The true disturbance distribution is a fourth-order Gram–Charlier expansion of the standard normal distribution as a function of the skewness and kurtosis coefficients. $\beta_{10} = 1$, $\beta_{20} = 1$, and $\sigma_0^2 = 0.25$.

at the 5% level, whereas NGPMLE is significant at the 1% level. These differences in coefficient significance also carry over to impact measures such as the average total, direct, and indirect impacts, which we report in the Supplementary Material. Overall, the application shows that more efficient estimation methods for the SARAR model can be valuable in practice.

6. CONCLUSIONS

This study considers the non-Gaussian PML estimation of the SARAR model. If the spatial weights matrix M_n in the SAR process of disturbances is row-normalized or the model reduces to the SAR model with no SAR process of disturbances, the NGPMLE for model parameters except the intercept term and the

TABLE 7. Empirical results for the hedonic pricing data.

	GPMLE		BGMME		NGPMLE	
	Estimate	SE	Estimate	SE	Estimate	SE
λ	0.188***	0.060	0.267***	0.055	0.121***	0.044
ρ	0.626***	0.061	0.612***	0.062	0.673***	0.048
CRIM	−0.187***	0.023	−0.177***	0.023	−0.166***	0.015
ZN	0.065**	0.031	0.063**	0.031	0.046**	0.021
INDUS	0.016	0.046	−0.001	0.046	0.001	0.031
CHAS	−0.007	0.021	−0.010	0.021	−0.014	0.014
NOX^2	−0.191***	0.055	−0.310***	0.055	−0.071*	0.038
RM^2	0.199***	0.024	0.193***	0.024	0.415***	0.016
AGE	−0.046	0.036	−0.074**	0.037	−0.161***	0.025
DIS	−0.256***	0.055	−0.207***	0.055	−0.172***	0.039
RAD	0.342***	0.061	0.387***	0.061	0.202***	0.041
TAX	−0.259***	0.057	−0.234***	0.057	−0.213***	0.038
PTRATIO	−0.127***	0.030	−0.103***	0.030	−0.079***	0.021
B	0.119***	0.026	0.131***	0.026	0.152***	0.018
LSTAT	−0.378***	0.035	−0.373***	0.035	−0.155***	0.023

Test for normality of innovations:
Test statistic: 67.742; p-value: 0.000.

Test for skewness of innovations:
Test statistic: 0.755; p-value: 0.450; estimated skewness coefficient = 0.267.

Test for excess kurtosis of innovations:
Test statistic: 4.115; p-value: 0.000; estimated kurtosis coefficient = 5.751.

Notes: *, **, and *** denote significance at, respectively, the 10%, 5%, and 1% levels.

variance parameter σ^2 is consistent. If M_n is not row-normalized but innovations are symmetric, the NGPMLE for model parameters except σ^2 is consistent. With neither row-normalization of M_n nor the symmetry of innovations, a location parameter can be added to the non-Gaussian pseudo log-likelihood function to obtain consistent estimation of model parameters except σ^2. We formally prove the convergence and asymptotic normality of the NGPMLE. An advantage of the NGPMLE is that it can have a significant efficiency improvement upon the GPMLE and BGMME. We also propose a non-Gaussian score test for spatial dependence, which is locally more powerful than the Gaussian score test when the NGPMLE is more efficient than the GPMLE. Using Student's t distribution to formulate the non-Gaussian likelihood function, our numerical integration and Monte Carlo results show that the NGPMLE with no added parameter can have a significant efficiency improvement upon the GPMLE and BGMME, but the NGPMLE with an added parameter can be less efficient than the GPMLE. The non-Gaussian score test based on the NGPMLE with no added parameter is more powerful than the Gaussian score test in finite samples. Therefore, we recommend the use of the NGPMLE with no added parameter and the non-Gaussian score test based on it when they are applicable.

APPENDIX A. Expressions for Asymptotic Variances

In this appendix, we present the expressions for asymptotic variances of NGPMLEs in Theorem 2.

A.1. Row-Normalized M_n

For model (1) with a row-normalized M_n, the NGPMLE maximizes $\ln L_n(\gamma)$ in (2). Note that $v_i(\theta_*) = \frac{\sigma_0}{\sigma_\infty} v_i + \frac{\sigma_0}{\sigma_\infty} c_v$, where $c_v = -\frac{1}{\sigma_0}(1-\rho_0)(\beta_{1\infty} - \beta_{10})$. Denote $\zeta_{1i} = \frac{\sigma_0}{\sigma_\infty}\frac{\partial \ln f(v_i(\theta_*), \eta_\infty)}{\partial v}$, $\zeta_{2i} = \zeta_{1i} v_i + 1$, $\zeta_{3i} = -\frac{\partial \ln f(v_i(\theta_*), \eta_\infty)}{\partial \eta}$, $\zeta_{4i} = -\frac{\sigma_0^2}{\sigma_\infty^2}\frac{\partial^2 \ln f(v_i(\theta_*), \eta_\infty)}{\partial v^2}$, $\zeta_{5i} = \frac{\sigma_0}{\sigma_\infty}\frac{\partial^2 \ln f(v_i(\theta_*), \eta_\infty)}{\partial v \partial \eta}$, $\zeta_{6i} = -\frac{\partial^2 \ln f(v_i(\theta_*), \eta_\infty)}{\partial \eta \partial \eta'}$, $D_n = R_n W_n S_n^{-1} R_n^{-1} = [d_{n,ij}]$, $Z_n = M_n R_n^{-1} = [z_{n,ij}]$, $Q_n = \frac{1}{\sigma_0} R_n W_n S_n^{-1} X_n \beta_0 = [q_{ni}]$, and $c_\beta = -\frac{1}{\sigma_0}(\beta_{1\infty} - \beta_{10})$. By Assumption 4(i)(c), $\mathrm{E}(\zeta_{ji}) = 0$, for $j = 1, 2, 3$. For any two subvectors δ_1 and δ_2 of δ, let $\mathcal{B}_{\delta_1\delta_2} = \frac{1}{n}\mathrm{E}(\frac{\partial \ln L_n(\gamma_*)}{\partial \delta_1}\frac{\partial \ln L_n(\gamma_*)}{\partial \delta_2'})$ and $\mathcal{A}_{\delta_1\delta_2} = -\frac{1}{n}\mathrm{E}(\frac{\partial^2 \ln L_n(\gamma_*)}{\partial \delta_1 \partial \delta_2'})$.

For the expression of $\mathcal{B}$, using the reduced form of Y_n, we have $\frac{\partial \ln L_n(\gamma_*)}{\partial \lambda} = -\sum_{i=1}^n q_{ni}\zeta_{1i} - \sum_{i=1}^n d_{n,ii}\zeta_{2i} - \sum_{i=1}^n \zeta_{1i}\sum_{j\neq i} d_{n,ij} v_j$, $\frac{\partial \ln L_n(\gamma_*)}{\partial \rho} = -\sum_{i=1}^n c_\beta \zeta_{1i} - \sum_{i=1}^n z_{n,ii}\zeta_{2i} - \sum_{i=1}^n \zeta_{1i}\sum_{j\neq i} z_{n,ij} v_j$, $\frac{\partial \ln L_n(\gamma_*)}{\partial \beta} = -\frac{1}{\sigma_0}\sum_{i=1}^n \zeta_{1i} X_n' R_n' e_i$, $\frac{\partial \ln L_n(\gamma_*)}{\partial \sigma^2} = -\frac{1}{2\sigma_\infty^2}\sum_{i=1}^n (c_v \zeta_{1i} + \zeta_{2i})$, and $\frac{\partial \ln L_n(\gamma_*)}{\partial \eta} = -\sum_{i=1}^n \zeta_{3i}$. Then,

$$
\begin{aligned}
\mathcal{B}_{\lambda\lambda} &= \frac{1}{n}\mathrm{E}(\zeta_{1i}^2)\sum_{i=1}^n q_{ni}^2 + \frac{2}{n}\mathrm{E}(\zeta_{1i}\zeta_{2i})\sum_{i=1}^n q_{ni} d_{n,ii} + \frac{1}{n}\mathrm{E}(\zeta_{2i}^2)\sum_{i=1}^n d_{n,ii}^2 + \frac{1}{n}\mathrm{E}(\zeta_{1i}^2)\sum_{i=1}^n\sum_{j\neq i} d_{n,ij}^2 \\
&\quad + \frac{1}{n}\sum_{i=1}^n\sum_{j\neq i} d_{n,ij} d_{n,ji} \\
&= \frac{1}{n}\mathrm{E}(\zeta_{1i}^2) Q_n' Q_n + \frac{2}{n}\mathrm{E}(\zeta_{1i}\zeta_{2i}) Q_n' \mathrm{vec}_{\mathrm{D}}(D_n) \\
&\quad + \frac{1}{n}[\mathrm{E}(\zeta_{2i}^2) - \mathrm{E}(\zeta_{1i}^2) - 1]\mathrm{vec}_{\mathrm{D}}{}'(D_n)\mathrm{vec}_{\mathrm{D}}(D_n) + \frac{1}{n}\mathrm{E}(\zeta_{1i}^2)\operatorname{tr}(D_n' D_n) + \frac{1}{n}\operatorname{tr}(D_n^2), \\
\mathcal{B}_{\lambda\rho} &= \frac{c_\beta}{n}\mathrm{E}(\zeta_{1i}^2) Q_n' 1_n + \frac{1}{n}\mathrm{E}(\zeta_{1i}\zeta_{2i})[Q_n' \mathrm{vec}_{\mathrm{D}}(Z_n) + c_\beta \operatorname{tr}(D_n)] \\
&\quad + \frac{1}{n}[\mathrm{E}(\zeta_{2i}^2) - \mathrm{E}(\zeta_{1i}^2) - 1]\mathrm{vec}_{\mathrm{D}}{}'(D_n)\mathrm{vec}_{\mathrm{D}}(Z_n) + \frac{1}{n}\mathrm{E}(\zeta_{1i}^2)\operatorname{tr}(D_n' Z_n) + \frac{1}{n}\operatorname{tr}(D_n Z_n),
\end{aligned}
$$

$\mathcal{B}_{\lambda\beta} = \frac{1}{n\sigma_0}\mathrm{E}(\zeta_{1i}^2) Q_n' R_n X_n + \frac{1}{n\sigma_0}\mathrm{E}(\zeta_{1i}\zeta_{2i})\mathrm{vec}_{\mathrm{D}}{}'(D_n) R_n X_n$, $\mathcal{B}_{\lambda\sigma^2} = \frac{1}{2n\sigma_\infty^2}[c_v \mathrm{E}(\zeta_{1i}^2) + \mathrm{E}(\zeta_{1i}\zeta_{2i})] Q_n' 1_n + \frac{1}{2n\sigma_\infty^2}[c_v \mathrm{E}(\zeta_{1i}\zeta_{2i}) + \mathrm{E}(\zeta_{2i}^2)]\operatorname{tr}(D_n)$, $\mathcal{B}_{\lambda\eta} = \frac{1}{n}\mathrm{E}(\zeta_{1i}\zeta_{3i}') Q_n' 1_n + \frac{1}{n}\mathrm{E}(\zeta_{2i}\zeta_{3i}')\operatorname{tr}(D_n)$,

$$\mathcal{B}_{\rho\rho} = c_\beta^2 \mathrm{E}(\zeta_{1i}^2) + \frac{2c_\beta}{n}\mathrm{E}(\zeta_{1i}\zeta_{2i})\,\mathrm{tr}(Z_n) + \frac{1}{n}[\mathrm{E}(\zeta_{2i}^2) - \mathrm{E}(\zeta_{1i}^2) - 1]\mathrm{vec}_{\mathrm{D}}{}'(Z_n)\mathrm{vec}_{\mathrm{D}}(Z_n) + \frac{1}{n}\mathrm{E}(\zeta_{1i}^2)\,\mathrm{tr}(Z_n'Z_n) + \frac{1}{n}\mathrm{tr}(Z_n^2),$$

$\mathcal{B}_{\rho\beta} = \frac{c_\beta}{n\sigma_0}\mathrm{E}(\zeta_{1i}^2)1_n'R_nX_n + \frac{1}{n\sigma_0}\mathrm{E}(\zeta_{1i}\zeta_{2i})\mathrm{vec}_{\mathrm{D}}{}'(Z_n)R_nX_n$, $\mathcal{B}_{\rho\sigma^2} = \frac{c_\beta}{2\sigma_\infty^2}[c_v\mathrm{E}(\zeta_{1i}^2) + \mathrm{E}(\zeta_{1i}\zeta_{2i})] + \frac{1}{2n\sigma_\infty^2}[c_v\mathrm{E}(\zeta_{1i}\zeta_{2i}) + \mathrm{E}(\zeta_{2i}^2)]\,\mathrm{tr}(Z_n)$, $\mathcal{B}_{\rho\eta} = c_\beta\mathrm{E}(\zeta_{1i}\zeta_{3i}') + \frac{1}{n}\mathrm{E}(\zeta_{2i}\zeta_{3i}')\,\mathrm{tr}(Z_n)$, $\mathcal{B}_{\beta\beta} = \frac{1}{n\sigma_0^2}\mathrm{E}(\zeta_{1i}^2)X_n'R_n'R_nX_n$, $\mathcal{B}_{\beta\sigma^2} = \frac{1}{2n\sigma_\infty^2\sigma_0}[c_v\mathrm{E}(\zeta_{1i}^2) + \mathrm{E}(\zeta_{1i}\zeta_{2i})]X_n'R_n'1_n$, $\mathcal{B}_{\beta\eta} = \frac{1}{n\sigma_0}X_n'R_n'1_n\mathrm{E}(\zeta_{1i}\zeta_{3i}')$, $\mathcal{B}_{\sigma^2\sigma^2} = \frac{1}{4\sigma_\infty^4}[c_v^2\mathrm{E}(\zeta_{1i}^2) + 2c_v\mathrm{E}(\zeta_{1i}\zeta_{2i}) + \mathrm{E}(\zeta_{2i}^2)]$, $\mathcal{B}_{\sigma^2\eta} = \frac{1}{2\sigma_\infty^2}[c_v\mathrm{E}(\zeta_{1i}\zeta_{3i}') + \mathrm{E}(\zeta_{2i}\zeta_{3i}')]$ and $\mathcal{B}_{\eta\eta} = \mathrm{E}(\zeta_{3i}\zeta_{3i}')$.

For the expression of $\mathcal{A}$, using the explicit form of $\frac{\partial \ln L_n(\gamma)}{\partial\gamma\partial\gamma'}$ in the Supplementary Material and the reduced form of Y_n, we have

$$\mathcal{A}_{\lambda\lambda} = \frac{1}{n}\mathrm{E}(\zeta_{4i})Q_n'Q_n + \frac{2}{n}\mathrm{E}(\zeta_{4i}v_i)Q_n'\mathrm{vec}_{\mathrm{D}}(D_n) + \frac{1}{n}[\mathrm{E}(\zeta_{4i}v_i^2) - \mathrm{E}(\zeta_{4i})]\mathrm{vec}_{\mathrm{D}}{}'(D_n)\mathrm{vec}_{\mathrm{D}}(D_n) + \frac{1}{n}\mathrm{E}(\zeta_{4i})\,\mathrm{tr}(D_n'D_n) + \frac{1}{n}\mathrm{tr}(D_n^2),$$

$$\mathcal{A}_{\lambda\rho} = \frac{c_\beta}{n}\mathrm{E}(\zeta_{4i})Q_n'1_n + \frac{1}{n}\mathrm{E}(\zeta_{4i}v_i)[Q_n'\mathrm{vec}_{\mathrm{D}}(Z_n) + c_\beta\,\mathrm{tr}(D_n)] + \frac{1}{n}[\mathrm{E}(\zeta_{4i}v_i^2) - \mathrm{E}(\zeta_{4i})]\mathrm{vec}_{\mathrm{D}}{}'(D_n)\mathrm{vec}_{\mathrm{D}}(Z_n) + \frac{1}{n}\mathrm{E}(\zeta_{4i})\,\mathrm{tr}(D_n'Z_n) + \frac{1}{n}\mathrm{tr}(D_nZ_n),$$

$$\mathcal{A}_{\lambda\beta} = \frac{1}{n\sigma_0}\mathrm{E}(\zeta_{4i})Q_n'R_nX_n + \frac{1}{n\sigma_0}\mathrm{E}(\zeta_{4i}v_i)\mathrm{vec}_{\mathrm{D}}{}'(D_n)R_nX_n,$$

$$\mathcal{A}_{\lambda\sigma^2} = \frac{1}{2n\sigma_\infty^2}[c_v\mathrm{E}(\zeta_{4i}) + \mathrm{E}(\zeta_{4i}v_i)]Q_n'1_n + \frac{1}{2n\sigma_\infty^2}[c_v\mathrm{E}(\zeta_{4i}v_i) + \mathrm{E}(\zeta_{4i}v_i^2) + 1]\,\mathrm{tr}(D_n),$$

$$\mathcal{A}_{\lambda\eta} = \frac{1}{n}\mathrm{E}(\zeta_{5i}')Q_n'1_n + \frac{1}{n}\mathrm{E}(v_i\zeta_{5i}')\,\mathrm{tr}(D_n),$$

$$\mathcal{A}_{\rho\rho} = c_\beta^2\mathrm{E}(\zeta_{4i}) + \frac{2c_\beta}{n}\mathrm{E}(\zeta_{4i}v_i)\,\mathrm{tr}(Z_n) + \frac{1}{n}[\mathrm{E}(\zeta_{4i}v_i^2) - \mathrm{E}(\zeta_{4i})]\mathrm{vec}_{\mathrm{D}}{}'(Z_n)\mathrm{vec}_{\mathrm{D}}(Z_n) + \frac{1}{n}\mathrm{E}(\zeta_{4i})\,\mathrm{tr}(Z_n'Z_n) + \frac{1}{n}\mathrm{tr}(Z_n^2),$$

$\mathcal{A}_{\rho\beta} = \frac{c_\beta}{n\sigma_0}\mathrm{E}(\zeta_{4i})1_n'R_nX_n + \frac{1}{n\sigma_0}\mathrm{E}(\zeta_{4i}v_i)\mathrm{vec}_{\mathrm{D}}{}'(Z_n)R_nX_n$, $\mathcal{A}_{\rho\sigma^2} = \frac{c_\beta}{2\sigma_\infty^2}[c_v\mathrm{E}(\zeta_{4i}) + \mathrm{E}(\zeta_{4i}v_i)] + \frac{1}{2n\sigma_\infty^2}[c_v\mathrm{E}(\zeta_{4i}v_i) + \mathrm{E}(\zeta_{4i}v_i^2) + 1]\,\mathrm{tr}(Z_n)$, $\mathcal{A}_{\rho\eta} = c_\beta\mathrm{E}(\zeta_{5i}') + \frac{1}{n}\mathrm{E}(v_i\zeta_{5i}')\,\mathrm{tr}(Z_n)$, $\mathcal{A}_{\beta\beta} = \frac{1}{n\sigma_0^2}\mathrm{E}(\zeta_{4i})X_n'R_n'R_nX_n$, $\mathcal{A}_{\beta\sigma^2} = \frac{1}{2n\sigma_\infty^2\sigma_0}[c_v\mathrm{E}(\zeta_{4i}) + \mathrm{E}(\zeta_{4i}v_i)]X_n'R_n'1_n$, $\mathcal{A}_{\beta\eta} = \frac{1}{n\sigma_0}X_n'R_n'1_n\mathrm{E}(\zeta_{5i}')$, $\mathcal{A}_{\sigma^2\sigma^2} = \frac{1}{4\sigma_\infty^4}[c_v^2\mathrm{E}(\zeta_{4i}) + 2c_v\mathrm{E}(\zeta_{4i}v_i) + \mathrm{E}(\zeta_{4i}v_i^2) + 1]$, $\mathcal{A}_{\sigma^2\eta} = \frac{1}{2\sigma_\infty^2}[c_v\mathrm{E}(\zeta_{5i}') + \mathrm{E}(v_i\zeta_{5i}')]$, and $\mathcal{A}_{\eta\eta} = \mathrm{E}(\zeta_{6i})$.

A.2. Symmetric v_i

As in the last subsection, the NGPMLE in this case maximizes $\ln L_n(\gamma)$ in (2). In this and the next subsections, let ζ_{1i} to ζ_{6i} be as defined in the last subsection except that $v_i(\theta^*)$ is replaced by $\frac{\sigma_0}{\sigma_\infty}v_i$. It is shown in the proof of Theorem 2 that $\mathrm{E}(\zeta_{ji}) = 0$, for $j = 1, 2, 3$. Then the expressions of $\mathcal{A}$ and $\mathcal{B}$ are the same as those in the last subsection, except the additional restrictions $c_v = 0$ and $c_\beta = 0$.

With symmetric v_i, it is shown in the proof of Corollary 1 that $\mathrm{E}(\zeta_{1i}\zeta_{2i}) = 0$, $\mathrm{E}(\zeta_{1i}\zeta_{3i}) = 0$, $\mathrm{E}(\zeta_{4i}v_i) = 0$, and $\mathrm{E}(\zeta_{5i}) = 0$. Then $\mathcal{A}_{\beta\rho} = 0$, $\mathcal{A}_{\beta\sigma^2} = 0$, $\mathcal{A}_{\beta\eta} = 0$, $\mathcal{B}_{\beta\rho} = 0$, $\mathcal{B}_{\beta\sigma^2} = 0$, and $\mathcal{B}_{\beta\eta} = 0$.

In the case that $\tau_0 = 0$, $D_n = W_n$ and $Z_n = M_n$. As W_n and M_n have zero diagonals, $\mathrm{vec_D}(D_n) = 0$, $\mathrm{vec_D}(Z_n) = 0$, $\mathrm{tr}(D_n) = 0$, and $\mathrm{tr}(T_n) = 0$. Then some components of $\mathcal{A}$ and $\mathcal{B}$ can be simplified accordingly. In particular, $\mathcal{A}_{\rho\sigma^2} = 0$, $\mathcal{A}_{\rho\eta} = 0$, $\mathcal{B}_{\rho\sigma^2} = 0$, and $\mathcal{B}_{\rho\eta} = 0$.

A.3. Non-Row-Normalized M_n and Asymmetric v_i

In this case, the NGPMLE maximizes $\ln L_n(\delta)$ in (4). By Assumption 4(i)(c), $\mathrm{E}(\zeta_{ji}) = 0$, for $j = 1,2,3$. The expressions of $\mathcal{B}_{\delta_1\delta_2} = \frac{1}{n}\mathrm{E}(\frac{\partial \ln L_n(\delta_\#)}{\partial \delta_1}\frac{\partial \ln L_n(\delta_\#)}{\partial \delta_2'})$ and $\mathcal{A}_{\delta_1\delta_2} = -\frac{1}{n}\mathrm{E}(\frac{\partial^2 \ln L_n(\delta_\#)}{\partial \delta_1 \partial \delta_2'})$ for δ_1 and δ_2 not containing α can be derived by imposing $c_v = -\frac{\alpha_\infty}{\sigma_0}$ and $c_\beta = 0$ in the corresponding expressions in Appendix A.1. The remaining components of $\mathcal{B}$ are $\mathcal{B}_{\alpha\lambda} = \frac{1}{n\sigma_0}\mathrm{E}(\zeta_{1i}^2)Q_n' 1_n + \frac{1}{n\sigma_0}\mathrm{E}(\zeta_{1i}\zeta_{2i})\,\mathrm{tr}(D_n)$, $\mathcal{B}_{\alpha\rho} = \frac{1}{n\sigma_0}\mathrm{E}(\zeta_{1i}\zeta_{2i})\,\mathrm{tr}(Z_n)$, $\mathcal{B}_{\alpha\beta} = \frac{1}{n\sigma_0^2}\mathrm{E}(\zeta_{1i}^2)1_n' R_n X_n$, $\mathcal{B}_{\alpha\sigma^2} = \frac{1}{2\sigma_\infty^2\sigma_0}[-\frac{\alpha_\infty}{\sigma_0}\mathrm{E}(\zeta_{1i}^2) + \mathrm{E}(\zeta_{1i}\zeta_{2i})]$, $\mathcal{B}_{\alpha\alpha} = \frac{1}{\sigma_0^2}\mathrm{E}(\zeta_{1i}^2)$, and $\mathcal{B}_{\alpha\eta} = \frac{1}{\sigma_0}\mathrm{E}(\zeta_{1i}\zeta_{3i}')$. The remaining components of $\mathcal{A}$ are $\mathcal{A}_{\alpha\lambda} = \frac{1}{n\sigma_0}\mathrm{E}(\zeta_{4i})Q_n' 1_n + \frac{1}{n\sigma_0}\mathrm{E}(\zeta_{4i}v_i)\,\mathrm{tr}(D_n)$, $\mathcal{A}_{\alpha\rho} = \frac{1}{n\sigma_0}\mathrm{E}(\zeta_{4i}v_i)\,\mathrm{tr}(Z_n)$, $\mathcal{A}_{\alpha\beta} = \frac{1}{n\sigma_0^2}\mathrm{E}(\zeta_{4i})1_n' R_n X_n$, $\mathcal{A}_{\alpha\sigma^2} = \frac{1}{2\sigma_\infty^2\sigma_0}[-\frac{\alpha_\infty}{\sigma_0}\mathrm{E}(\zeta_{4i}) + \mathrm{E}(\zeta_{4i}v_i)]$, $\mathcal{A}_{\alpha\alpha} = \frac{1}{\sigma_0^2}\mathrm{E}(\zeta_{4i})$, and $\mathcal{A}_{\alpha\eta} = \frac{1}{\sigma_0}\mathrm{E}(\zeta_{5i}')$.

APPENDIX B. Lemmas

The following Lemma B.1 provides more primitive conditions for $g_n(\tau) > 0$ at $\tau \neq \tau_0$ in a neighborhood of τ_0, where $g_n(\tau)$ is in Assumption 3. The matrices T_{1n} and T_{2n} below are defined after Assumption 3.

LEMMA B.1. *Suppose that $W_n = M_n$ and that T_{1n} and T_{2n} are linearly independent. If W_n is symmetric or is row-normalized from a symmetric matrix, then $g_n(\tau) > 0$ at $\tau \neq \tau_0$ in a neighborhood of τ_0.*

Proof. As explained below Assumption 3, we need to show that $\frac{\partial^2 g_n(\tau_0)}{\partial\tau\partial\tau'}$ is positive-definite, which requires that $\mathrm{tr}(T_{1n}^2) > 0$, $\mathrm{tr}(T_{2n}^2) > 0$, and $\mathrm{tr}(T_{1n}^2)\,\mathrm{tr}(T_{2n}^2) > \mathrm{tr}^2(T_{1n}T_{2n})$, by some calculation.

If W_n is symmetric, with $W_n = M_n$, it is obvious that T_{1n} and T_{2n} are symmetric. Then $\mathrm{tr}(T_{jn}^2) = \mathrm{tr}(T_{jn}'T_{jn}) \geq 0$, for $j = 1,2$. By the Cauchy–Schwarz inequality, $\mathrm{tr}(T_{1n}^2)\,\mathrm{tr}(T_{2n}^2) = \mathrm{tr}(T_{1n}'T_{1n})\,\mathrm{tr}(T_{2n}'T_{2n}) \geq \mathrm{tr}^2(T_{1n}'T_{2n}) = \mathrm{tr}^2(T_{1n}T_{2n})$. The inequality is strict when T_{1n} and T_{2n} are linearly independent, which also implies that $\mathrm{tr}(T_{jn}^2) > 0$, for $j = 1,2$.

If W_n is row-normalized from a symmetric matrix such that $W_n = H_nA_n$, where $H_n = \mathrm{diag}(1/(e_{n1}'A_n1_n), \ldots, 1/(e_{nn}'A_n1_n))$ and A_n is symmetric, let $B_n = H_n^{1/2}A_nH_n^{1/2}$

and $C_n(\lambda) = I_n - \lambda B_n$. Then B_n and $C_n(\lambda)$ are symmetric and satisfy $B_n C_n(\lambda) = C_n(\lambda) B_n$. We have $S_n(\lambda) = H_n^{1/2} C_n(\lambda) H_n^{-1/2}$, $A_{1n} = H_n A_n \cdot H_n^{1/2} C_n^{-1}(\rho_0) H_n^{-1/2} = H_n^{1/2} B_n C_n^{-1}(\rho_0) H_n^{-1/2}$, and $A_{2n} = H_n^{1/2} C_n(\rho_0) H_n^{-1/2} \cdot H_n A_n \cdot H_n^{1/2} C_n^{-1}(\lambda_0) H_n^{-1/2} \cdot H_n^{1/2} C_n^{-1}(\rho_0) H_n^{-1/2} = H_n^{1/2} B_n C_n^{-1}(\lambda_0) H_n^{-1/2}$. For $n \times n$ matrices E_{1n} and E_{2n}, if E_{2n} is diagonal, then $\mathrm{diag}(E_{1n}E_{2n}) = \mathrm{diag}(E_{1n})E_{2n}$. Thus, $T_{1n} = H_n^{1/2} D_{1n} H_n^{-1/2}$, $T_{2n} = H_n^{1/2} D_{2n} H_n^{-1/2}$, and $T_{1n}T_{2n} = H_n^{1/2} D_{1n} D_{2n} H_n^{-1/2}$, where $D_{1n} = B_n C_n^{-1}(\rho_0) - \mathrm{diag}(B_n C_n^{-1}(\rho_0))$ and $D_{2n} = B_n C_n^{-1}(\lambda_0) - \mathrm{diag}(B_n C_n^{-1}(\lambda_0))$ are symmetric. Thus, $\mathrm{tr}(T_{jn}^2) = \mathrm{tr}(D_{jn}^2) = \mathrm{tr}(D_{jn}' D_{jn}) \geq 0$, for $j = 1, 2$. Furthermore, $\mathrm{tr}(T_{1n}^2)\,\mathrm{tr}(T_{2n}^2) = \mathrm{tr}(D_{1n}^2)$ $\mathrm{tr}(D_{2n}^2) = \mathrm{tr}(D_{1n}' D_{1n})\,\mathrm{tr}(D_{2n}' D_{2n}) \geq \mathrm{tr}^2(D_{1n}' D_{2n}) = \mathrm{tr}^2(T_{1n}T_{2n})$ by the Cauchy–Schwarz inequality. The inequality is strict when D_{1n} and D_{2n} are linearly independent, i.e., T_{1n} and T_{2n} are linearly independent, which also implies that $\mathrm{tr}(T_{jn}^2) > 0$, for $j = 1, 2$. □

LEMMA B.2. *For $j = 1, \ldots, l$, let A_{jn} be $n \times n$ nonstochastic matrices that are bounded in the row-sum norm, and let $U_{jn} = [u_{jn,1}, \ldots, u_{jn,n}]'$ be $n \times 1$ vectors such that $\sup_{i,j,n} \mathrm{E}(|u_{jn,i}|^{a_j}) < \infty$, for $a_j > 1$. Then $\sup_{i,n} \mathrm{E}[(\prod_{j=1}^l |e_{ni}' A_{jn} U_{jn}|)^{1/\sum_{j=1}^l \frac{1}{a_j}}] < \infty$.*

Proof. This is a special case of Lemma 1(ii) in Jin and Lee (2019). □

LEMMA B.3. *Suppose that $h(x)$ is a scalar function, v_i's in $V_n = [v_1, \ldots, v_n]'$ are i.i.d. with mean zero and variance σ_0^2, $A_n = [a_{n,ij}]$ and $B_n = [b_{n,ij}]$ are $n \times n$ nonstochastic matrices that are bounded in both the row- and column-sum norms, and $\mathrm{E}(|v_i|^{c_v}) < \infty$ and $\mathrm{E}(|h(v_i)|^{c_h}) < \infty$, for some $c_v > 0$ and $c_h > 0$. Then $c_{1n} - \mathrm{E}(c_{1n}) = o_P(1)$ if $\frac{1}{c_h} + \frac{2}{c_v} < 1$, and $c_{2n} - \mathrm{E}(c_{2n}) = o_P(1)$ if $\frac{1}{c_h} + \frac{1}{c_v} < 1$, where $c_{1n} = \frac{1}{n}\sum_{i=1}^n h(v_i)(\sum_{j=1}^n a_{n,ij} v_j)(\sum_{k=1}^n b_{n,ik} v_k)$ and $c_{2n} = \frac{1}{n}\sum_{i=1}^n h(v_i)$ $(\sum_{j=1}^n a_{n,ij} v_j)$.*

Proof. This lemma is proved by an LLN for martingale differences. The details are in the Supplementary Material. □

LEMMA B.4. *Suppose that $A_n = [a_{n,ij}]$ is an $n \times n$ nonstochastic matrix that is bounded in both the row- and column-sum norms; $b_n = [b_{ni}]$ is an $n \times 1$ vector of uniformly bounded constants; $\varepsilon_n = [\epsilon_{ni}]$, $V_n = [v_{ni}]$, and $\Psi_n = [\psi_{ni}]$ are $n \times 1$ random vectors with mean zero; $[\epsilon_{ni}, v_{ni}, \psi_{ni}]$, for $i = 1, \ldots, n$, are independent; and $\sup_{i,n} \mathrm{E}(|\epsilon_{ni} v_{ni}|^{2+\iota}) + \sup_{i,n} \mathrm{E}(|\epsilon_{ni}|^{2+\iota}) + \sup_{i,n} \mathrm{E}(|v_{ni}|^{2+\iota}) + \sup_{i,n} \mathrm{E}(|\psi_{ni}|^{2+\iota}) < \infty$, for some $\iota > 0$. Let $\omega_n = \varepsilon_n' A_n V_n + b_n' \Psi_n - \mathrm{E}(\varepsilon_n' A_n V_n)$ and $\sigma_{\omega_n}^2 = \mathrm{var}(\omega_n)$. If $\inf_n \frac{1}{n}\sigma_{\omega_n}^2 > 0$, then $\frac{\omega_n}{\sigma_{\omega_n}} \xrightarrow{d} N(0,1)$.*

Proof. This lemma is a special case of Lemma 6 in Yang and Lee (2017). □

LEMMA B.5. *Suppose that Assumption 1 holds. Let each of $A_n = [a_{n,ij}]$ and $B_n = [b_{n,ij}]$ be one of the matrices W_n, M_n, R_n, and S_n. Denote $C_n = A_n B_n = [c_{n,ij}]$. If $\lim_{r\to\infty} \sup_{i,n} \sum_{j:\, d(i,j)>r} |a_{n,ij}| = 0$, $\lim_{r\to\infty} \sup_{i,n} \sum_{j:\, d(i,j)>r} |b_{n,ij}| = 0$, and $\sup_n \|A_n\|_\infty + \sup_n \|B_n\|_\infty < \infty$, then $\lim_{r\to\infty} \sup_{i,n} \sum_{j:\, d(i,j)>r} |c_{n,ij}| = 0$.*

Proof. As $c_{n,ij} = \sum_{k=1}^{n} a_{n,ik} b_{n,kj}$,

$$
\begin{aligned}
\sup_{i,n} \sum_{j:\, d(i,j)>r} |c_{n,ij}| \leq{}& \sup_{i,n} \sum_{j:\, d(i,j)>r} \sum_{k:\, d(j,k)>r/2} |a_{n,ik} b_{n,kj}| \\
&+ \sup_{i,n} \sum_{j:\, d(i,j)>r} \sum_{k:\, d(j,k)\leq r/2} |a_{n,ik} b_{n,kj}| \\
\leq{}& \sup_{i,n} \sum_{k=1}^{n} |a_{n,ik}| \sum_{j:\, d(j,k)>r/2} |b_{n,kj}| + \sup_{i,n} \sum_{k:\, d(i,k)>r/2} |a_{n,ik}| \sum_{j=1}^{n} |b_{n,kj}| \\
\leq{}& \sup_{n} \|A_n\|_\infty \sup_{k,n} \sum_{j:\, d(j,k)>r/2} |b_{n,kj}| + \sup_{i,n} \sum_{k:\, d(i,k)>r/2} |a_{n,ik}| \cdot \sup_{n} \|B_n\|_\infty,
\end{aligned}
$$

where the second inequality holds because $d(i,j) > r$ and $d(j,k) \leq r/2$ imply that $d(i,k) > r/2$. Thus, $\lim_{r\to\infty} \sup_{i,n} \sum_{j:\, d(i,j)>r} |c_{n,ij}| = 0$. □

For any matrix $A = [a_{ij}]$, denote $\mathrm{abs}(A) = [|a_{ij}|]$.

LEMMA B.6. (*i*) *If Assumptions 1 and 2(iii) hold, then* $\sup_n \|S_n^{-1}\|_\infty < \infty$ *and* $\sup_n \|R_n^{-1}\|_\infty < \infty$. (*ii*) *If Assumptions 1, 2(ii), and 7 hold, then* $\lim_{r\to\infty} \sup_{i,n} \sum_{j:\, d(i,j)>r} |w_{n,ij}| = 0$, $\lim_{r\to\infty} \sup_{i,n} \sum_{j:\, d(i,j)>r} |m_{n,ij}| = 0$, $\sup_n \|W_n\|_1 < \infty$, *and* $\sup_n \|M_n\|_1 < \infty$. (*iii*) *If Assumptions 1, 2(ii) and (iii), and 7(ii) hold, then* $\sup_n \|S_n^{-1}\|_1 < \infty$ *and* $\sup_n \|R_n^{-1}\|_1 < \infty$.

Proof. (*i*) As $\|\lambda_0 W_n\|_\infty \leq c_0 < 1$, $S_n^{-1} = \sum_{k=0}^{\infty} (\lambda_0 W_n)^k$. Thus, by the triangle inequality, $\sup_n \|S_n^{-1}\|_\infty \leq \sup_n \sum_{k=0}^{\infty} (\|\lambda_0 W_n\|_\infty)^k \leq \sum_{k=0}^{\infty} c_0^k = \frac{1}{1-c_0} < \infty$. Similarly, $\sup_n \|R_n^{-1}\|_\infty < \infty$.

(*ii*) Under Assumption 7(i), $w_{n,ij} = 0$ if $d(i,j) > \bar{d}_0$. Then $\lim_{r\to\infty} \sup_{i,n} \sum_{j:\, d(i,j)>r} |w_{n,ij}| = 0$. By Lemma A.1 in Jenish and Prucha (2009), $|\{j\colon k \leq d(i,j) < k+1\}| \leq c k^{c_d - 1}$, for $k \geq 1$, and some constant $c > 0$, where $|A|$ for a set A denotes its cardinality. Then $\sup_n \|W_n\|_1 = \sup_{j,n} \sum_{i:\, d(i,j)\leq \bar{d}_0} |w_{n,ij}| \leq \sup_{j,n} \sum_{k=1}^{[\bar{d}_0]+1} \sum_{i:\, k\leq d(i,j)<k+1} c_w = c c_w \sum_{k=1}^{[\bar{d}_0]+1} k^{c_d-1} < \infty$, where $c_w = \sup_n \|W_n\|_\infty < \infty$ under Assumption 2(ii) and $[\bar{d}_0]$ is the smallest integer that is nongreater than $\bar{d}_0$.

Under Assumption 7(ii), $\sup_{i,n} \sum_{j:\, d(i,j)>r} |w_{n,ij}| \leq \sup_{i,n} \sum_{k=[r]}^{\infty} \sum_{j:\, k\leq d(i,j)<k+1} |w_{n,ij}| \leq \sup_{i,n} \sum_{k=[r]}^{\infty} \sum_{j:\, k\leq d(i,j)<k+1} \pi_1 k^{-\pi_2} \leq \sum_{k=[r]}^{\infty} c\pi_1 k^{c_d-\pi_2-1}$. As $\pi_2 > c_d$, $\sum_{k=1}^{\infty} c\pi_1 k^{c_d-\pi_2-1} < \infty$. Then $\lim_{r\to\infty} \sum_{k=[r]}^{\infty} c\pi_1 k^{c_d-\pi_2-1} = 0$. It follows that $\lim_{r\to\infty} \sup_{i,n} \sum_{j:d(i,j)>r} |w_{n,ij}| = 0$. Similarly, $\sup_n \|W_n\|_1 = \sup_{j,n} \sum_{i:\, d(i,j)\geq 1} |w_{n,ij}| \leq \sum_{k=1}^{\infty} c\pi_1 k^{c_d-\pi_2-1} < \infty$.

The results on M_n can be similarly proved.

(*iii*) Under the maintained assumptions, we have $\|\lambda_0^l [\mathrm{abs}(W_n)]^l\|_1 \leq \max\{lN, 1\}\omega c_0^{l-1}$, where $\omega = |\lambda_0| \sup_n \|W_n\|_1 < \infty$, as in the proof of Lemma 1 of Xu and Lee (2015). The only difference is that our upper bound $\max\{lN, 1\}\omega c_0^{l-1}$ has c_0^{l-1} instead of ζ^{l-1}, where ζ is the upper bound of the compact parameter space of λ. Since we have a linear SAR process, there is no need to introduce ζ and the proof is similar. Then

$\sup_n \|S_n^{-1}\|_1 \leq \sup_n \sum_{k=0}^{\infty}(\|\lambda_0 W_n\|_1)^k \leq c(1+\sum_{k=1}^{\infty} k c_0^{k-1}) < \infty$ for some constant c. Similarly, $\sup_n \|R_n^{-1}\|_1 < \infty$. □

LEMMA B.7. *Under Assumptions 1, 2(i)–(iii), and 7,* $\{e_{ni}' A_n V_n\}$ *is* L_2*-NED on* $\{v_1, \ldots, v_n\}$*, where* A_n *is either* $S_n^{-1}R_n^{-1}$, $W_n S_n^{-1}R_n^{-1}$, $M_n S_n^{-1}R_n^{-1}$, *or* $W_n M_n S_n^{-1} R_n^{-1}$.

Proof. As $\|\lambda_0 W_n\|_\infty \leq c_0 < 1$, $S_n^{-1} = \sum_{k=0}^{\infty}(\lambda_0 W_n)^k$. Then $\mathrm{abs}(S_n^{-1}) \leq^* \sum_{k=0}^{\infty} [\mathrm{abs}(\lambda_0 W_n)]^k \leq^* [I_n - \mathrm{abs}(\lambda_0 W_n)]^{-1}$, where $A_n \leq^* B_n$ for two $n\times n$ matrices $A_n = [a_{n,ij}]$ and $B_n = [b_{n,ij}]$ means that $a_{n,ij} \leq b_{n,ij}$ for any i,j. Since the proof of Proposition 1 in Xu and Lee (2015, p. 274) shows that $[I_n - \mathrm{abs}(\lambda_0 W_n)]^{-1}$ satisfies $\lim_{r\to\infty}\sup_{i,n}\sum_{j:d(i,j)>r}|e_{ni}'[I_n - \mathrm{abs}(\lambda_0 W_n)]^{-1} e_{nj}| = 0$ under Assumptions 1, 2(iii), and 7, we also have $\lim_{r\to\infty}\sup_{i,n}\sum_{j:d(i,j)>r}|e_{ni}' S_n^{-1} e_{nj}| = 0$. Similarly, $\lim_{r\to\infty}\sup_{i,n}\sum_{j:d(i,j)>r}|e_{ni}' R_n^{-1} e_{nj}| = 0$. By Lemma B.6, $\lim_{r\to\infty}\sup_{i,n}\sum_{j:d(i,j)>r} |w_{n,ij}| = 0$, $\lim_{r\to\infty}\sup_{i,n}\sum_{j:d(i,j)>r}|m_{n,ij}| = 0$, $\sup_n \|S_n^{-1}\|_\infty < \infty$, and $\sup_n \|R_n^{-1}\|_\infty < \infty$. Thus, by Lemma B.5, $\lim_{r\to\infty}\sup_{i,n}\sum_{j:d(i,j)>r}[\mathrm{abs}(A_n)]_{ij} = 0$, where A_n is either $S_n^{-1}R_n^{-1}$, $W_n S_n^{-1}R_n^{-1}$, $M_n S_n^{-1}R_n^{-1}$, or $W_n M_n S_n^{-1}R_n^{-1}$. Hence, by Proposition 1 in Jenish and Prucha (2012), $\{e_{ni}' A_n V_n\}$ is L_2-NED on $\{v_1, \ldots, v_n\}$. □

APPENDIX C. Proofs

For the following proofs of Propositions 1 and 2, denote $\Psi_{ni}(\theta) = \sigma_0 e_{ni}' T_n(\tau) V_n - \sigma_0 v_i t_{n,ii}(\tau) + e_{ni}' R_n(\rho)[S_n(\lambda) S_n^{-1} X_n \beta_0 - X_n \beta]$, which does not depend on v_i. As $Y_n = S_n^{-1}(X_n\beta_0 + \sigma_0 R_n^{-1} V_n)$, $v_i(\theta) = \frac{1}{\sigma}\Psi_{ni}(\theta) + \frac{1}{\sigma}\sigma_0 v_i t_{n,ii}(\tau)$.

Proof of Proposition 1. (*i*) We first prove the result under Assumption 4(i). As $T_n(\tau) = I_n + (\rho_0 - \rho)A_{1n} + (\lambda_0 - \lambda)A_{2n} + (\rho_0 - \rho)(\lambda_0 - \lambda)A_{3n}$, under Assumption 3(iii), $t_{n,ii}(\tau) \neq 0$ for any i and τ. Since M_n is row-normalized, $R_n 1_n = (1-\rho_0)1_n$. Then the nonsingularity of R_n implies that $\rho_0 \neq 1$. Denote $\mathcal{Q}(\sigma, \beta_1, \eta) = \mathrm{E}[\ln f(\frac{\sigma_0 v_i - (1-\rho_0)(\beta_1 - \beta_{10})}{\sigma}, \eta)] - \frac{1}{2}\ln(\sigma^2)$, $\sigma_{ni} = \frac{\sigma}{t_{n,ii}(\tau)}$, and $\beta_{1,ni} = \beta_{10} - \frac{1}{(1-\rho_0)t_{n,ii}(\tau)}\Psi_{ni}(\theta)$. Since $\mathrm{E}[\ln f(\frac{\sigma_0 v_i - \alpha}{\sigma}, \eta)] - \frac{1}{2}\ln(\sigma^2)$ is uniquely maximized at $(\sigma_\infty, \alpha_\infty, \eta_\infty)$, $\mathcal{Q}(\sigma, \beta_1, \eta)$ is uniquely maximized at $(\sigma_\infty, \beta_{1\infty}, \eta_\infty)$, where $\beta_{1\infty} = \beta_{10} + \frac{\alpha_\infty}{1-\rho_0}$. Let $\mathrm{E}_{-i}(\cdot)$ be the conditional expectation given $v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_n$. Then,

$$
\begin{aligned}
\mathrm{E}[\ln L_n(\gamma)] &= \sum_{i=1}^{n} \mathrm{E}\{\mathrm{E}_{-i}[\ln f(v_i(\theta), \eta)]\} - \frac{n}{2}\ln(\sigma^2) + \ln|S_n(\lambda)| + \ln|R_n(\rho)| \\
&= \sum_{i=1}^{n} \mathrm{E}[\mathcal{Q}(\sigma_{ni}, \beta_{1,ni}, \eta)] - \sum_{i=1}^{n} \ln|t_{n,ii}(\tau)| + \ln|S_n(\lambda)| + \ln|R_n(\rho)| \\
&\leq n\mathcal{Q}(\sigma_\infty, \beta_{1\infty}, \eta_\infty) - \sum_{i=1}^{n} \ln|t_{n,ii}(\tau)| + \ln|S_n(\lambda)| + \ln|R_n(\rho)| \qquad \textbf{(C.1)} \\
&= n\mathcal{Q}(\sigma_\infty, \beta_{1\infty}, \eta_\infty) - \sum_{i=1}^{n} \ln|t_{n,ii}(\tau)| + \ln|T_n(\tau)| + \ln|S_n| + \ln|R_n| \\
&\leq \mathrm{E}[\ln L_n(\gamma_*)], \qquad \textbf{(C.2)}
\end{aligned}
$$

where (C.1) uses the property that $\mathcal{Q}(\sigma,\beta_1,\eta)$ is uniquely maximized at $(\sigma_\infty,\beta_{1\infty},\eta_\infty)$ and (C.2) uses the assumption that $\ln|T_n(\tau)| \le \sum_{i=1}^n \ln|t_{n,ii}(\tau)|$. The inequality in (C.2) is strict if $\tau \ne \tau_0$. With $\tau = \tau_0$, we have $T_n(\tau) = I_n$, $t_{n,ii}(\tau) = 1$, $\sigma_{ni} = \sigma$, and $\beta_{1,ni} = \beta_{10} - \frac{1}{1-\rho_0} e'_{ni} R_n X_n(\beta_0 - \beta) = \beta_1 - \frac{1}{1-\rho_0} e'_{ni} R_n X_{2n}(\beta_{20} - \beta_2)$. Since $R_n X_n$ has full column rank, $\beta_{1,ni} \ne \beta_{1\infty}$ for some i if $\beta_2 \ne \beta_{20}$. Thus, with $\tau = \tau_0$, the inequality in (C.1) is strict if $(\beta_2,\sigma,\eta) \ne (\beta_{20},\sigma_\infty,\eta_\infty)$. It follows that $\mathrm{E}[\ln L_n(\gamma)]$ is uniquely maximized at $\gamma = \gamma_*$. (*ii*) We next prove the result under Assumption 4(ii). Because v_i's are symmetrically distributed around zero with unimodal density, by Lemma A in Newey and Steigerwald (1997), $\mathrm{E}[\ln f(v_i(\theta),\eta)] = \mathrm{E}\{\mathrm{E}_{-i}[\ln f(v_i(\theta),\eta)]\} \le \mathrm{E}\{\mathrm{E}_{-i}[\ln f(\frac{\sigma_0}{\sigma} v_i t_{n,ii}(\tau),\eta)]\} = \mathrm{E}[\ln f(\frac{\sigma_0}{\sigma} v_i t_{n,ii}(\tau),\eta)]$, where the inequality is strict if $\Psi_{ni}(\theta) \ne 0$. Denote $\mathcal{Q}(\sigma,\eta) = \mathrm{E}[\ln f(\frac{\sigma_0 v_i}{\sigma},\eta)] - \frac{1}{2}\ln(\sigma^2)$. Then,

$$
\begin{aligned}
\mathrm{E}[\ln L_n(\gamma)] &\le \sum_{i=1}^n \mathcal{Q}(\sigma_{ni},\eta) - \sum_{i=1}^n \ln|t_{n,ii}(\tau)| + \ln|S_n(\lambda)| + \ln|R_n(\rho)| && \textbf{(C.3)}\\
&\le n\mathcal{Q}(\sigma_\infty,\eta_\infty) - \sum_{i=1}^n \ln|t_{n,ii}(\tau)| + \ln|S_n(\lambda)| + \ln|R_n(\rho)| && \textbf{(C.4)}\\
&\le \mathrm{E}[\ln L_n(\gamma_\#)], && \textbf{(C.5)}
\end{aligned}
$$

where (C.4) uses the assumption that $\mathcal{Q}(\sigma,\eta)$ is uniquely maximized at $(\sigma_\infty,\eta_\infty)$, and (C.5) uses the assumption that $\ln|T_n(\tau)| \le \sum_{i=1}^n \ln|t_{n,ii}(\tau)|$ as in the proof for (*i*) above. Furthermore, the inequality in (C.5) is strict if $\tau \ne \tau_0$. With $\tau = \tau_0$, the inequality in (C.4) is strict if $(\sigma,\eta) \ne (\sigma_\infty,\eta_\infty)$. With $(\tau,\sigma,\eta) = (\tau_0,\sigma_\infty,\eta_\infty)$, we have $T_n(\tau) = I_n$ and $\Psi_{ni}(\theta) = e'_{ni} R_n X_n(\beta_0 - \beta)$. Since $R_n X_n$ has full column rank, with $(\tau,\sigma,\eta) = (\tau_0,\sigma_\infty,\eta_\infty)$, the inequality in (C.3) is strict if $\beta \ne \beta_0$. Hence, $\mathrm{E}[\ln L_n(\gamma)]$ is uniquely maximized at $\gamma_\#$. □

Proof of Proposition 2. Denote $\mathcal{Q}(\sigma,\alpha,\eta) = \mathrm{E}[\ln f(\frac{\sigma_0 v_i - \alpha}{\sigma},\eta)] - \frac{1}{2}\ln(\sigma^2)$, $\sigma_{ni} = \frac{\sigma}{t_{n,ii}(\tau)}$, and $\alpha_{ni} = \frac{\alpha - \Psi_{ni}(\theta)}{t_{n,ii}(\tau)}$. Then,

$$
\begin{aligned}
\mathrm{E}[\ln L_n(\delta)] &= \sum_{i=1}^n \mathrm{E}\big\{\mathrm{E}_{-i}\big[\ln f\big(v_i(\theta) - \frac{\alpha}{\sigma},\eta\big)\big]\big\} - \frac{n}{2}\ln(\sigma^2) + \ln|S_n(\lambda)| + \ln|R_n(\rho)|\\
&= \sum_{i=1}^n \mathrm{E}[\mathcal{Q}(\sigma_{ni},\alpha_{ni},\eta)] - \sum_{i=1}^n \ln|t_{n,ii}(\tau)| + \ln|S_n(\lambda)| + \ln|R_n(\rho)|\\
&\le n\mathcal{Q}(\sigma_\infty,\alpha_\infty,\eta_\infty) - \sum_{i=1}^n \ln|t_{n,ii}(\tau)| + \ln|S_n(\lambda)| + \ln|R_n(\rho)| && \textbf{(C.6)}\\
&= n\mathcal{Q}(\sigma_\infty,\alpha_\infty,\eta_\infty) - \sum_{i=1}^n \ln|t_{n,ii}(\tau)| + \ln|T_n(\tau)| + \ln|S_n| + \ln|R_n|\\
&\le \mathrm{E}[\ln L_n(\delta_\#)], && \textbf{(C.7)}
\end{aligned}
$$

where (C.6) uses the property that $Q_n(\sigma,\alpha,\eta)$ is uniquely maximized at $(\sigma_\infty,\alpha_\infty,\eta_\infty)$ and (C.7) uses the assumption that $\ln|T_n(\tau)| \le \sum_{i=1}^n \ln|t_{n,ii}(\tau)|$. The inequality in (C.7) is strict if $\tau \ne \tau_0$. With $\tau = \tau_0$, we have $T_n(\tau) = I_n$, $t_{n,ii}(\tau) = 1$, $\sigma_{ni} = \sigma$, and $\alpha_{ni} = \alpha - e'_{ni} R_n X_n(\beta_0 - \beta)$. Since $R_n X_n$ has full column rank and does not contain an intercept term, $\alpha_{ni} \ne \alpha_\infty$ for some i if $\beta \ne \beta_0$. Thus, with $\tau = \tau_0$, the inequality in (C.6) is strict if $(\beta,\sigma,\alpha,\eta) \neq (\beta_0,\sigma_\infty,\alpha_\infty,\eta_\infty)$. It follows that $\mathrm{E}[\ln L_n(\delta)]$ is uniquely maximized at $\delta = \delta_\#$. □

Proof of Theorem 1. We only prove the convergence of $\hat{\gamma}$ in the case with symmetric v_i, since the proofs for other cases are similar. As $Y_n = S_n^{-1}(X_n\beta_0 + \sigma_0 R_n^{-1}V_n)$, $R_n(\rho) = R_n + (\rho_0-\rho)M_n$, and $S_n(\lambda) = S_n + (\lambda_0-\lambda)W_n$, we have

$$
\begin{aligned}
&R_n(\rho)[S_n(\lambda)Y_n - X_n\beta] \\
&= \sigma_0 V_n + (\lambda_0-\lambda)R_n W_n S_n^{-1}X_n\beta_0 + \sigma_0(\lambda_0-\lambda)R_n W_n S_n^{-1}R_n^{-1}V_n + R_n X_n(\beta_0-\beta) \\
&\quad + \sigma_0(\rho_0-\rho)M_n R_n^{-1}V_n + (\rho_0-\rho)(\lambda_0-\lambda)M_n W_n S_n^{-1}X_n\beta_0 \\
&\quad + \sigma_0(\rho_0-\rho)(\lambda_0-\lambda)M_n W_n S_n^{-1}R_n^{-1}V_n + (\rho_0-\rho)M_n X_n(\beta_0-\beta).
\end{aligned}
$$

Under Assumption 2(iii), by Lemma B.6, R_n^{-1} and S_n^{-1} are bounded in the row-sum norm. As W_n and M_n are also bounded in the row-sum norm, so are the products of W_n, M_n, R_n^{-1}, and S_n^{-1}. With $\sup_i \mathrm{E}(|v_i|^{2+2c_t+\iota}) < \infty$ in Assumption 8(ii), $v_i(\theta) = \frac{1}{\sigma}e'_{ni}R_n(\rho)[S_n(\lambda)Y_n - X_n\beta]$ is uniformly $L_{(2+2c_t+\iota)}$ bounded by Lemma B.2. Furthermore, by Lemma B.7, $\{v_i(\theta)\}$ is L_2-NED on $\{v_1,\ldots,v_n\}$. With $|\frac{\partial \ln f(x,\eta)}{\partial x}| \le c_f(|x|^{c_t}+1)$ for $c_t = 0$ in Assumption 8(i), i.e., $\frac{\partial \ln f(x,\eta)}{\partial x}$ is bounded, by Proposition 2 of Jenish and Prucha (2012), $\ln f(v_i(\theta),\eta)$ is L_2-NED on $\{v_1,\ldots,v_n\}$; on the other hand, with $|\frac{\partial \ln f(x,\eta)}{\partial x}| \le c_f(|x|^{c_t}+1)$ for $c_t = 1$ in Assumption 8(i), by Lemma A.4 in Xu and Lee (2015), $\ln f(v_i(\theta),\eta)$ is uniformly L_2-NED on $\{v_1,\ldots,v_n\}$. By the mean value theorem, $\ln f(v_i(\theta),\eta) = \ln f(0,\eta) + \frac{\partial \ln f(cv_i(\theta),\eta)}{\partial v}v_i(\theta)$, where c is some constant between 0 and 1. Thus, with $\sup_i \mathrm{E}(|v_i|^{2+2c_t+\iota}) < \infty$, $\ln f(v_i(\theta),\eta)$ is uniformly L_2 bounded by Lemma B.2. It follows by the LLN in Theorem 1 of Jenish and Prucha (2012) that $\frac{1}{n}\ln L_n(\gamma) - \frac{1}{n}\mathrm{E}[\ln L_n(\gamma)] = o_p(1)$.

We next prove that $\frac{1}{n}\ln L_n(\gamma)$ is stochastically equicontinuous (SE) and $\frac{1}{n}\mathrm{E}[L_n(\gamma)]$ is equicontinuous. With $|\frac{\partial \ln f(x,\eta)}{\partial x}| \le c_f(|x|^{c_t}+1)$,

$$
\frac{1}{n}\mathrm{E}\Big|\frac{\partial \ln L_n(\gamma)}{\partial \lambda}\Big| \le \frac{c_f}{n\sigma}\mathrm{E}\sum_{i=1}^{n}[|v_i(\theta)|^{c_t}+1]\cdot|e'_{ni}R_n(\rho)W_nY_n| + \frac{1}{n}|\mathrm{tr}[W_nS_n^{-1}(\lambda)]|, \tag{C.8}
$$

where $\frac{c_f}{n\sigma}\mathrm{E}\sum_{i=1}^{n}[|v_i(\theta)|^{c_t}+1]\cdot|e'_{ni}R_n(\rho)W_nY_n| = O(1)$ by $Y_n = S_n^{-1}(X_n\beta_0 + \sigma_0R_n^{-1}V_n)$ and Lemma B.2, and $\frac{1}{n}|\mathrm{tr}[W_nS_n^{-1}(\lambda)]| = O(1)$ since $\sup_n \|W_n\|_\infty < \infty$ by Assumption 2(ii), $\sup_n \|W_n\|_1 < \infty$ by Lemma B.6 and $S_n^{-1}(\lambda)$ is bounded in either the row- or column-sum norm. Thus, $\frac{1}{n}\mathrm{E}|\frac{\partial \ln L_n(\gamma)}{\partial \lambda}| = O(1)$ and $\frac{1}{n}\frac{\partial \ln L_n(\gamma)}{\partial \lambda} = O_p(1)$. As $\sigma v_i(\theta)$ is linear in every element of θ and the parameter space of γ is compact, by (C.8), $\mathrm{E}\sup_{\gamma\in\Gamma}|\frac{1}{n}\frac{\partial \ln L_n(\gamma)}{\partial \lambda}| = O(1)$ and $\sup_{\gamma\in\Gamma}|\frac{1}{n}\frac{\partial \ln L_n(\gamma)}{\partial \lambda}| = O_p(1)$. Similarly, for other elements γ_j of γ, $\mathrm{E}\sup_{\gamma\in\Gamma}|\frac{1}{n}\frac{\partial \ln L_n(\gamma)}{\partial \gamma_j}| = O(1)$ and $\sup_{\gamma\in\Gamma}|\frac{1}{n}\frac{\partial \ln L_n(\gamma)}{\partial \gamma_j}| = O_p(1)$. Hence, $\mathrm{E}\sup_{\gamma\in\Gamma}\|\frac{1}{n}\frac{\partial \ln L_n(\gamma)}{\partial \gamma}\| = O(1)$ and $\sup_{\gamma\in\Gamma}\|\frac{1}{n}\frac{\partial \ln L_n(\gamma)}{\partial \gamma}\| = O_p(1)$. By Lemma 3.6 in Newey and McFadden (1994), $\mathrm{E}\sup_{\gamma\in\Gamma}\|\frac{1}{n}\frac{\partial \ln L_n(\gamma)}{\partial \gamma}\| = O(1)$ implies that $\frac{1}{n}\frac{\partial \mathrm{E}[\ln L_n(\gamma)]}{\partial \gamma} = \frac{1}{n}\mathrm{E}(\frac{\partial \ln L_n(\gamma)}{\partial \gamma})$. Therefore, by the mean value theorem and Theorem 21.10 in Davidson (1994), $\frac{1}{n}\ln L_n(\gamma)$ is SE, and $\frac{1}{n}\mathrm{E}[L_n(\gamma)]$ is equicontinuous.

The pointwise convergence $\frac{1}{n}\ln L_n(\gamma) - \frac{1}{n}\mathrm{E}[\ln L_n(\gamma)] = o_p(1)$ and the SE of $\frac{1}{n}\ln L_n(\gamma)$ imply that $\sup_{\gamma\in\Gamma}|\frac{1}{n}\ln L_n(\gamma) - \frac{1}{n}\mathrm{E}[\ln L_n(\gamma)]| = o_p(1)$. As $\frac{1}{n}\mathrm{E}[\ln L_n(\gamma)]$ is equicontinuous and $\lim_{n\to\infty}\frac{1}{n}\mathrm{E}[\ln L_n(\gamma)]$ is uniquely maximized at $\gamma=\gamma_\#$, we have $\hat\gamma=\gamma_\#+o_p(1)$ (White, 1994, Theorem 3.4). □

Proof of Theorem 2. We only prove the asymptotic distribution of $\hat\gamma$ in the case with symmetric v_i, and omit similar proofs for other cases. By the mean value theorem, $0=\frac{\partial\ln L_n(\hat\gamma)}{\partial\gamma}=\frac{\partial\ln L_n(\gamma_\#)}{\partial\gamma}+\frac{\partial^2\ln L_n(\tilde\gamma)}{\partial\gamma\partial\gamma'}(\hat\gamma-\gamma_\#)$, where $\tilde\gamma$ lies between $\hat\gamma$ and $\gamma_\#$. Then,

$$\sqrt{n}(\hat\gamma-\gamma_\#)=-\Big(\frac{1}{n}\frac{\partial^2\ln L_n(\tilde\gamma)}{\partial\gamma\partial\gamma'}\Big)^{-1}\frac{1}{\sqrt{n}}\frac{\partial\ln L_n(\gamma_\#)}{\partial\gamma}. \tag{C.9}$$

We prove that (i) $\frac{1}{n}\frac{\partial^2\ln L_n(\tilde\gamma)}{\partial\gamma\partial\gamma'}=\frac{1}{n}\frac{\partial^2\ln L_n(\gamma_\#)}{\partial\gamma\partial\gamma'}+o_p(1)$ and (ii) $\frac{1}{n}\frac{\partial^2\ln L_n(\gamma_\#)}{\partial\gamma\partial\gamma'}=\frac{1}{n}\mathrm{E}(\frac{\partial^2\ln L_n(\gamma_\#)}{\partial\gamma\partial\gamma'})+o_p(1)$ so that $\frac{1}{n}\frac{\partial^2\ln L_n(\tilde\gamma)}{\partial\gamma\partial\gamma'}=\frac{1}{n}\mathrm{E}(\frac{\partial^2\ln L_n(\gamma_\#)}{\partial\gamma\partial\gamma'})+o_p(1)$.

For (i), we prove that every element of $\frac{1}{n}\frac{\partial^2\ln L_n(\gamma)}{\partial\gamma\partial\gamma'}$ is SE under Assumption 9(ii) and (iii). With $\|\frac{\partial^3\ln f(v_i(\theta),\eta)}{\partial z\partial z'\partial z_i}\|\le c_f(|v_i(\theta)|^{3c_t}+1)$ in Assumption 9(ii), we could show that $\sup_{\gamma\in\Gamma}\|\frac{\partial^3\ln L_n(\gamma)}{\partial\gamma\partial\gamma'\partial\gamma_j}\|=O_p(1)$, where γ_j is the jth element of γ. As an example, consider

$$\frac{\partial^3\ln L_n(\gamma)}{\partial\lambda^3}=-\frac{1}{\sigma^3}\sum_{i=1}^n\frac{\partial^3\ln f(v_i(\theta),\eta)}{\partial v^3}[e_{ni}'R_n(\rho)W_nY_n]^3-2\,\mathrm{tr}\{[W_nS_n^{-1}(\lambda)]^3\},$$

where $|\frac{\partial^3\ln f(v_i(\theta),\eta)}{\partial v^3}|\le c[|v_i(\theta)|^{3c_t}+1]$. With the reduced form $Y_n=S_n^{-1}(X_n\beta_0+\sigma_0R_n^{-1}V_n)$ and $\mathrm{E}(|v_i|^{3+3c_t})<\infty$, $\frac{1}{n}\frac{\partial^3\ln L_n(\gamma)}{\partial\lambda^3}=O_p(1)$ by Lemma B.2. As $v_i(\theta)=\frac{1}{\sigma}e_{ni}'R_n(\rho)[S_n(\lambda)Y_n-X_n\beta]$ is linear in each element of $[\lambda,\rho,\beta']'$, $\{S_n^{-1}(\lambda)\}$ is bounded in either the row-sum or column-sum norm uniformly on the parameter space of λ and Γ is compact, $\sup_{\gamma\in\Gamma}|\frac{1}{n}\frac{\partial^3\ln L_n(\gamma)}{\partial\lambda^3}|=O_p(1)$. Hence, (i) holds by the mean value theorem.

We prove (ii) by Lemma B.3. As an example, consider

$$\frac{1}{n}\frac{\partial^2\ln L_n(\gamma_\#)}{\partial\lambda^2}=\frac{1}{n\sigma_\infty^2}\sum_{i=1}^n\frac{\partial^2\ln f(\frac{\sigma_0}{\sigma_\infty}v_i,\eta_\infty)}{\partial v^2}(e_{ni}'R_nW_nY_n)^2-\frac{1}{n}\mathrm{tr}[(W_nS_n^{-1})^2].$$

Under Assumption 9(ii), $\frac{\partial^2\ln f(\frac{\sigma_0}{\sigma_\infty}v_i,\eta_\infty)}{\partial v^2}$ is either bounded or $|\frac{\partial^2\ln f(\frac{\sigma_0}{\sigma_\infty}v_i,\eta_\infty)}{\partial v^2}|\le c_f(\frac{\sigma_0^2}{\sigma_\infty^2}|v_i|^2+1)$. In the latter case, as $\sup_i\mathrm{E}(|v_i|^{4+\iota})<\infty$, $\mathrm{E}[|\frac{\partial^2\ln f(\frac{\sigma_0}{\sigma_\infty}v_i,\eta_\infty)}{\partial v^2}|^{2+\iota/2}]<\infty$. Then, using $Y_n=S_n^{-1}(X_n\beta_0+\sigma_0R_n^{-1}V_n)$ and $\sup_i\mathrm{E}(|v_i|^{2+2c_t+\iota})<\infty$, where $c_t=0$ for the case with bounded $\frac{\partial^2\ln f(\frac{\sigma_0}{\sigma_\infty}v_i,\eta_\infty)}{\partial v^2}$ and $c_t=1$ for the case with $|\frac{\partial^2\ln f(\frac{\sigma_0}{\sigma_\infty}v_i,\eta_\infty)}{\partial v^2}|\le c_f(\frac{\sigma_0^2}{\sigma_\infty^2}|v_i|^2+1)$, we have $\frac{1}{n}\frac{\partial^2\ln L_n(\gamma_\#)}{\partial\lambda^2}-\mathrm{E}(\frac{1}{n}\frac{\partial^2\ln L_n(\gamma_\#)}{\partial\lambda})=o_p(1)$ by Lemma B.3.

With (i) and (ii), by (C.9), $\sqrt{n}(\hat\gamma-\gamma_\#)=-\big(\frac{1}{n}\mathrm{E}\frac{\partial^2\ln L_n(\gamma_\#)}{\partial\gamma\partial\gamma'}\big)^{-1}\frac{1}{\sqrt{n}}\frac{\partial\ln L_n(\gamma_\#)}{\partial\gamma}+o_p(1)$. Under Assumption 4(ii), $\mathrm{E}[\ln f(\frac{\sigma_0}{\sigma}v_i+c,\eta)]$ is uniquely maximized at $c=0$ for any σ and η, by Lemma A in Newey and Steigerwald (1997). Then $\mathrm{E}(\zeta_{1i})=0$, where $\zeta_{1i}=\frac{\sigma_0}{\sigma_\infty}\frac{\partial\ln f(\frac{\sigma_0}{\sigma_\infty}v_i,\eta_\infty)}{\partial v}$. By Assumption 4(ii)(c), $\mathrm{E}(\zeta_{2i})=0$ and $\mathrm{E}(\zeta_{3i})=0$, where $\zeta_{2i}=\zeta_{1i}v_i+1$ and $\zeta_{3i}=-\frac{\partial\ln f(\frac{\sigma_0}{\sigma_\infty}v_i,\eta_\infty)}{\partial\eta}$. Hence, every element of $\frac{\partial\ln L_n(\gamma_\#)}{\partial\gamma}$ is a special case of the general linear-quadratic form ω_n in Lemma B.4. By Assumptions 2(ii) and 9(iv) and Lemma B.6, the involved matrices $S_n^{-1}R_n^{-1}$, $W_nS_n^{-1}R_n^{-1}$, $M_nS_n^{-1}R_n^{-1}$, and $W_nM_nS_n^{-1}R_n^{-1}$ in ω_n are bounded in both the row- and column-sum norms. As $|\frac{\partial \ln f(\frac{\sigma_0}{\sigma_\infty}v_i,\eta_\infty)}{\partial v}| \leq c_f(|\frac{\sigma_0}{\sigma_\infty}v_i|^{c_t}+1)$ and $\sup_i \mathrm{E}(|v_i|^{2+2c_t+\iota}) < \infty$, we have $\mathrm{E}[|\frac{\partial \ln f(\frac{\sigma_0}{\sigma_\infty}v_i,\eta_\infty)}{\partial v}v_i|^{2+\iota/(1+c_t)}] < \infty$ and $\mathrm{E}[|\frac{\partial \ln f(\frac{\sigma_0}{\sigma_\infty}v_i,\eta_\infty)}{\partial v}|^{2+2c_t+\iota}] < \infty$ for $c_t = 0$ or 1. As $\|\frac{\partial \ln f(\frac{\sigma_0}{\sigma_\infty}v_i,\eta_\infty)}{\partial \eta}\| \leq c_f(|\frac{\sigma_0}{\sigma_\infty}v_i|^{1+c_t}+1)$ and $\sup_i \mathrm{E}(|v_i|^{2+2c_t+\iota}) < \infty$, $\mathrm{E}[\|\frac{\partial \ln f(\frac{\sigma_0}{\sigma_\infty}v_i,\eta_\infty)}{\partial \eta}\|^{2+\iota/(1+c_t)}] < \infty$. Then Lemma B.4 implies that $\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\gamma_\#)}{\partial \gamma} \xrightarrow{d} N(0, \lim_{n\to\infty}\mathcal{B})$, where $\mathcal{B} = \frac{1}{n}\mathrm{E}(\frac{\partial \ln L_n(\gamma_\#)}{\partial \gamma}\frac{\partial \ln L_n(\gamma_\#)}{\partial \gamma'})$. Hence, $\sqrt{n}(\hat{\gamma}-\gamma_\#) \xrightarrow{d} N(0, \lim_{n\to\infty}\mathcal{A}^{-1}\mathcal{B}\mathcal{A}^{-1})$, where $\mathcal{A} = -\frac{1}{n}\mathrm{E}(\frac{\partial^2 \ln L_n(\gamma_\#)}{\partial\gamma\partial\gamma'})$. □

Proof of Corollary 1. We first prove that: (*i*) $\mathrm{E}(\zeta_{1i}\zeta_{2i}) = 0$, (*ii*) $\mathrm{E}(\zeta_{1i}\zeta_{3i}) = 0$, (*iii*) $\mathrm{E}(\zeta_{4i}v_i) = 0$; and (*iv*) $\mathrm{E}(\zeta_{5i}) = 0$, where ζ_{1i} to ζ_{5i} are defined in Appendix A.2 and they satisfy $\mathrm{E}(\zeta_{1i}) = 0$, $\mathrm{E}(\zeta_{2i}) = 0$, and $\mathrm{E}(\zeta_{3i}) = 0$, as shown in the proof of Theorem 2.

(*i*) Note that for any even function $h_1(v)$ of v, $h_1(v) = h_1(|v|) = h_2(v^2)$, where $h_2(z) \equiv h_1(z^{1/2})$, for $z \geq 0$. Then a symmetrically distributed v_i is also spherically symmetric (Fang, Kotz, and Ng, 1990, p. 35). Define $g(\varsigma,\eta) = f(\varsigma^{1/2},\eta)$, for $\varsigma \geq 0$, so that $f(v,\eta) = f(|v|,\eta) = g(v^2,\eta)$. Then $\frac{\partial \ln f(v,\eta)}{\partial v} = 2\frac{\partial \ln g(v^2,\eta)}{\partial \varsigma}v$ and $\mathrm{E}(\frac{\partial \ln f(\frac{\sigma_0}{\sigma_\infty}v_i,\eta_\infty)}{\partial v}) = \frac{2\sigma_0}{\sigma_\infty}\mathrm{E}(\frac{\partial \ln g(\frac{\sigma_0^2}{\sigma_\infty^2}v_i^2,\eta_\infty)}{\partial \varsigma}v_i)$. Let $v_i = |v_i|\cdot\varpi_i$. It follows that $|v_i|$ and ϖ_i are independent (Fang et al., 1990, p. 30). Then $\mathrm{E}(\zeta_{1i}\zeta_{2i}) = \mathrm{E}[\zeta_{1i}(\zeta_{1i}v_i+1)] = \mathrm{E}(\zeta_{1i}^2v_i) = \frac{4\sigma_0^4}{\sigma_\infty^4}\mathrm{E}[(\frac{\partial \ln g(\frac{\sigma_0^2}{\sigma_\infty^2}v_i^2,\eta_\infty)}{\partial \varsigma})^2v_i^3] = \frac{4\sigma_0^4}{\sigma_\infty^4}\mathrm{E}[(\frac{\partial \ln g(\frac{\sigma_0^2}{\sigma_\infty^2}v_i^2,\eta_\infty)}{\partial \varsigma})^2|v_i|^3\cdot\varpi_i^3] = \frac{4\sigma_0^4}{\sigma_\infty^4}\mathrm{E}[(\frac{\partial \ln g(\frac{\sigma_0^2}{\sigma_\infty^2}v_i^2,\eta_\infty)}{\partial \varsigma})^2|v_i|^3]\mathrm{E}(\varpi_i^3)$. Since $0 = \mathrm{E}(v_i^3) = \mathrm{E}(|v_i|^3\cdot\varpi_i^3) = \mathrm{E}(|v_i|^3)\mathrm{E}(\varpi_i^3)$, $\mathrm{E}(\varpi_i^3) = 0$. Thus, $\mathrm{E}(\zeta_{1i}\zeta_{2i}) = 0$.

(*ii*) $\mathrm{E}(\zeta_{1i}\zeta_{3i}) = \mathrm{E}(\frac{2\sigma_0}{\sigma_\infty}\frac{\partial \ln g(\frac{\sigma_0^2}{\sigma_\infty^2}v_i^2,\eta_\infty)}{\partial \varsigma}v_i \cdot \frac{\partial \ln g(\frac{\sigma_0^2}{\sigma_\infty^2}v_i^2,\eta_\infty)}{\partial \eta}) = \frac{2\sigma_0}{\sigma_\infty}\mathrm{E}(\frac{\partial \ln g(\frac{\sigma_0^2}{\sigma_\infty^2}v_i^2,\eta_\infty)}{\partial \varsigma}\frac{\partial \ln g(\frac{\sigma_0^2}{\sigma_\infty^2}v_i^2,\eta_\infty)}{\partial \eta}|v_i|\cdot\varpi_i) = \frac{2\sigma_0}{\sigma_\infty}\mathrm{E}(\frac{\partial \ln g(\frac{\sigma_0^2}{\sigma_\infty^2}v_i^2,\eta_\infty)}{\partial \varsigma}\frac{\partial \ln g(\frac{\sigma_0^2}{\sigma_\infty^2}v_i^2,\eta_\infty)}{\partial \eta}|v_i|)\mathrm{E}(\varpi_i) = 0$, where we use $\mathrm{E}(\varpi_i) = 0$ implied by $0 = \mathrm{E}(v_i) = \mathrm{E}(|v_i|)\mathrm{E}(\varpi_i)$.

(*iii*) As $\frac{\partial \ln f(v,\eta)}{\partial v} = 2\frac{\partial \ln g(v^2,\eta)}{\partial \varsigma}v$, $\frac{\partial^2 \ln f(v,\eta)}{\partial v^2} = 4\frac{\partial^2 \ln g(v^2,\eta)}{\partial \varsigma^2}v^2 + 2\frac{\partial \ln g(v^2,\eta)}{\partial \varsigma}$. Then $\mathrm{E}(\zeta_{4i}v_i) = -\frac{4\sigma_0^4}{\sigma_\infty^4}\mathrm{E}(\frac{\partial^2 \ln g(\frac{\sigma_0^2}{\sigma_\infty^2}v_i^2,\eta_\infty)}{\partial \varsigma^2}v_i^3) - \frac{2\sigma_0^2}{\sigma_\infty^2}\mathrm{E}(\frac{\partial \ln g(\frac{\sigma_0^2}{\sigma_\infty^2}v_i^2,\eta_\infty)}{\partial \varsigma}v_i) = 0$.

(*iv*) As $\frac{\partial \ln f(v,\eta)}{\partial v} = 2\frac{\partial \ln g(v^2,\eta)}{\partial \varsigma}v$, $\frac{\partial^2 \ln f(v,\eta)}{\partial v\partial\eta} = 2\frac{\partial^2 \ln g(v^2,\eta)}{\partial \varsigma\partial\eta}v$. Then $\mathrm{E}(\zeta_{5i}) = \frac{2\sigma_0}{\sigma_\infty}\mathrm{E}(\frac{\partial^2 \ln g(\frac{\sigma_0^2}{\sigma_\infty^2}v_i^2,\eta_\infty)}{\partial \varsigma\partial\eta}v_i) = 0$.

By (*i*)–(*iv*) and Appendix A, we have $\mathcal{A}_{\beta\rho} = 0$, $\mathcal{A}_{\beta\sigma^2} = 0$, $\mathcal{A}_{\beta\eta} = 0$, $\mathcal{B}_{\beta\rho} = 0$, $\mathcal{B}_{\beta\sigma^2} = 0$, and $\mathcal{B}_{\beta\eta} = 0$. Hence, for the spatial error model, by Theorem 1, the asymptotic variance of the NGPMLE $\hat{\beta}$ is $\lim_{n\to\infty}\mathcal{A}_{\beta\beta}^{-1}\mathcal{B}_{\beta\beta}\mathcal{A}_{\beta\beta}^{-1} = \lim_{n\to\infty}[\frac{1}{n\sigma_0^2}\mathrm{E}(\zeta_{4i})X_n'R_n'R_nX_n]^{-1}\cdot\frac{1}{n\sigma_0^2}\mathrm{E}(\zeta_{1i}^2)X_n'R_n'R_nX_n\cdot[\frac{1}{n\sigma_0^2}\mathrm{E}(\zeta_{4i})X_n'R_n'R_nX_n]^{-1} = \lim_{n\to\infty}\frac{\sigma_0^2\mathrm{E}(\zeta_{1i}^2)}{[\mathrm{E}(\zeta_{4i})]^2}(\frac{1}{n}X_n'R_n'R_nX_n)^{-1}$.

The GPMLE is a special case of the NGPMLE with $f(v,\eta) = \frac{1}{\sqrt{2\pi}} e^{-v^2/2}$ and $\sigma_\infty^2 = \sigma_0^2$. Then, for the GPMLE, $\zeta_{1i} = -v_i$, $\zeta_{4i} = 1$, and the asymptotic variance for β is $\lim_{n\to\infty} \sigma_0^2 (\frac{1}{n} X_n' R_n' R_n X_n)^{-1}$. The BGMME of β has the same asymptotic variance as the GPMLE, by Corollary 3 in Liu et al. (2010). $\square$

Proof of Theorem 3. We could show that $\check{\gamma} = \gamma_\infty + o_p(1)$ as the proof of Theorem 2. By the mean value theorem, $0 = \frac{\partial \ln L_n(\check{\gamma})}{\partial \gamma_u} = \frac{\partial \ln L_n(\gamma_n)}{\partial \gamma_u} - \frac{\partial^2 \ln L_n(\bar{\gamma})}{\partial \gamma_u \partial \tau'} \tau_n + \frac{\partial^2 \ln L_n(\bar{\gamma})}{\partial \gamma_u \partial \gamma_u'} (\check{\gamma}_u - \gamma_{u\infty})$, where $\gamma_n = [\tau_n', \gamma_{u\infty}']'$ and $\bar{\gamma}$ lies between $\check{\gamma}$ and γ_∞. Thus, $\sqrt{n}(\check{\gamma}_u - \gamma_{u\infty}) = -(\frac{1}{n} \frac{\partial^2 \ln L_n(\bar{\gamma})}{\partial \gamma_u \partial \gamma_u'})^{-1} (\frac{1}{\sqrt{n}} \frac{\partial \ln L_n(\gamma_n)}{\partial \gamma_u} - \frac{1}{n} \frac{\partial^2 \ln L_n(\bar{\gamma})}{\partial \gamma_u \partial \tau'} \cdot \sqrt{n}\tau_n)$. As in the proof of Theorem 2, we could show that $\frac{1}{n} \frac{\partial^2 \ln L_n(\bar{\gamma})}{\partial \gamma_u \partial \gamma_u'} = \frac{1}{n} \mathrm{E}(\frac{\partial^2 \ln L_n(\gamma_n)}{\partial \gamma_u \partial \gamma_u'}) + o_p(1)$ and $\frac{1}{n} \frac{\partial^2 \ln L_n(\bar{\gamma})}{\partial \gamma_u \partial \tau'} = \frac{1}{n} \mathrm{E}(\frac{\partial^2 \ln L_n(\gamma_n)}{\partial \gamma_u \partial \tau'}) + o_p(1)$. Hence,

$$\sqrt{n}(\check{\gamma}_u - \gamma_{u\infty}) = -\Big(\frac{1}{n} \mathrm{E} \frac{\partial^2 \ln L_n(\gamma_n)}{\partial \gamma_u \partial \gamma_u'}\Big)^{-1} \times \Big[\frac{1}{\sqrt{n}} \frac{\partial \ln L_n(\gamma_n)}{\partial \gamma_u} - \frac{1}{n} \mathrm{E}\Big(\frac{\partial^2 \ln L_n(\gamma_n)}{\partial \gamma_u \partial \tau'}\Big) \cdot \sqrt{n}\tau_n\Big] + o_p(1). \tag{C.10}$$

Similarly,

$$\frac{1}{\sqrt{n}} \frac{\partial \ln L_n(\check{\gamma})}{\partial \tau} = \frac{1}{\sqrt{n}} \frac{\partial \ln L_n(\gamma_n)}{\partial \tau} - \frac{1}{n} \frac{\partial^2 \ln L_n(\bar{\gamma})}{\partial \tau \partial \tau'} \cdot \sqrt{n}\tau_n + \frac{1}{n} \frac{\partial^2 \ln L_n(\bar{\gamma})}{\partial \tau \partial \gamma_u'} \cdot \sqrt{n}(\check{\gamma}_u - \gamma_{u\infty}), \tag{C.11}$$

where $\frac{1}{n} \frac{\partial^2 \ln L_n(\bar{\gamma})}{\partial \tau \partial \tau'} = \frac{1}{n} \mathrm{E}(\frac{\partial^2 \ln L_n(\gamma_n)}{\partial \tau \partial \tau'}) + o_p(1)$ and $\frac{1}{n} \frac{\partial^2 \ln L_n(\bar{\gamma})}{\partial \tau \partial \gamma_u'} = \frac{1}{n} \mathrm{E}(\frac{\partial^2 \ln L_n(\gamma_n)}{\partial \tau \partial \gamma_u'}) + o_p(1)$. Plugging (C.10) into (C.11) yields $\frac{1}{\sqrt{n}} \frac{\partial \ln L_n(\check{\gamma})}{\partial \tau} = \Delta \cdot \frac{1}{\sqrt{n}} \frac{\partial \ln L_n(\gamma_n)}{\partial \gamma} + \frac{1}{n} \Lambda \cdot \sqrt{n}\tau_n + o_p(1)$. Since $\frac{1}{\sqrt{n}} \frac{\partial \ln L_n(\gamma_n)}{\partial \gamma} \xrightarrow{d} N(0, \lim_{n\to\infty} \mathcal{B})$, the result in the proposition follows. $\square$

SUPPLEMENTARY MATERIAL

Fei Jin and Yuqin Wang (2023): Supplement to "Consistent non-Gaussian pseudo maximum likelihood estimators of spatial autoregressive models," Econometric Theory Supplementary Material. To view, please visit: https://doi.org/10.1017/S0266466623000026

REFERENCES

Anselin, L. (1988) *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers.

Anselin, L. (2010) Thirty years of spatial econometrics. *Papers in Regional Science* 89, 3–25.

Anselin, L. & A. Bera (1998) Spatial dependence in linear regression models with an introduction to spatial econometrics. In A. Ullah and D.E. Giles (eds.), *Handbook of Applied Economic Statistics*, pp. 237–289. Marcel Dekker.

Arbia, G. (2014) *A Primer for Spatial Econometrics with Applications in R*. Springer.

Arbia, G. (2016) Spatial econometrics: A broad view. *Foundations and Trends in Econometrics* 8, 145–265.

Baltagi, B.H., P. Egger, & M. Pfaffermayr (2008) Estimating regional trade agreement effects on FDI in an interdependent world. *Journal of Econometrics* 145, 194–208.

Bao, Y. (2013) Finite-sample bias of the QMLE in spatial autoregressive models. *Econometric Theory* 29, 68–88.

Blommestein, H. (1983) Specification and estimation of spatial dependence: A discussion of alternative strategies for spatial economic modelling. *Regional Science and Urban Economics* 13, 251–270.

Blommestein, H. (1985) Elimination of circular routes in spatial dynamic regression equations. *Regional Science and Urban Economics* 15, 121–130.

Burridge, P. (1980) On the Cliff–Ord test for spatial autocorrelation. *Journal of the Royal Statistical Society, Series B* 42, 107–108.

Cliff, A. & J.K. Ord (1973) *Spatial Autocorrelation*. Pion.

Cliff, A. & J.K. Ord (1981) *Spatial Process: Models and Applications*. Pion.

Conley, T.G. (1999) GMM estimation with cross sectional dependence. *Journal of Econometrics* 92, 1–45.

Cressie, N. (1993) *Statistics for Spatial Data*. Wiley.

Davidson, J. (1994) *Stochastic Limit Theory: An Introduction for Econometricians*. Oxford University Press.

Doğan, O. & S. Taşpınar (2013) GMM estimation of spatial autoregressive models with moving average disturbances. *Regional Science and Urban Economics* 43, 903–926.

Fan, J., L. Qi, & D. Xiu (2014) Quasi-maximum likelihood estimation of GARCH models with heavy-tailed likelihoods. *Journal of Business & Economic Statistics* 32, 178–191.

Fang, K.T., S. Kotz, & K.W. Ng (1990) *Symmetric Multivariate and Related Distributions*. Chapman and Hall.

Fingleton, B. (2008) A generalized method of moments estimator for a spatial panel model with an endogenous spatial lag and spatial moving average errors. *Spatial Economic Analysis* 3, 27–44.

Fiorentini, G. & E. Sentana (2019) Consistent non-Gaussian pseudo maximum likelihood estimators. *Journal of Econometrics* 213, 321–358.

Francq, C., G. Lepage, & J.M. Zakoïan (2011) Two-stage non Gaussian QML estimation of GARCH models and testing the efficiency of the Gaussian QMLE. *Journal of Econometrics* 165, 246–257.

Giles, J.A. & D.E. Giles (1993) Pre-test estimation and testing in econometrics: Recent developments. *Journal of Economic Surveys* 7, 145–197.

Gilley, O.W. & R.K. Pace (1996) On the Harrison and Rubinfeld data. *Journal of Environmental Economics and Management* 31, 403–405.

Godfrey, L.G. & C.D. Orme (1991) Testing for skewness of regression disturbances. *Economics Letters* 37, 31–34.

Gouriéroux, C., A. Monfort, & A. Trognon (1984) Pseudo maximum likelihood methods: Theory. *Econometrica* 52, 681–700.

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

Haining, R.P. (1978) The moving average model for spatial interaction. *Transactions of the Institute of British Geographers* 3, 202–225.

Harrison, D.J. & D.L. Rubinfeld (1978) Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management* 5, 81–102.

Hillier, G. & F. Martellosio (2018) Exact and higher-order properties of the MLE in spatial autoregressive models, with applications to inference. *Journal of Econometrics* 205, 402–422.

Jarque, C.M. & A.K. Bera (1980) Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters* 6, 255–259.

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.F. Lee (2019) GEL estimation and tests of spatial autoregressive models. *Journal of Econometrics* 208, 585–612.

Jin, F., L.F. Lee, & K. Yang (2022) Best Linear and Quadratic Moments for Spatial Econometric Models and an Application to Spatial Interdependence Patterns of Employment Growth in US Counties. Working paper, School of Economics, Fudan University.

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

Kelejian, H.H. & I.R. Prucha (1999) A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review* 40, 509–533.

Kelejian, H.H. & I.R. Prucha (2001) On the asymptotic distribution of the Moran *I*test statistic with applications. *Journal of Econometrics* 104, 219–257.

Kelejian, H.H. & I.R. Prucha (2010) Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics* 157, 53–67.

Lee, J. & P.M. Robinson (2020) Adaptive inference on pure spatial models. *Journal of Econometrics* 216, 375–393.

Lee, L.F. (2002) Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models. *Econometric Theory* 18, 252–277.

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

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

LeSage, J. & R.K. Pace (2009) *Introduction to Spatial Econometrics*. Chapman & Hall/CRC.

LeSage, J.P. (1999) *Spatial Econometrics*. Department of Economics, University of Toledo.

LeSage, J.P. & R.K. Pace (2007) A matrix exponential spatial specification. *Journal of Econometrics* 140, 190–214.

Lin, M. & G. Sinnamon (2020) Revisiting a sharpened version of Hadamard's determinant inequality. *Linear Algebra and its Applications* 606, 192–200.

Liu, X., L.F. Lee, & C.R. Bollinger (2010) An efficient GMM estimator of spatial autoregressive models. *Journal of Econometrics* 159, 303–319.

Moran, P.A.P. (1950) Notes on continuous stochastic phenomena. *Biometrika* 35, 255–260.

Newey, W.K. & D. McFadden (1994) Large sample estimation and hypothesis testing. In R.F. Engle, D.L. McFadden (eds.), *Handbook of Econometrics*, vol. 4, Ch. 36, pp. 2111–2245. Elsevier.

Newey, W.K. & D.G. Steigerwald (1997) Asymptotic bias for quasi-maximum-likelihood estimators in conditional heteroskedasticity models. *Econometrica* 65, 587–599.

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

Pace, R.K. & R. Barry (1997) Quick computation of spatial autoregressive estimators. *Geographical Analysis* 29, 232–246.

Pace, R.K. & O.W. Gilley (1997) Using the spatial configuration of the data to improve estimation. *The Journal of Real Estate Finance and Economics* 14, 333–340.

Robinson, P.M. (2010) Efficient estimation of the semiparametric spatial autoregressive model. *Journal of Econometrics* 157, 6–17.

Spiring, F. (2011) The refined positive definite and unimodal regions for the Gram–Charlier and Edgeworth series expansion. *Advances in Decision Sciences* 2011, Article no. 463097.

White, H. (1994) *Estimation, Inference and Specification Analysis*. Cambridge University Press.

Xu, X. & L.F. Lee (2015) Maximum likelihood estimation of a spatial autoregressive Tobit model. *Journal of Econometrics* 188, 264–280.

Xu, X. & L.F. Lee (2018) Sieve maximum likelihood estimation of the spatial autoregressive Tobit model. *Journal of Econometrics* 203, 96–112.

Yang, K. & L.F. Lee (2017) Identification and QML estimation of multivariate and simultaneous equations spatial autoregressive models. *Journal of Econometrics* 196, 196–214.