Online supplement to “Consistent non-Gaussian pseudo maximum likelihood estimators of spatial autoregressive models”

Fei Jin[1] and Yuqin Wang*[2]

[1]School of Economics, Fudan University, and Shanghai Institute of International Finance and Economics, Shanghai 200433, China

[2]Institute for Advanced Research, Shanghai University of Finance and Economics, and Key Laboratory of Mathematical Economics (SUFE), Ministry of Education, Shanghai 200433, China

January 7, 2023

Contents

*Corresponding author. E-mail addresses: jinfei@fudan.edu.cn (F. Jin); wang.yuqin@sufe.edu.cn (Y. Wang).

1 Derivatives of the pseudo log likelihood function for the SARAR model

In this section, we present the first and second order derivatives of the pseudo log likelihood function $\ln L_n(\delta)$ in (4) for the SARAR model in (1). The derivatives of $\ln L_n(\gamma)$ in (2) for model (1) can be derived by removing irrelevant components of the derivatives below.

The first order derivatives of $\ln L_n(\delta)$ are

$$\frac{\partial \ln L_n(\delta)}{\partial \lambda} = -\frac{1}{\sigma}\sum_{i=1}^{n}\frac{\partial \ln f(v_i(\theta)-\frac{1}{\sigma}\alpha,\eta)}{\partial v}e'_{ni}R_n(\rho)W_nY_n - \mathrm{tr}[W_nS_n^{-1}(\lambda)], \tag{S1}$$

$$\frac{\partial \ln L_n(\delta)}{\partial \rho} = -\frac{1}{\sigma}\sum_{i=1}^{n}\frac{\partial \ln f(v_i(\theta)-\frac{1}{\sigma}\alpha,\eta)}{\partial v}e'_{ni}M_n[S_n(\lambda)Y_n - X_n\beta] - \mathrm{tr}[M_nR_n^{-1}(\rho)], \tag{S2}$$

$$\frac{\partial \ln L_n(\delta)}{\partial \beta} = -\frac{1}{\sigma}\sum_{i=1}^{n}\frac{\partial \ln f(v_i(\theta)-\frac{1}{\sigma}\alpha,\eta)}{\partial v}X'_nR'_n(\rho)e_{ni}, \tag{S3}$$

$$\frac{\partial \ln L_n(\delta)}{\partial \alpha} = -\frac{1}{\sigma}\sum_{i=1}^{n}\frac{\partial \ln f(v_i(\theta)-\frac{1}{\sigma}\alpha,\eta)}{\partial v}, \tag{S4}$$

$$\frac{\partial \ln L_n(\delta)}{\partial \sigma^2} = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}\frac{\partial \ln f(v_i(\theta)-\frac{1}{\sigma}\alpha,\eta)}{\partial v}[v_i(\theta)-\frac{1}{\sigma}\alpha] - \frac{n}{2\sigma^2}, \tag{S5}$$

$$\frac{\partial \ln L_n(\delta)}{\partial \eta} = \sum_{i=1}^{n}\frac{\partial \ln f(v_i(\theta)-\frac{1}{\sigma}\alpha,\eta)}{\partial \eta}. \tag{S6}$$

The second order derivatives are:

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

$$\frac{\partial^2 \ln L_n(\delta)}{\partial \lambda\partial\rho} = \frac{1}{\sigma^2}\sum_{i=1}^n \frac{\partial^2 \ln f(v_i(\theta)-\frac{1}{\sigma}\alpha,\eta)}{\partial v^2}e'_{ni}R_n(\rho)W_nY_ne'_{ni}M_n[S_n(\lambda)Y_n - X_n\beta]$$
$$+\frac{1}{\sigma}\sum_{i=1}^n \frac{\partial \ln f(v_i(\theta)-\frac{1}{\sigma}\alpha,\eta)}{\partial v}e'_{ni}M_nW_nY_n,$$

$$\frac{\partial^2 \ln L_n(\delta)}{\partial \lambda\partial\beta'} = \frac{1}{\sigma^2}\sum_{i=1}^n \frac{\partial^2 \ln f(v_i(\theta)-\frac{1}{\sigma}\alpha,\eta)}{\partial v^2}e'_{ni}R_n(\rho)W_nY_ne'_{ni}R_n(\rho)X_n,$$

$$\frac{\partial^2 \ln L_n(\delta)}{\partial \lambda\partial\alpha} = \frac{1}{\sigma^2}\sum_{i=1}^n \frac{\partial^2 \ln f(v_i(\theta)-\frac{1}{\sigma}\alpha,\eta)}{\partial v^2}e'_{ni}R_n(\rho)W_nY_n,$$

$$\frac{\partial^2 \ln L_n(\delta)}{\partial \lambda\partial\sigma^2} = \frac{1}{2\sigma^3}\sum_{i=1}^n \frac{\partial^2 \ln f(v_i(\theta)-\frac{1}{\sigma}\alpha,\eta)}{\partial v^2}e'_{ni}R_n(\rho)W_nY_n[v_i(\theta)-\frac{1}{\sigma}\alpha]$$
$$+\frac{1}{2\sigma^3}\sum_{i=1}^n \frac{\partial \ln f(v_i(\theta)-\frac{1}{\sigma}\alpha,\eta)}{\partial v}e'_{ni}R_n(\rho)W_nY_n,$$

$$\frac{\partial^2 \ln L_n(\delta)}{\partial \lambda\partial\eta'} = -\frac{1}{\sigma}\sum_{i=1}^n \frac{\partial^2 \ln f(v_i(\theta)-\frac{1}{\sigma}\alpha,\eta)}{\partial v\partial\eta'}e'_{ni}R_n(\rho)W_nY_n,$$

$$\frac{\partial^2 \ln L_n(\delta)}{\partial \rho^2} = \frac{1}{\sigma^2}\sum_{i=1}^n \frac{\partial^2 \ln f\big(v_i(\theta)-\frac{1}{\sigma}\alpha,\eta\big)}{\partial v^2}\{e'_{ni}M_n[S_n(\lambda)Y_n - X_n\beta]\}^2 - \mathrm{tr}\{[M_nR_n^{-1}(\rho)]^2\},$$

$$\frac{\partial^2 \ln L_n(\delta)}{\partial \rho\partial\beta'} = \frac{1}{\sigma^2}\sum_{i=1}^n \frac{\partial^2 \ln f(v_i(\theta)-\frac{1}{\sigma}\alpha,\eta)}{\partial v^2}e'_{ni}M_n[S_n(\lambda)Y_n - X_n\beta]e'_{ni}R_n(\rho)X_n$$
$$+\frac{1}{\sigma}\sum_{i=1}^n \frac{\partial \ln f(v_i(\theta)-\frac{1}{\sigma}\alpha,\eta)}{\partial v}e'_{ni}M_nX_n,$$

$$\frac{\partial^2 \ln L_n(\delta)}{\partial \rho\partial\alpha} = \frac{1}{\sigma^2}\sum_{i=1}^n \frac{\partial^2 \ln f(v_i(\theta)-\frac{1}{\sigma}\alpha,\eta)}{\partial v^2}e'_{ni}M_n[S_n(\lambda)Y_n - X_n\beta],$$

$$\frac{\partial^2 \ln L_n(\delta)}{\partial \rho\partial\sigma^2} = \frac{1}{2\sigma^3}\sum_{i=1}^n \frac{\partial^2 \ln f(v_i(\theta)-\frac{1}{\sigma}\alpha,\eta)}{\partial v^2}e'_{ni}M_n[S_n(\lambda)Y_n - X_n\beta][v_i(\theta)-\frac{1}{\sigma}\alpha]$$
$$+\frac{1}{2\sigma^3}\sum_{i=1}^n \frac{\partial \ln f(v_i(\theta)-\frac{1}{\sigma}\alpha,\eta)}{\partial v}e'_{ni}M_n[S_n(\lambda)Y_n - X_n\beta],$$

$$\frac{\partial^2 \ln L_n(\delta)}{\partial \rho\partial\eta'} = -\frac{1}{\sigma}\sum_{i=1}^n \frac{\partial^2 \ln f(v_i(\theta)-\frac{1}{\sigma}\alpha,\eta)}{\partial v\partial\eta'}e'_{ni}M_n[S_n(\lambda)Y_n - X_n\beta],$$

$$\frac{\partial^2 \ln L_n(\delta)}{\partial \beta\partial\beta'} = \frac{1}{\sigma^2}\sum_{i=1}^n \frac{\partial^2 \ln f(v_i(\theta)-\frac{1}{\sigma}\alpha,\eta)}{\partial v^2}X'_nR'_n(\rho)e_{ni}e'_{ni}R_n(\rho)X_n,$$

$$
\begin{aligned}
\frac{\partial^2 \ln L_n(\delta)}{\partial\beta\partial\alpha} &= \frac{1}{\sigma^2}\sum_{i=1}^{n}\frac{\partial^2 \ln f(v_i(\theta)-\frac{1}{\sigma}\alpha,\eta)}{\partial v^2}X_n'R_n'(\rho)e_{ni},\\
\frac{\partial^2 \ln L_n(\delta)}{\partial\beta\partial\sigma^2} &= \frac{1}{2\sigma^3}\sum_{i=1}^{n}\frac{\partial^2 \ln f(v_i(\theta)-\frac{1}{\sigma}\alpha,\eta)}{\partial v^2}X_n'R_n'(\rho)e_{ni}[v_i(\theta)-\frac{1}{\sigma}\alpha]\\
&\quad+\frac{1}{2\sigma^3}\sum_{i=1}^{n}\frac{\partial \ln f(v_i(\theta)-\frac{1}{\sigma}\alpha,\eta)}{\partial v}X_n'R_n'(\rho)e_{ni},\\
\frac{\partial^2 \ln L_n(\delta)}{\partial\beta\partial\eta'} &= -\frac{1}{\sigma}\sum_{i=1}^{n}\frac{\partial^2 \ln f(v_i(\theta)-\frac{1}{\sigma}\alpha,\eta)}{\partial v\partial\eta'}X_n'R_n'(\rho)e_{ni},\\
\frac{\partial^2 \ln L_n(\delta)}{\partial\alpha^2} &= \frac{1}{\sigma^2}\sum_{i=1}^{n}\frac{\partial^2 \ln f(v_i(\theta)-\frac{1}{\sigma}\alpha,\eta)}{\partial v^2},\\
\frac{\partial^2 \ln L_n(\delta)}{\partial\alpha\partial\sigma^2} &= \frac{1}{2\sigma^3}\sum_{i=1}^{n}\frac{\partial^2 \ln f(v_i(\theta)-\frac{1}{\sigma}\alpha,\eta)}{\partial v^2}[v_i(\theta)-\frac{1}{\sigma}\alpha]+\frac{1}{2\sigma^3}\sum_{i=1}^{n}\frac{\partial \ln f(v_i(\theta)-\frac{1}{\sigma}\alpha,\eta)}{\partial v},\\
\frac{\partial^2 \ln L_n(\delta)}{\partial\alpha\partial\eta'} &= -\frac{1}{\sigma}\sum_{i=1}^{n}\frac{\partial^2 \ln f(v_i(\theta)-\frac{1}{\sigma}\alpha,\eta)}{\partial v\partial\eta'},\\
\frac{\partial^2 \ln L_n(\delta)}{\partial(\sigma^2)^2} &= \frac{1}{4\sigma^4}\sum_{i=1}^{n}\frac{\partial^2 \ln f(v_i(\theta)-\frac{1}{\sigma}\alpha,\eta)}{\partial v^2}[v_i(\theta)-\frac{1}{\sigma}\alpha]^2\\
&\quad+\frac{3}{4\sigma^4}\sum_{i=1}^{n}\frac{\partial \ln f(v_i(\theta)-\frac{1}{\sigma}\alpha,\eta)}{\partial v}[v_i(\theta)-\frac{1}{\sigma}\alpha]+\frac{n}{2\sigma^4},\\
\frac{\partial^2 \ln L_n(\delta)}{\partial\sigma^2\partial\eta'} &= -\frac{1}{2\sigma^2}\sum_{i=1}^{n}\frac{\partial^2 \ln f(v_i(\theta)-\frac{1}{\sigma}\alpha,\eta)}{\partial v\partial\eta'}[v_i(\theta)-\frac{1}{\sigma}\alpha],\\
\frac{\partial^2 \ln L_n(\delta)}{\partial\eta\partial\eta'} &= \sum_{i=1}^{n}\frac{\partial^2 \ln f(v_i(\theta)-\frac{1}{\sigma}\alpha,\eta)}{\partial\eta\partial\eta'}.
\end{aligned}
$$

2 Tests for normality, skewness and excess kurtosis of the innovations in the SARAR model

In this section, we consider tests for normality, skewness and excess kurtosis of the innovations in the following SARAR model:

$$
Y_n = \lambda_0 W_n Y_n + X_n\beta_0 + U_n, \quad U_n = \rho_0 M_n U_n + \mathcal{V}_n, \tag{S7}
$$

where the elements ν_i's of $\mathcal{V}_n = [\nu_1, \cdots, \nu_n]'$ are independent with mean zero. The above model with $\mathcal{V}_n = \sigma_0 V_n$ is the same as that in (1) of the main text, but we write $\mathcal{V}_n$ as a whole term for analytic convenience. Note that by removing some irrelevant terms in the test statistics of the following subsections, we can easily obtain tests for the SAR model and the spatial error model, which are nested in the SARAR model.

2.1 Test for normality of the innovations

Assume that ν_i follows the Pearson distribution with the probability density function:

$$h(x,\eta)=\frac{\exp(-\int \frac{x+\eta_2}{\eta_3x^2+\eta_2x+\eta_1}\,\mathrm{d}x)}{\int_{-\infty}^{\infty}\exp(-\int \frac{x+\eta_2}{\eta_3x^2+\eta_2x+\eta_1}\,\mathrm{d}x)\,\mathrm{d}x}=\frac{\exp(-h_1(x,\eta))}{\int_{-\infty}^{\infty}\exp(-h_1(x,\eta))\,\mathrm{d}x},$$

where η_1, η_2 and η_3 are constants, $\eta=[\eta_1,\eta_2,\eta_3]'$, and $h_1(x,\eta)=\int \frac{x+\eta_2}{\eta_3x^2+\eta_2x+\eta_1}\,\mathrm{d}x$. The null hypothesis for the normality test is H_0: $\eta_2=\eta_3=0$. The log likelihood function of model (S7) is

$$\begin{aligned}\ln L_n(\gamma)&=\sum_{i=1}^n \ln h(\nu_i(\theta),\eta)+\ln|S_n(\lambda)|+\ln|R_n(\rho)|\\&=-\sum_{i=1}^n h_1(\nu_i(\theta),\eta)-n\ln\Big[\int_{-\infty}^{\infty}\exp(-h_1(x,\eta))\,\mathrm{d}x\Big]+\ln|S_n(\lambda)|+\ln|R_n(\rho)|,\end{aligned}$$

where $\theta=[\lambda,\rho,\beta']'$, $\gamma=[\theta',\eta']'$, and $\nu_i(\theta)=e_{ni}'R_n(\rho)[S_n(\lambda)Y_n-X_n\beta]$. We have the following derivatives of $\ln L_n(\gamma)$:

$$\begin{aligned}\frac{\partial \ln L_n(\gamma)}{\partial\lambda}&=\sum_{i=1}^n\frac{\partial h_1(\nu_i(\theta),\eta)}{\partial\nu}e_{ni}'R_n(\rho)W_nY_n-\mathrm{tr}[W_nS_n^{-1}(\lambda)],\\\frac{\partial \ln L_n(\gamma)}{\partial\rho}&=\sum_{i=1}^n\frac{\partial h_1(\nu_i(\theta),\eta)}{\partial\nu}e_{ni}'M_n[S_n(\lambda)Y_n-X_n\beta]-\mathrm{tr}[M_nR_n^{-1}(\rho)],\\\frac{\partial \ln L_n(\gamma)}{\partial\beta}&=\sum_{i=1}^n\frac{\partial h_1(\nu_i(\theta),\eta)}{\partial\nu}X_n'R_n'(\rho)e_{ni},\\\frac{\partial \ln L_n(\gamma)}{\partial\eta}&=-\sum_{i=1}^n\frac{\partial h_1(\nu_i(\theta),\eta)}{\partial\eta}+\frac{n}{\int_{-\infty}^{\infty}\exp(-h_1(x,\eta))\,\mathrm{d}x}\int_{-\infty}^{\infty}\exp(-h_1(x,\eta))\frac{\partial h_1(x,\eta)}{\partial\eta}\,\mathrm{d}x,\end{aligned}$$

where $\frac{\partial h_1(x,\eta)}{\partial x}=\frac{x+\eta_2}{h_2(x,\eta)}$ with $h_2(x,\eta)=\eta_3x^2+\eta_2x+\eta_1$, and

$$\frac{\partial h_1(x,\eta)}{\partial\eta}=\Big[-\int\frac{x+\eta_2}{(h_2(x,\eta))^2}\,\mathrm{d}x,\int\Big(\frac{1}{h_2(x,\eta)}-\frac{x(x+\eta_2)}{(h_2(x,\eta))^2}\Big)\,\mathrm{d}x,-\int\frac{x^2(x+\eta_2)}{(h_2(x,\eta))^2}\,\mathrm{d}x\Big]'.$$

The second order derivatives of $\ln L_n(\gamma)$ are

$$\frac{\partial^2 \ln L_n(\gamma)}{\partial\lambda^2} = -\sum_{i=1}^n \frac{\partial^2 h_1(\nu_i(\theta),\eta)}{\partial\nu^2}[e'_{ni}R_n(\rho)W_nY_n]^2 - \operatorname{tr}[W_nS_n^{-1}(\lambda)W_nS_n^{-1}(\lambda)],$$

$$\frac{\partial^2 \ln L_n(\gamma)}{\partial\lambda\partial\rho} = -\sum_{i=1}^n \frac{\partial^2 h_1(\nu_i(\theta),\eta)}{\partial\nu^2}e'_{ni}R_n(\rho)W_nY_ne'_{ni}M_n[S_n(\lambda)Y_n - X_n\beta] - \sum_{i=1}^n \frac{\partial h_1(\nu_i(\theta),\eta)}{\partial\nu}e'_{ni}M_nW_nY_n,$$

$$\frac{\partial^2 \ln L_n(\gamma)}{\partial\lambda\partial\beta'} = -\sum_{i=1}^n \frac{\partial^2 h_1(\nu_i(\theta),\eta)}{\partial\nu^2}e'_{ni}R_n(\rho)W_nY_ne'_{ni}R_n(\rho)X_n,$$

$$\frac{\partial^2 \ln L_n(\gamma)}{\partial\lambda\partial\eta'} = \sum_{i=1}^n \frac{\partial^2 h_1(\nu_i(\theta),\eta)}{\partial\nu\partial\eta'}e'_{ni}R_n(\rho)W_nY_n,$$

$$\frac{\partial^2 \ln L_n(\gamma)}{\partial\rho^2} = -\sum_{i=1}^n \frac{\partial^2 h_1(\nu_i(\theta),\eta)}{\partial\nu^2}\{e'_{ni}M_n[S_n(\lambda)Y_n - X_n\beta]\}^2 - \operatorname{tr}[M_nR_n^{-1}(\rho)M_nR_n^{-1}(\rho)],$$

$$\frac{\partial^2 \ln L_n(\gamma)}{\partial\rho\partial\beta'} = -\sum_{i=1}^n \frac{\partial^2 h_1(\nu_i(\theta),\eta)}{\partial\nu^2}e'_{ni}M_n[S_n(\lambda)Y_n - X_n\beta]e'_{ni}R_n(\rho)X_n - \sum_{i=1}^n \frac{\partial h_1(\nu_i(\theta),\eta)}{\partial\nu}e'_{ni}M_nX_n,$$

$$\frac{\partial^2 \ln L_n(\gamma)}{\partial\rho\partial\eta'} = \sum_{i=1}^n \frac{\partial^2 h_1(\nu_i(\theta),\eta)}{\partial\nu\partial\eta'}e'_{ni}M_n[S_n(\lambda)Y_n - X_n\beta],$$

$$\frac{\partial^2 \ln L_n(\gamma)}{\partial\beta\partial\beta'} = -\sum_{i=1}^n \frac{\partial^2 h_1(\nu_i(\theta),\eta)}{\partial\nu^2}X'_nR'_n(\rho)e_{ni}e'_{ni}R_n(\rho)X_n,$$

$$\frac{\partial^2 \ln L_n(\gamma)}{\partial\beta\partial\eta'} = \sum_{i=1}^n X'_nR'_n(\rho)e_{ni}\frac{\partial^2 h_1(\nu_i(\theta),\eta)}{\partial\nu\partial\eta'},$$

$$\frac{\partial^2 \ln L_n(\gamma)}{\partial\eta\partial\eta'} = -\sum_{i=1}^n \frac{\partial^2 h_1(\nu_i(\theta),\eta)}{\partial\eta\partial\eta'} + \frac{n}{[\int_{-\infty}^{\infty}\exp(-h_1(x,\eta))\,\mathrm{d}x]^2}h_3(\eta)h'_3(\eta) - \frac{n}{\int_{-\infty}^{\infty}\exp(-h_1(x,\eta))\,\mathrm{d}x}\int_{-\infty}^{\infty}\exp(-h_1(x,\eta))\frac{\partial h_1(x,\eta)}{\partial\eta}\frac{\partial h_1(x,\eta)}{\partial\eta'}\,\mathrm{d}x + \frac{n}{\int_{-\infty}^{\infty}\exp(-h_1(x,\eta))\,\mathrm{d}x}\int_{-\infty}^{\infty}\exp(-h_1(x,\eta))\frac{\partial^2 h_1(x,\eta)}{\partial\eta\partial\eta'}\,\mathrm{d}x,$$

where $\frac{\partial^2 h_1(x,\eta)}{\partial x^2} = \frac{1}{h_2(x,\eta)} - \frac{(x+\eta_2)(2\eta_3 x+\eta_2)}{(h_2(x,\eta))^2}$, $\frac{\partial^2 h_1(x,\eta)}{\partial x \partial \eta'} = [-\frac{x+\eta_2}{(h_2(x,\eta))^2}, \frac{1}{h_2(x,\eta)} - \frac{x(x+\eta_2)}{(h_2(x,\eta))^2}, -\frac{x^2(x+\eta_2)}{(h_2(x,\eta))^2}]$,

$$\frac{\partial^2 h_1(x,\eta)}{\partial \eta \partial \eta'} = \begin{pmatrix} \int \frac{2(x+\eta_2)}{(h_2(x,\eta))^3}\,\mathrm{d}x & * & * \\ \int [-\frac{1}{(h_2(x,\eta))^2} + \frac{2x(x+\eta_2)}{(h_2(x,\eta))^3}]\,\mathrm{d}x & \int [-\frac{2x}{(h_2(x,\eta))^2} + \frac{2x^2(x+\eta_2)}{(h_2(x,\eta))^3}]\,\mathrm{d}x & * \\ \int \frac{2x^2(x+\eta_2)}{(h_2(x,\eta))^3}\,\mathrm{d}x & \int [-\frac{x^2}{(h_2(x,\eta))^2} + \frac{2x^3(x+\eta_2)}{(h_2(x,\eta))^3}]\,\mathrm{d}x & \int \frac{2x^4(x+\eta_2)}{(h_2(x,\eta))^3}\,\mathrm{d}x \end{pmatrix},$$

and $h_3(\eta) = \int_{-\infty}^{\infty} \exp(-h_1(x,\eta)) \frac{\partial h_1(x,\eta)}{\partial \eta}\,\mathrm{d}x$.

Let $\check{\gamma} = [\check{\theta}', \check{\eta}']'$ be the restricted MLE with $\eta_2 = 0$ and $\eta_3 = 0$ imposed, where $\check{\eta} = [\check{\eta}_1, 0, 0]'$. With $\eta_2 = 0$ and $\eta_3 = 0$, since $h(x,\eta) = \frac{1}{\sqrt{2\pi\eta_1}} \exp(-\frac{x^2}{2\eta_1})$ is the probability density function of the normal distribution $N(0, \eta_1)$, by Lee (2004), $\check{\eta}_1 = \frac{1}{n} \check{\mathcal{V}}_n' \check{\mathcal{V}}_n$, where $\check{\mathcal{V}}_n = R_n(\check{\rho})[S_n(\check{\lambda})Y_n - X_n\check{\beta}]$. With $\eta = \check{\eta}$, we have $h_2(x,\check{\eta}) = \check{\eta}_1$, $h_1(x,\check{\eta}) = \frac{1}{2\check{\eta}_1}x^2$, $\frac{\partial h_1(x,\check{\eta})}{\partial x} = \frac{1}{\check{\eta}_1}x$, $\frac{\partial^2 h_1(x,\check{\eta})}{\partial x^2} = \frac{1}{\check{\eta}_1}$, $\frac{\partial h_1(x,\check{\eta})}{\partial \eta} = [-\frac{1}{2\check{\eta}_1^2}x^2, \frac{1}{\check{\eta}_1}x - \frac{1}{3\check{\eta}_1^2}x^3, -\frac{1}{4\check{\eta}_1^2}x^4]'$, $\frac{\partial^2 h_1(x,\check{\eta})}{\partial x \partial \eta'} = [-\frac{1}{\check{\eta}_1^2}x, \frac{1}{\check{\eta}_1} - \frac{1}{\check{\eta}_1^2}x^2, -\frac{1}{\check{\eta}_1^2}x^3]$, and

$$\frac{\partial^2 h_1(x,\check{\eta})}{\partial \eta \partial \eta'} = \begin{pmatrix} \frac{1}{\check{\eta}_1^3}x^2 & * & * \\ -\frac{1}{\check{\eta}_1^2}x + \frac{2}{3\check{\eta}_1^3}x^3 & -\frac{1}{\check{\eta}_1^2}x^2 + \frac{1}{2\check{\eta}_1^3}x^4 & * \\ \frac{1}{2\check{\eta}_1^3}x^4 & -\frac{1}{3\check{\eta}_1^2}x^3 + \frac{2}{5\check{\eta}_1^3}x^5 & \frac{1}{3\check{\eta}_1^3}x^6 \end{pmatrix}.$$

Denote $\gamma_1 = [\theta', \eta_1]'$ and $\gamma_2 = [\eta_2, \eta_3]'$. Then $\frac{\partial \ln L_n(\check{\gamma})}{\partial \gamma_1} = 0$, and

$$\frac{\partial \ln L_n(\check{\gamma})}{\partial \eta_2} = \frac{1}{3\check{\eta}_1^2} \sum_{i=1}^n \nu_i^3(\check{\theta}) - \frac{1}{\check{\eta}_1} \sum_{i=1}^n \nu_i(\check{\theta}), \tag{S8}$$

$$\frac{\partial \ln L_n(\check{\gamma})}{\partial \eta_3} = \frac{1}{4\check{\eta}_1^2} \sum_{i=1}^n \nu_i^4(\check{\theta}) - \frac{3}{4}n. \tag{S9}$$

At the true parameter value $\gamma_0 = [\theta_0', \eta_{10}, 0, 0]'$, we have the reduced form $Y_n = S_n^{-1}(X_n\beta_0 + R_n^{-1}\mathcal{V}_n)$, where $\mathcal{V}_n \sim N(0, \eta_{10} I_n)$. Denote $D_n = R_n W_n S_n^{-1} R_n^{-1}$, $Z_n = M_n R_n^{-1}$, $\Upsilon_n = R_n W_n S_n^{-1} X_n \beta_0$, and $\mathcal{A}_{ij} = -\frac{1}{n} \mathrm{E}(\frac{\partial^2 \ln L_n(\gamma_0)}{\partial \gamma_i \partial \gamma_j'})$ for $i, j = 1, 2$. Then $\mathcal{A}_{22} = \begin{pmatrix} \frac{2}{3\eta_{10}} & 0 \\ 0 & 6 \end{pmatrix}$,

$$\mathcal{A}_{21} = \begin{pmatrix} 0 & 0 & 0_{1\times k_x} & 0 \\ \frac{3}{n}\operatorname{tr}(D_n) & \frac{3}{n}\operatorname{tr}(Z_n) & 0_{1\times k_x} & \frac{3}{2\eta_{10}} \end{pmatrix},$$

and

$$\mathcal{A}_{11} = \begin{pmatrix} \frac{1}{n\eta_{10}}\Upsilon_n'\Upsilon_n + \frac{1}{n}\operatorname{tr}(D_n^s D_n) & \frac{1}{n}\operatorname{tr}(D_n^s Z_n) & \frac{1}{n\eta_{10}}\Upsilon_n' R_n X_n & \frac{1}{n\eta_{10}}\operatorname{tr}(D_n) \\ * & \frac{1}{n}\operatorname{tr}(Z_n^s Z_n) & 0_{1\times k_x} & \frac{1}{n\eta_{10}}\operatorname{tr}(Z_n) \\ * & * & \frac{1}{n\eta_{10}} X_n' R_n' R_n X_n & 0 \\ * & * & * & \frac{1}{2\eta_{10}^2} \end{pmatrix},$$

where $A^s = A + A'$ for any square matrix A. Let $\check{\mathcal{A}}_{ij}$ be the matrices derived by replacing $\gamma_{10} = [\theta_0', \eta_{10}]'$ in $\mathcal{A}_{ij}$ with $\check{\gamma}_1 = [\check{\theta}', \check{\eta}_1]'$. Then by the Lagrange multiplier principle, the normality test statistic

$$\frac{1}{n}\frac{\partial \ln L_n(\check{\gamma})}{\partial \gamma_2'}(\check{\mathcal{A}}_{22} - \check{\mathcal{A}}_{21}\check{\mathcal{A}}_{11}^{-1}\check{\mathcal{A}}_{12})^{-1}\frac{\partial \ln L_n(\check{\gamma})}{\partial \gamma_2}$$

is asymptotically chi-squared distributed with 2 degrees of freedom. We omit the detailed proof for this asymptotic distribution. The analysis will follow by the asymptotic properties of the maximum likelihood estimator $\check{\gamma}_1$ for the SARAR model, which are provided in, e.g., Jin and Lee (2013).

2.2 Test for skewness of the innovations

In this and the next subsections, let σ^2 be the variance of ν_i in model (S7) and $\theta = [\lambda, \rho, \beta', \sigma^2]'$. Denote the GPMLE of θ by $\check{\theta}$, $\check{\nu}_i = \nu_i(\check{\theta}) = e_{ni}' R_n(\check{\rho})[S_n(\check{\lambda})Y_n - X_n\check{\beta}]$, and $\mu_{k0} = \mathrm{E}(\nu_i^k)$ for $k = 3, \dots, 8$. The μ_{k0} can be estimated by $\check{\mu}_k \equiv \frac{1}{n}\sum_{i=1}^n \check{\nu}_i^k$. Under typical assumptions in the spatial econometric literature, we have $\check{\mu}_k = \mu_{k0} + o_p(1)$ for $k = 3, \dots, 8$, for which the proof is omitted for simplicity. To test for skewness of ν_i, we investigate the asymptotic distribution of the third moment $\check{\mu}_3$. By the mean value theorem,

$$\check{\mu}_3 = \frac{1}{n}\sum_{i=1}^n \nu_i^3 + \frac{3}{n}\sum_{i=1}^n \nu_i^2(\tilde{\theta})\frac{\partial \nu_i(\tilde{\theta})}{\partial \theta'}(\check{\theta} - \theta_0), \tag{S10}$$

where $\tilde{\theta}$ lies between θ_0 and $\check{\theta}$. Let $\ln \mathcal{L}_n(\theta)$ be the pseudo Gaussian log likelihood function of model (S7). By Jin and Lee (2018), under regularity conditions, $\sqrt{n}(\check{\theta} - \theta_0) = \mathbb{A}^{-1}\frac{1}{\sqrt{n}}\frac{\partial \ln \mathcal{L}_n(\theta_0)}{\partial \theta} + o_p(1)$, where

$$\frac{\partial \ln \mathcal{L}_n(\theta_0)}{\partial \theta} = \Big[\frac{1}{\sigma_0^2}\mathcal{V}_n' D_n \mathcal{V}_n - \operatorname{tr}(D_n) + \frac{1}{\sigma_0^2}\Upsilon_n'\mathcal{V}_n, \frac{1}{\sigma_0^2}\mathcal{V}_n' Z_n \mathcal{V}_n - \operatorname{tr}(Z_n), \frac{1}{\sigma_0^2}\mathcal{V}_n' R_n X_n, \frac{1}{2\sigma_0^4}\mathcal{V}_n'\mathcal{V}_n - \frac{n}{2\sigma_0^2}\Big]',$$

and $\mathbb{A} \equiv -\,\mathrm{E}(\frac{1}{n}\frac{\partial^2 \ln \mathcal{L}_n(\theta_0)}{\partial\theta\partial\theta'})$ has the expression

$$\mathbb{A} = \begin{pmatrix} \frac{1}{n\sigma_0^2}\Upsilon_n'\Upsilon_n + \frac{1}{n}\operatorname{tr}(D_n^s D_n) & \frac{1}{n}\operatorname{tr}(D_n^s Z_n) & \frac{1}{n\sigma_0^2}\Upsilon_n' R_n X_n & \frac{1}{n\sigma_0^2}\operatorname{tr}(D_n) \\ * & \frac{1}{n}\operatorname{tr}(Z_n^s Z_n) & 0 & \frac{1}{n\sigma_0^2}\operatorname{tr}(Z_n) \\ * & * & \frac{1}{n\sigma_0^2}X_n' R_n' R_n X_n & 0 \\ * & * & * & \frac{1}{2\sigma_0^4} \end{pmatrix},$$

where D_n, Z_n and Υ_n are defined in the last subsection. Thus, by (S10),

$$\sqrt{n}\check{\mu}_3 = \Big[1, \frac{3}{n}\sum_{i=1}^n \nu_i^2(\tilde{\theta})\frac{\partial \nu_i(\tilde{\theta})}{\partial \theta'}\mathbb{A}^{-1}\Big]\begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^n \nu_i^3 \\ \frac{1}{\sqrt{n}}\frac{\partial \ln \mathcal{L}_n(\theta_0)}{\partial\theta} \end{pmatrix} + o_p(1).$$

Under the condition that $\mu_{30} = 0$, by Lemma B.4, $[\frac{1}{\sqrt{n}}\sum_{i=1}^n \nu_i^3, \frac{1}{\sqrt{n}}\frac{\partial \ln \mathcal{L}_n(\theta_0)}{\partial\theta'}]'$ is asymptotically normal with mean zero. The variance $\mathbb{B} \equiv \operatorname{var}(\frac{1}{\sqrt{n}}\frac{\partial \ln \mathcal{L}_n(\theta_0)}{\partial\theta})$ has the expression

$$\mathbb{B} = \begin{pmatrix} \mathbb{B}_{11} & \mathbb{B}_{12} & \frac{\mu_{30}}{n\sigma_0^4}\operatorname{vec}_{\mathrm{D}}'(D_n)R_n X_n + \frac{1}{n\sigma_0^2}\Upsilon_n' R_n X_n & \frac{1}{2n\sigma_0^2}(\frac{\mu_{40}}{\sigma_0^4} - 1)\operatorname{tr}(D_n) + \frac{\mu_{30}}{2n\sigma_0^6}\Upsilon_n' 1_n \\ * & \mathbb{B}_{22} & \frac{\mu_{30}}{n\sigma_0^4}\operatorname{vec}_{\mathrm{D}}'(Z_n)R_n X_n & \frac{1}{2n\sigma_0^2}(\frac{\mu_{40}}{\sigma_0^4} - 1)\operatorname{tr}(Z_n) \\ * & * & \frac{1}{n\sigma_0^2}X_n' R_n' R_n X_n & \frac{\mu_{30}}{2n\sigma_0^6}X_n' R_n' 1_n \\ * & * & * & \frac{1}{4\sigma_0^4}(\frac{\mu_{40}}{\sigma_0^4} - 1) \end{pmatrix},$$

where $\operatorname{vec}_{\mathrm{D}}(A)$ for a square matrix A denotes a column vector formed by the diagonal elements of A, $\mathbb{B}_{11} = \frac{1}{n}(\frac{\mu_{40}}{\sigma_0^4} - 3)\operatorname{vec}_{\mathrm{D}}'(D_n)\operatorname{vec}_{\mathrm{D}}(D_n) + \frac{1}{n}\operatorname{tr}(D_n^s D_n) + \frac{2\mu_{30}}{n\sigma_0^4}\Upsilon_n'\operatorname{vec}_{\mathrm{D}}(D_n) + \frac{1}{n\sigma_0^2}\Upsilon_n'\Upsilon_n$, $\mathbb{B}_{12} = \frac{1}{n}(\frac{\mu_{40}}{\sigma_0^4} - 3)\operatorname{vec}_{\mathrm{D}}'(D_n)\operatorname{vec}_{\mathrm{D}}(Z_n) + \frac{1}{n}\operatorname{tr}(D_n^s Z_n) + \frac{\mu_{30}}{n\sigma_0^4}\Upsilon_n'\operatorname{vec}_{\mathrm{D}}(Z_n)$, and $\mathbb{B}_{22} = \frac{1}{n}(\frac{\mu_{40}}{\sigma_0^4} - 3)\operatorname{vec}_{\mathrm{D}}'(Z_n)\operatorname{vec}_{\mathrm{D}}(Z_n) + \frac{1}{n}\operatorname{tr}(Z_n^s Z_n)$. Using $\mu_{30} = 0$, some terms of $\mathbb{B}$ can be simplified. With $\mu_{30} = 0$, we have $\operatorname{var}(\frac{1}{\sqrt{n}}\sum_{i=1}^n \nu_i^3) = \mu_{60}$,

$$\mathrm{E}\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^n \nu_i^3 \cdot \frac{1}{\sqrt{n}}\frac{\partial \ln \mathcal{L}_n(\theta_0)}{\partial\theta'}\Big) = \Big[\frac{\mu_{50}}{n\sigma_0^2}\operatorname{tr}(D_n) + \frac{\mu_{40}}{n\sigma_0^2}\Upsilon_n' 1_n, \frac{\mu_{50}}{n\sigma_0^2}\operatorname{tr}(Z_n), \frac{\mu_{40}}{n\sigma_0^2}1_n' R_n X_n, \frac{\mu_{50}}{2\sigma_0^4}\Big].$$

Furthermore, under the assumption that $\mu_{30} = 0$,

$$\frac{3}{n}\sum_{i=1}^n \nu_i^2(\tilde{\theta})\frac{\partial \nu_i(\tilde{\theta})}{\partial\theta'} = \frac{3}{n}\sum_{i=1}^n \nu_i^2\frac{\partial \nu_i(\theta_0)}{\partial\theta'} + o_p(1) = \mathbb{F} + o_p(1),$$

where $\mathbb{F} = -\frac{3\sigma_0^2}{n}[1_n'\Upsilon_n, 0, 1_n' R_n X_n, 0]$. Let $\check{\mathbb{C}}$ be an estimate of $\mathbb{C} \equiv \operatorname{var}([\frac{1}{\sqrt{n}}\sum_{i=1}^n \nu_i^3, \frac{1}{\sqrt{n}}\frac{\partial \ln \mathcal{L}_n(\theta_0)}{\partial\theta'}]')$ by replacing θ_0, μ_{40}, μ_{50}, μ_{60} in $\mathbb{C}$ with, respectively $\check{\theta}$, $\check{\mu}_4$, $\check{\mu}_5$, $\check{\mu}_6$. Define $\check{\mathbb{A}}$ and $\check{\mathbb{F}}$ similarly.

Then

$$\frac{\sqrt{n}\check{\mu}_3}{\sqrt{[1, \check{\mathbb{F}}\check{\mathbb{A}}^{-1}]\check{\mathbb{C}}[1, \check{\mathbb{F}}\check{\mathbb{A}}^{-1}]'}}$$

is asymptotically standard normal under the condition that $\mu_{30} = 0$. This statistic can be used to test skewness of ν_i.

2.3 Test for excess kurtosis of the innovations

To test whether the kurtosis of ν_i is greater than 3 or not, we consider the asymptotic distribution of $\check{\mu}_4 - 3\check{\sigma}^4$. As $3\sigma^4 = 3(\sigma^2)^2$, by expanding $3\check{\sigma}^4$ at $\sigma^2 = \sigma_0^2$, we have $3\check{\sigma}^4 = 3\sigma_0^4 + 6\tilde{\sigma}^2(\check{\sigma}^2 - \sigma_0^2)$, where $\tilde{\sigma}^2$ lies between $\check{\sigma}^2$ and σ_0^2. Then by the mean value theorem,

$$\begin{aligned}
\sqrt{n}(\check{\mu}_4 - 3\check{\sigma}^4) &= \sqrt{n}(\check{\mu}_4 - 3\sigma_0^4) - 6\tilde{\sigma}^2 \cdot \sqrt{n}(\check{\sigma}^2 - \sigma_0^2) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\nu_i^4 - 3\sigma_0^4) - 6\tilde{\sigma}^2 \cdot \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\nu_i^2 - \sigma_0^2) \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}[4\nu_i^3(\tilde{\theta}) - 12\tilde{\sigma}^2 v_i(\tilde{\theta})]\frac{\partial \nu_i(\tilde{\theta})}{\partial \theta'}\sqrt{n}(\check{\theta} - \theta_0) \\
&= \Big[1, -6\tilde{\sigma}^2, \frac{1}{n}\sum_{i=1}^{n}[4\nu_i^3(\tilde{\theta}) - 12\tilde{\sigma}^2 v_i(\tilde{\theta})]\frac{\partial \nu_i(\tilde{\theta})}{\partial \theta'}\mathbb{A}^{-1}\Big]\begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\nu_i^4 - 3\sigma_0^4) \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\nu_i^2 - \sigma_0^2) \\ \frac{1}{\sqrt{n}}\frac{\partial \ln \mathcal{L}_n(\theta_0)}{\partial \theta} \end{pmatrix} + o_p(1),
\end{aligned}$$

where $\tilde{\theta}$ lies between θ_0 and $\check{\theta}$, and $\mathbb{A}$ is in the last subsection. With $\mu_{40} = 3\sigma_0^4$, we have $\mathrm{var}[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\nu_i^4 - 3\sigma_0^4)] = \mu_{80} - 9\sigma_0^8$, $\mathrm{E}[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\nu_i^4 - 3\sigma_0^4) \cdot \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\nu_i^2 - \sigma_0^2)] = \mu_{60} - 3\sigma_0^6$,

$$\begin{aligned}
&\mathrm{E}\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\nu_i^4 - 3\sigma_0^4) \cdot \frac{1}{\sqrt{n}}\frac{\partial \ln \mathcal{L}_n(\theta_0)}{\partial \theta'}\Big) \\
&= \Big[\frac{\mu_{60} - 3\sigma_0^6}{n\sigma_0^2}\mathrm{tr}(D_n) + \frac{\mu_{50}}{n\sigma_0^2}\Upsilon_n' 1_n, \frac{\mu_{60} - 3\sigma_0^6}{n\sigma_0^2}\mathrm{tr}(Z_n), \frac{\mu_{50}}{n\sigma_0^2}1_n' R_n X_n, \frac{\mu_{60} - 3\sigma_0^6}{2\sigma_0^4}\Big],
\end{aligned}$$

$\mathrm{var}[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\nu_i^2 - \sigma_0^2)] = 2\sigma_0^4$,

$$\mathrm{E}\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\nu_i^2 - \sigma_0^2) \cdot \frac{1}{\sqrt{n}}\frac{\partial \ln \mathcal{L}_n(\theta_0)}{\partial \theta'}\Big) = \Big[\frac{2\sigma_0^2}{n}\mathrm{tr}(D_n) + \frac{\mu_{30}}{n\sigma_0^2}\Upsilon_n' 1_n, \frac{2\sigma_0^2}{n}\mathrm{tr}(Z_n), \frac{\mu_{30}}{n\sigma_0^2}1_n' R_n X_n, 1\Big],$$

and $\mathbb{B} \equiv \mathrm{var}(\frac{1}{\sqrt{n}} \frac{\partial \ln \mathcal{L}_n(\theta_0)}{\partial \theta})$ in the last subsection can be simplified using $\mu_{40} = 3\sigma_0^4$. Furthermore,

$$\frac{1}{n} \sum_{i=1}^{n} [4\nu_i^3(\tilde{\theta}) - 12\tilde{\sigma}^2 v_i(\tilde{\theta})] \frac{\partial \nu_i(\tilde{\theta})}{\partial \theta'} = \frac{1}{n} \sum_{i=1}^{n} (4\nu_i^3 - 12\sigma_0^2 v_i) \frac{\partial \nu_i(\theta_0)}{\partial \theta'} + o_p(1) = \mathbb{G} + o_p(1),$$

where $\mathbb{G} = -\frac{4\mu_{30}}{n}[1_n' \Upsilon_n, 0, 1_n' R_n X_n, 0]$. Let $\check{\mathbb{H}}$ be an estimate of $\mathbb{H} \equiv \mathrm{var}([\frac{1}{\sqrt{n}} \sum_{i=1}^n (\nu_i^4 - 3\sigma_0^4), \frac{1}{\sqrt{n}} \sum_{i=1}^n (\nu_i^2 - \sigma_0^2), \frac{1}{\sqrt{n}} \frac{\partial \ln \mathcal{L}_n(\theta_0)}{\partial \theta'}]')$ by replacing θ_0, μ_{30}, ..., μ_{80} in $\mathbb{H}$ with, respectively, $\check{\theta}$, $\check{\mu}_3$, ..., $\check{\mu}_8$. Define $\check{\mathbb{G}}$ similarly. Then,

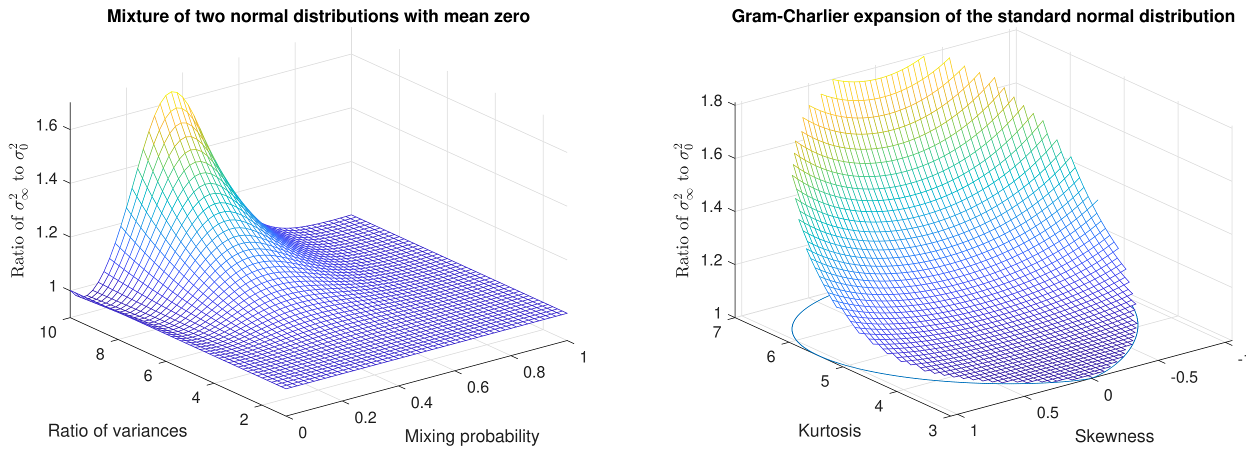
$$\frac{\sqrt{n}(\check{\mu}_4 - 3\check{\sigma}^4)}{\sqrt{[1, -6\check{\sigma}^2, \check{\mathbb{G}}\check{\mathbb{A}}^{-1}]\check{\mathbb{H}}[1, -6\check{\sigma}^2, \check{\mathbb{G}}\check{\mathbb{A}}^{-1}]'}}$$

is asymptotically standard normal under the condition that $\mu_{40} = 3\sigma_0^4$. This statistic can be used to test for excess kurtosis.

3 Ratios of the pseudo-true value of the variance parameter to its true value

Figure S1 reports the ratios of the pseudo-true value of the variance parameter to its true value, where the density function of Student's t distribution is used for the NGPMLE. For the case that the true disturbance distribution is a mixture of two normal distributions with mean zero so that it is symmetric, the pseudo-true value is derived from the maximization of $\mathrm{E}[\ln f(\frac{\sigma_0 v_i}{\sigma}, \eta)] - \ln(\sigma)$ by choosing σ and η; for the case that true disturbance distribution is a Gram-Charlier expansion of the standard normal distribution so that it is asymmetric, the pseudo-true value is derived from the maximization of $\mathrm{E}[\ln f(\frac{\sigma_0 v_i - (1-\rho_0)(\beta_1 - \beta_{10})}{\sigma}, \eta)] - \ln(\sigma)$ by choosing σ, β_1 and η. The latter case corresponds to the situation when the spatial weights matrix M_n for the spatial error dependence process is row-normalized and X_n contains an intercept term, so that the NGPMLE of model parameters except the variance parameter can be consistent. Since maximizing $\mathrm{E}[\ln f(\frac{\sigma_0 v_i - (1-\rho_0)(\beta_1 - \beta_{10})}{\sigma}, \eta)] - \ln(\sigma)$ by choosing σ, β_1 and η is equivalent to maximizing $\mathrm{E}[\ln f(\frac{\sigma_0 v_i - \alpha}{\sigma}, \eta)] - \ln(\sigma)$ by choosing σ, α and η, it is unnecessary to specify the values of ρ_0 and β_{10}. The pseudo-true values are generally different from the true values, even when the true disturbance disturbance distribution is symmetric.

Figure S1: Ratios of the pseudo-true value of the variance parameter to its true value



4 Efficiency comparisons for impact estimators in the SARAR model

Let $\hat{\lambda}$ and $\hat{\beta}_k$ be estimators of the parameters, respectively, λ and β_k in the SARAR model, where β_k is the kth element of β. The average total impact (ATI) computed with $[\hat{\lambda}, \hat{\beta}_k]'$ is $\frac{1}{n} 1_n' S_n^{-1}(\hat{\lambda}) 1_n \hat{\beta}_k$ (LeSage and Pace, 2009). Suppose that $\hat{\varphi}_k \equiv [\hat{\lambda}, \hat{\beta}_k]'$ is asymptotically normal such that $\sqrt{n}(\hat{\varphi}_k - \varphi_{k0}) \xrightarrow{d} N(0, \Delta_k)$. By the mean value theorem,

$$\frac{1}{n} 1_n' S_n^{-1}(\hat{\lambda}) 1_n \hat{\beta}_k = \frac{1}{n} 1_n' S_n^{-1} 1_n \beta_{k0} + \Big[\frac{1}{n} 1_n' S_n^{-1}(\check{\lambda}) W_n S_n^{-1}(\check{\lambda}) 1_n \check{\beta}_k, \frac{1}{n} 1_n' S_n^{-1}(\check{\lambda}) 1_n\Big](\hat{\gamma}_k - \gamma_{k0}),$$

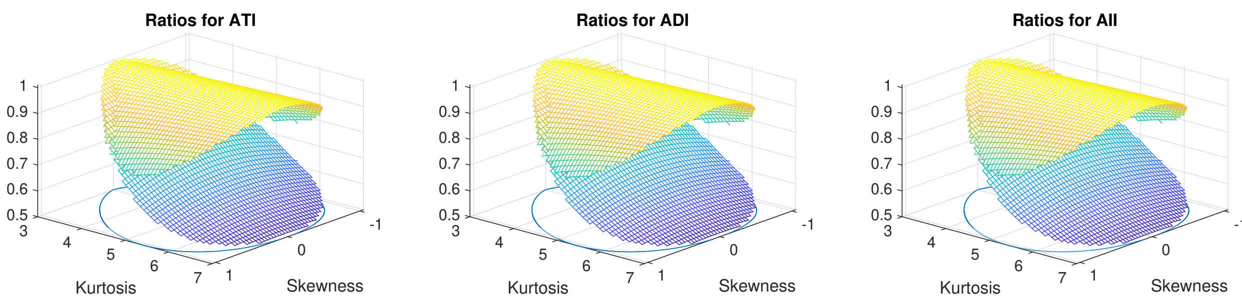
where $\check{\varphi}_k = [\check{\lambda}, \check{\beta}_k]'$ lies between φ_{k0} and $\hat{\varphi}_k$. Then,

$$\begin{aligned}
&\sqrt{n}\Big[\frac{1}{n} 1_n' S_n^{-1}(\hat{\lambda}) 1_n \hat{\beta}_k - \frac{1}{n} 1_n' S_n^{-1} 1_n \beta_{k0}\Big] \\
&= \Big[\frac{1}{n} 1_n' S_n^{-1} W_n S_n^{-1} 1_n \beta_{k0}, \frac{1}{n} 1_n' S_n^{-1} 1_n\Big] \sqrt{n}(\hat{\gamma}_k - \gamma_{k0}) + o_p(1) \\
&\xrightarrow{d} N\bigg(0, \lim_{n\to\infty} \Big[\frac{1}{n} 1_n' S_n^{-1} W_n S_n^{-1} 1_n \beta_{k0}, \frac{1}{n} 1_n' S_n^{-1} 1_n\Big] \Delta_k \Big[\frac{1}{n} 1_n' S_n^{-1} W_n S_n^{-1} 1_n \beta_{k0}, \frac{1}{n} 1_n' S_n^{-1} 1_n\Big]'\bigg).
\end{aligned}$$

Thus, if the NGPMLE is asymptotically more efficient than other estimators, so is the ATI computed with the NGPMLE than the ATIs computed with other estimators. Similar analysis applies to the average direct impacts (ADI) and the average indirect impacts (AII). With $\hat{\varphi}_k$, the ADI is $\frac{1}{n}\operatorname{tr}[S_n^{-1}(\hat{\lambda})]\hat{\beta}_k$, and the AII is $\{\frac{1}{n} 1_n' S_n^{-1}(\hat{\lambda}) 1_n - \frac{1}{n}\operatorname{tr}[S_n^{-1}(\hat{\lambda})]\}\hat{\beta}_k$.

Figures S2–S4 compare the efficiencies of impact estimators computed with the NGPMLE, GPMLE and BGMME using numerical integration, as in the main text. The patterns are similar to those for the efficiency comparisons between the NGPMLE, GPMLE and BG-

Figure S2: Efficiency comparisons of different impact 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 sub-figure shows the ratios of the asymptotic variance of an impact estimator computed with NGPMLE$_o$ to that computed with GPMLE, while the upper mesh shows the ratios of the asymptotic variance of the impact estimator computed with BGMME to that computed with GPMLE.



MME. The original NGPMLE has uniform efficiency improvement over the GPMLE, and has significantly larger efficiency improvement than the BGMME in most cases.

Tables S1–S2 further report Monte Carlo results on impact estimators computed with different estimators. The patterns are consistent with those in Figures S2–S4.

5 More Monte Carlo results

5.1 SAR model

Tables S3–S5 report more Monte Carlo results for the estimation of the SAR model, where the true disturbance distribution is a mixture of two normal distributions with mean zero. Each table corresponds to a different ratio of variances (RV) of the two normal distributions. The patterns are similar to those in the main text.

5.2 Bias of the NGPMLE when Assumption 4 is not satisfied

We investigate the possible bias of the NGPMLE when Assumption 4 is not satisfied. We consider two cases: (i) M_n is not row-normalized and v_i is asymmetric; and (ii) M_n is row-normalized, X_n does not contain an intercept term and v_i is asymmetric. For case (ii), X_n still contains 2 exogenous variables, but both are randomly drawn from the standard normal distribution. Other settings are the same as in the main text.

Tables S6–S7 report the results. We observe that NGPMLE has similar bias as GPMLE for different sets of parameters and different sample sizes. Then it is possible that the

Figure S3: Efficiency comparisons of different impact 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 sub-figures, the lower mesh in each sub-figure shows the ratios of the asymptotic variance of an impact estimator computed with $\mathrm{NGPMLE_o}$ to that computed with GPMLE, while the upper mesh shows the ratios of the asymptotic variance of the impact estimator computed with BGMME to that computed with GPMLE. For the fourth to sixth sub-figures, the mesh in each sub-figure shows the ratios of the asymptotic variance of an impact estimator computed with $\mathrm{NGPMLE_a}$ to that computed with GPMLE.

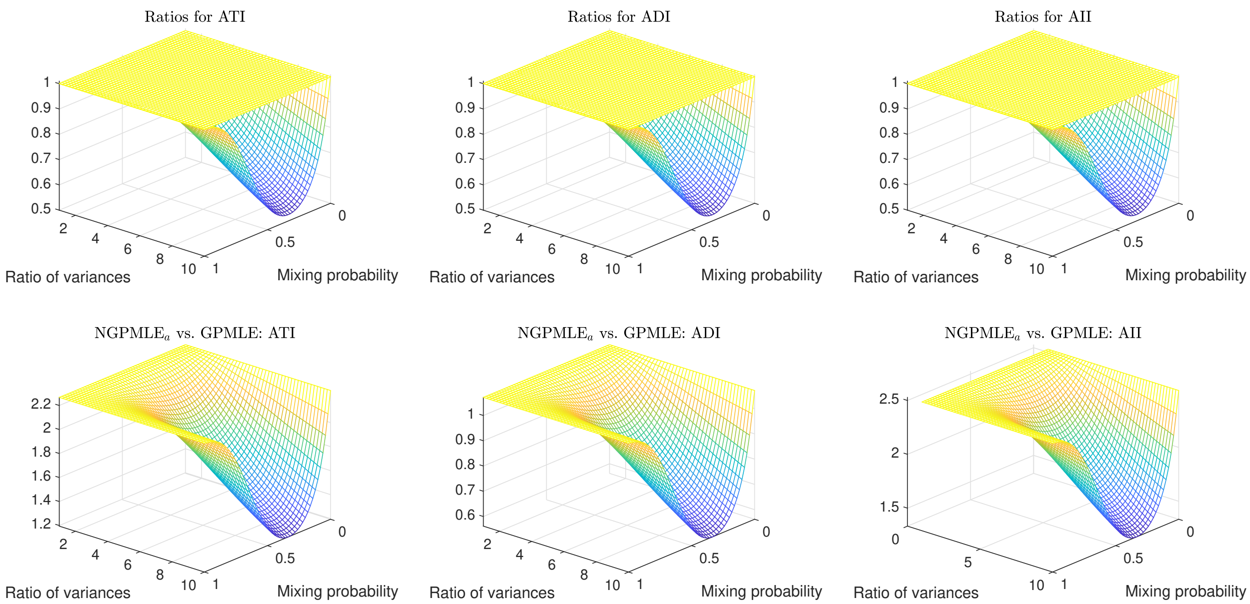


Figure S4: Efficiency comparisons of different impact estimators for the SARAR model with a non-row-normalized M_n and asymmetric innovations. The v_i is an admissible fourth order Gram-Charlier expansion of the standard normal distribution as a function of the skewness and kurtosis coefficients. The mesh in each sub-figure shows the ratios of the asymptotic variance of an impact estimator computed with $\mathrm{NGPMLE_a}$ to that computed with GPMLE.

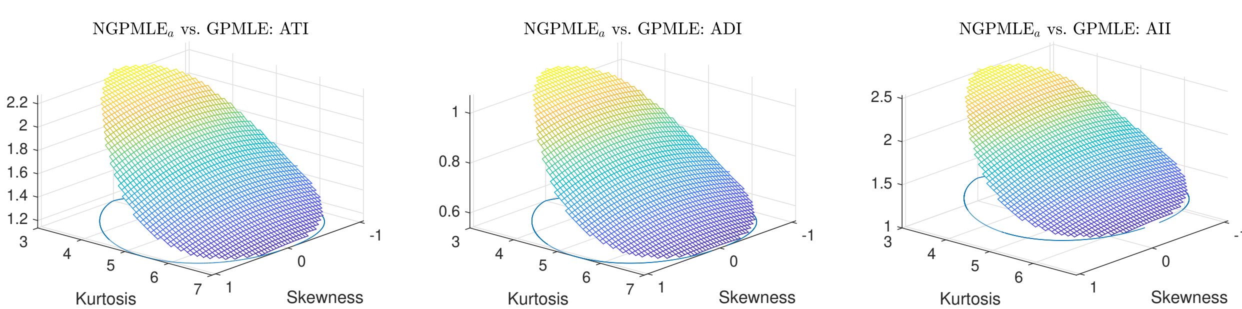


Table S1: Performance of various impact estimators for the SARAR model with a row-normalized M_n and asymmetric v_i

			ATI			ADI			AII		
Kurtosis	Skewness		Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
		Panel A: $n = 147$									
6	0.8	GPMLE	0.003	0.220	0.220	0.000	0.048	0.048	0.003	0.190	0.190
		BGMME	0.007	0.212	0.213	0.001	0.046	0.046	0.006	0.183	0.184
		NGPMLE	-0.002	0.173	0.173	-0.001	0.037	0.037	-0.001	0.149	0.149
6	0.05	GPMLE	0.006	0.221	0.221	0.000	0.048	0.048	0.006	0.191	0.191
		BGMME	0.010	0.229	0.229	0.000	0.049	0.049	0.010	0.198	0.198
		NGPMLE	0.002	0.174	0.174	-0.001	0.038	0.038	0.002	0.151	0.151
4	0.4	GPMLE	0.005	0.219	0.219	-0.001	0.048	0.048	0.006	0.190	0.190
		BGMME	0.010	0.223	0.224	0.000	0.048	0.048	0.011	0.194	0.195
		NGPMLE	0.007	0.213	0.213	-0.001	0.046	0.046	0.007	0.185	0.185
4	0.05	GPMLE	0.003	0.224	0.224	0.000	0.048	0.048	0.003	0.194	0.194
		BGMME	0.008	0.232	0.232	0.000	0.050	0.050	0.008	0.201	0.201
		NGPMLE	0.001	0.216	0.216	0.000	0.047	0.047	0.002	0.187	0.187
3.05	0.05	GPMLE	0.001	0.221	0.221	-0.001	0.048	0.048	0.002	0.191	0.191
		BGMME	0.006	0.231	0.231	0.000	0.049	0.049	0.007	0.200	0.200
		NGPMLE	0.000	0.226	0.226	-0.001	0.049	0.049	0.001	0.195	0.195
		Panel B: $n = 294$									
6	0.8	GPMLE	0.001	0.153	0.153	-0.001	0.034	0.034	0.002	0.132	0.132
		BGMME	0.002	0.144	0.145	-0.001	0.032	0.032	0.002	0.125	0.125
		NGPMLE	0.001	0.118	0.118	0.000	0.026	0.026	0.001	0.102	0.102
6	0.05	GPMLE	0.003	0.153	0.153	-0.001	0.033	0.033	0.004	0.133	0.133
		BGMME	0.005	0.156	0.156	0.000	0.034	0.034	0.006	0.135	0.135
		NGPMLE	0.004	0.120	0.120	0.000	0.026	0.026	0.004	0.104	0.104
4	0.4	GPMLE	0.002	0.151	0.151	0.000	0.033	0.033	0.002	0.131	0.131
		BGMME	0.004	0.150	0.150	0.000	0.033	0.033	0.004	0.130	0.130
		NGPMLE	0.003	0.145	0.145	0.000	0.032	0.032	0.003	0.126	0.126
4	0.05	GPMLE	0.001	0.149	0.149	0.000	0.033	0.033	0.001	0.129	0.129
		BGMME	0.003	0.153	0.153	0.000	0.033	0.033	0.003	0.132	0.132
		NGPMLE	0.001	0.144	0.144	0.000	0.032	0.032	0.001	0.125	0.125
3.05	0.05	GPMLE	0.001	0.151	0.151	0.000	0.033	0.033	0.001	0.131	0.131
		BGMME	0.003	0.155	0.155	0.000	0.034	0.034	0.003	0.135	0.135
		NGPMLE	0.000	0.154	0.154	0.000	0.034	0.034	0.000	0.134	0.134

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 non-intercept variable in X_n. $\lambda_0 = 0.4$, $\rho_0 = 0.2$, $\beta_{10} = 1$, $\beta_{20} = 1$ and $\sigma_0^2 = 0.25$.

Table S2: Performance of various impact estimators for the SARAR model with a non-row-normalized M_n and symmetric v_i

RV		ATI			ADI			AII		
		Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
	Panel A: $n = 147$									
9	GPMLE	0.003	0.149	0.149	-0.001	0.044	0.044	0.004	0.125	0.125
	BGMME	0.005	0.154	0.154	-0.001	0.045	0.045	0.006	0.131	0.131
	NGPMLE$_o$	0.000	0.113	0.113	-0.001	0.034	0.034	0.001	0.096	0.096
	NGPMLE$_a$	0.003	0.174	0.174	-0.001	0.035	0.035	0.004	0.154	0.154
6	GPMLE	0.004	0.147	0.147	0.000	0.044	0.044	0.004	0.124	0.124
	BGMME	0.005	0.155	0.155	-0.001	0.045	0.045	0.005	0.132	0.132
	NGPMLE$_o$	0.003	0.127	0.127	0.000	0.038	0.038	0.003	0.107	0.107
	NGPMLE$_a$	0.003	0.196	0.196	-0.001	0.040	0.040	0.004	0.175	0.175
3	GPMLE	0.006	0.148	0.148	-0.001	0.044	0.044	0.007	0.126	0.126
	BGMME	0.008	0.156	0.157	-0.001	0.046	0.046	0.009	0.133	0.133
	NGPMLE$_o$	0.004	0.148	0.148	-0.001	0.043	0.043	0.006	0.126	0.126
	NGPMLE$_a$	0.107	4.895	4.896	0.002	0.162	0.162	0.105	4.739	4.740
1.1	GPMLE	0.002	0.147	0.147	0.000	0.044	0.044	0.002	0.125	0.125
	BGMME	0.004	0.155	0.155	0.000	0.046	0.046	0.005	0.132	0.132
	NGPMLE$_o$	0.001	0.151	0.151	0.000	0.045	0.045	0.001	0.128	0.128
	NGPMLE$_a$	0.017	0.245	0.246	0.000	0.051	0.051	0.017	0.216	0.217
	Panel B: $n = 294$									
9	GPMLE	0.002	0.101	0.101	0.000	0.031	0.031	0.002	0.086	0.086
	BGMME	0.003	0.103	0.103	0.000	0.031	0.031	0.003	0.088	0.088
	NGPMLE$_o$	0.002	0.078	0.078	0.000	0.024	0.024	0.002	0.066	0.066
	NGPMLE$_a$	-0.001	0.119	0.119	-0.001	0.025	0.025	0.000	0.106	0.106
6	GPMLE	0.002	0.103	0.103	0.000	0.031	0.031	0.002	0.087	0.087
	BGMME	0.003	0.105	0.105	0.000	0.032	0.032	0.003	0.089	0.089
	NGPMLE$_o$	0.001	0.087	0.087	0.000	0.027	0.027	0.001	0.073	0.073
	NGPMLE$_a$	0.000	0.132	0.132	0.000	0.028	0.028	0.000	0.118	0.118
3	GPMLE	0.001	0.102	0.102	0.000	0.031	0.031	0.001	0.086	0.086
	BGMME	0.002	0.104	0.104	0.000	0.031	0.031	0.002	0.088	0.088
	NGPMLE$_o$	0.001	0.099	0.099	0.000	0.030	0.030	0.001	0.083	0.083
	NGPMLE$_a$	0.003	0.188	0.188	0.000	0.033	0.033	0.003	0.169	0.169
1.1	GPMLE	0.003	0.101	0.101	0.000	0.030	0.030	0.003	0.085	0.085
	BGMME	0.004	0.104	0.104	0.000	0.031	0.031	0.004	0.088	0.088
	NGPMLE$_o$	0.002	0.104	0.104	0.000	0.031	0.031	0.002	0.089	0.089
	NGPMLE$_a$	0.166	5.372	5.375	0.005	0.170	0.170	0.161	5.205	5.208
	Panel C: Normal innovations, $n = 147$									
	GPMLE	0.004	0.149	0.149	0.000	0.045	0.045	0.004	0.126	0.126
	BGMME	0.006	0.157	0.157	0.000	0.046	0.046	0.006	0.133	0.133
	NGPMLE$_o$	0.003	0.171	0.171	0.000	0.046	0.046	0.004	0.146	0.146
	NGPMLE$_a$	0.163	6.776	6.778	0.005	0.239	0.239	0.158	6.542	6.544
	Panel D: Normal innovations, $n = 294$									
	GPMLE	0.003	0.102	0.102	0.000	0.031	0.031	0.003	0.086	0.087
	BGMME	0.005	0.105	0.105	0.000	0.031	0.031	0.004	0.088	0.089
	NGPMLE$_o$	0.003	0.104	0.104	0.000	0.031	0.031	0.003	0.088	0.088
	NGPMLE$_a$	0.114	5.132	5.133	0.004	0.165	0.165	0.110	4.971	4.972

Notes: 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 non-intercept variable in X_n. $\lambda_0 = 0.4$, $\rho_0 = 0.2$, $\beta_{10} = 1$, $\beta_{20} = 1$ and $\sigma_0^2 = 0.25$.

Table S3: Performance of various estimators for the SAR model with symmetric v_i (RV = 6)

	λ			β_2		
	Bias	SD	RMSE	Bias	SD	RMSE
GPMLE	-0.012	0.055	0.056	0.001	0.042	0.042
BGMME	-0.010	0.055	0.056	-0.001	0.042	0.042
NGPMLE	-0.009	0.048	0.049	0.001	0.036	0.036
AEs with GPMLE as the initial estimate						
$\text{AE}_a(p,1)$	0.099	0.067	0.119	-0.009	0.043	0.044
$\text{AE}_a(b,1)$	0.083	0.059	0.101	-0.008	0.038	0.039
$\text{AE}_b(p,1)$	0.324	0.105	0.341	-0.031	0.049	0.058
$\text{AE}_b(b,1)$	0.257	0.090	0.272	-0.024	0.042	0.049
$\text{AE}_a(p,2)$	0.094	0.068	0.116	-0.009	0.044	0.045
$\text{AE}_a(b,2)$	0.080	0.061	0.100	-0.008	0.040	0.040
$\text{AE}_b(p,2)$	0.308	0.104	0.325	-0.030	0.050	0.058
$\text{AE}_b(b,2)$	0.249	0.090	0.265	-0.024	0.044	0.050
$\text{AE}_a(p,4)$	0.081	0.065	0.104	-0.008	0.043	0.044
$\text{AE}_a(b,4)$	0.070	0.067	0.097	-0.007	0.044	0.045
$\text{AE}_b(p,4)$	0.260	0.095	0.277	-0.025	0.047	0.053
$\text{AE}_b(b,4)$	0.224	0.094	0.243	-0.022	0.047	0.052
AEs with OLSE as the initial estimate						
$\text{AE}_a(p,1)$	0.043	0.060	0.074	-0.004	0.042	0.042
$\text{AE}_a(b,1)$	0.027	0.053	0.060	-0.003	0.038	0.038
$\text{AE}_b(p,1)$	0.315	0.110	0.334	-0.030	0.049	0.057
$\text{AE}_b(b,1)$	0.238	0.092	0.255	-0.022	0.042	0.048
$\text{AE}_a(p,2)$	0.040	0.061	0.073	-0.004	0.043	0.044
$\text{AE}_a(b,2)$	0.026	0.056	0.061	-0.002	0.039	0.039
$\text{AE}_b(p,2)$	0.299	0.109	0.318	-0.028	0.050	0.057
$\text{AE}_b(b,2)$	0.231	0.092	0.249	-0.022	0.043	0.048
$\text{AE}_a(p,4)$	0.026	0.060	0.066	-0.003	0.042	0.043
$\text{AE}_a(b,4)$	0.024	0.063	0.068	-0.003	0.044	0.044
$\text{AE}_b(p,4)$	0.243	0.097	0.262	-0.023	0.047	0.052
$\text{AE}_b(b,4)$	0.212	0.095	0.232	-0.020	0.047	0.051

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 6, and the mixing probability is 0.3. β_2 is the coefficient on the non-intercept variable in X_n. $\lambda_0 = 0.4$, $\rho_0 = 0.2$, $\beta_{10} = 1$, $\beta_{20} = 1$ and $\sigma_0^2 = 0.25$.

Table S4: Performance of various estimators for the SAR model with symmetric v_i (RV = 3)

	λ			β_2		
	Bias	SD	RMSE	Bias	SD	RMSE
GPMLE	-0.012	0.054	0.055	0.002	0.042	0.042
BGMME	-0.010	0.055	0.056	0.000	0.043	0.043
NGPMLE	-0.011	0.054	0.055	0.002	0.042	0.042
AEs with GPMLE as the initial estimate						
$\text{AE}_a(p,1)$	0.098	0.066	0.118	-0.009	0.043	0.044
$\text{AE}_a(b,1)$	0.092	0.068	0.114	-0.008	0.045	0.046
$\text{AE}_b(p,1)$	0.321	0.103	0.337	-0.031	0.049	0.058
$\text{AE}_b(b,1)$	0.305	0.104	0.322	-0.029	0.050	0.058
$\text{AE}_a(p,2)$	0.095	0.068	0.116	-0.008	0.045	0.046
$\text{AE}_a(b,2)$	0.089	0.069	0.113	-0.007	0.047	0.047
$\text{AE}_b(p,2)$	0.310	0.103	0.327	-0.030	0.050	0.058
$\text{AE}_b(b,2)$	0.297	0.104	0.315	-0.028	0.052	0.059
$\text{AE}_a(p,4)$	0.089	0.069	0.113	-0.008	0.047	0.047
$\text{AE}_a(b,4)$	0.082	0.073	0.110	-0.007	0.050	0.050
$\text{AE}_b(p,4)$	0.290	0.102	0.308	-0.028	0.051	0.058
$\text{AE}_b(b,4)$	0.268	0.104	0.287	-0.025	0.054	0.059
AEs with OLSE as the initial estimate						
$\text{AE}_a(p,1)$	0.043	0.059	0.073	-0.003	0.042	0.043
$\text{AE}_a(b,1)$	0.036	0.062	0.072	-0.002	0.045	0.045
$\text{AE}_b(p,1)$	0.311	0.108	0.329	-0.029	0.049	0.057
$\text{AE}_b(b,1)$	0.293	0.108	0.313	-0.027	0.050	0.057
$\text{AE}_a(p,2)$	0.041	0.061	0.074	-0.003	0.044	0.044
$\text{AE}_a(b,2)$	0.036	0.064	0.073	-0.002	0.046	0.046
$\text{AE}_b(p,2)$	0.301	0.108	0.320	-0.028	0.050	0.058
$\text{AE}_b(b,2)$	0.287	0.108	0.307	-0.026	0.052	0.058
$\text{AE}_a(p,4)$	0.038	0.064	0.075	-0.003	0.046	0.046
$\text{AE}_a(b,4)$	0.038	0.069	0.079	-0.002	0.049	0.050
$\text{AE}_b(p,4)$	0.282	0.106	0.301	-0.027	0.051	0.058
$\text{AE}_b(b,4)$	0.262	0.107	0.283	-0.024	0.054	0.059

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 3, and the mixing probability is 0.3. β_2 is the coefficient on the non-intercept variable in X_n. $\lambda_0 = 0.4$, $\rho_0 = 0.2$, $\beta_{10} = 1$, $\beta_{20} = 1$ and $\sigma_0^2 = 0.25$.

Table S5: Performance of various estimators for the SAR model with symmetric v_i (RV = 1.1)

	λ			β_2		
	Bias	SD	RMSE	Bias	SD	RMSE
GPMLE	-0.012	0.054	0.055	0.001	0.042	0.042
BGMME	-0.010	0.055	0.056	-0.001	0.043	0.043
NGPMLE	-0.012	0.054	0.056	0.001	0.043	0.043
AEs with GPMLE as the initial estimate						
$\mathrm{AE}_a(p, 1)$	0.099	0.065	0.118	-0.010	0.044	0.045
$\mathrm{AE}_a(b, 1)$	0.099	0.072	0.122	-0.010	0.049	0.050
$\mathrm{AE}_b(p, 1)$	0.322	0.101	0.338	-0.031	0.050	0.059
$\mathrm{AE}_b(b, 1)$	0.342	0.110	0.359	-0.033	0.055	0.065
$\mathrm{AE}_a(p, 2)$	0.097	0.066	0.117	-0.010	0.045	0.046
$\mathrm{AE}_a(b, 2)$	0.096	0.074	0.121	-0.010	0.051	0.051
$\mathrm{AE}_b(p, 2)$	0.316	0.101	0.332	-0.031	0.051	0.059
$\mathrm{AE}_b(b, 2)$	0.333	0.110	0.351	-0.032	0.056	0.065
$\mathrm{AE}_a(p, 4)$	0.091	0.072	0.115	-0.009	0.049	0.050
$\mathrm{AE}_a(b, 4)$	0.086	0.075	0.114	-0.009	0.051	0.052
$\mathrm{AE}_b(p, 4)$	0.296	0.104	0.313	-0.029	0.054	0.061
$\mathrm{AE}_b(b, 4)$	0.286	0.107	0.306	-0.028	0.056	0.063
AEs with OLSE as the initial estimate						
$\mathrm{AE}_a(p, 1)$	0.043	0.058	0.073	-0.004	0.043	0.043
$\mathrm{AE}_a(b, 1)$	0.042	0.066	0.078	-0.004	0.048	0.049
$\mathrm{AE}_b(p, 1)$	0.313	0.105	0.330	-0.030	0.050	0.058
$\mathrm{AE}_b(b, 1)$	0.336	0.115	0.355	-0.032	0.055	0.064
$\mathrm{AE}_a(p, 2)$	0.043	0.060	0.074	-0.004	0.044	0.044
$\mathrm{AE}_a(b, 2)$	0.042	0.069	0.081	-0.004	0.050	0.050
$\mathrm{AE}_b(p, 2)$	0.308	0.105	0.325	-0.030	0.051	0.059
$\mathrm{AE}_b(b, 2)$	0.329	0.115	0.348	-0.031	0.056	0.065
$\mathrm{AE}_a(p, 4)$	0.044	0.067	0.080	-0.005	0.049	0.049
$\mathrm{AE}_a(b, 4)$	0.043	0.070	0.083	-0.005	0.051	0.051
$\mathrm{AE}_b(p, 4)$	0.292	0.107	0.311	-0.028	0.054	0.061
$\mathrm{AE}_b(b, 4)$	0.285	0.110	0.306	-0.028	0.056	0.062

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 1.1, and the mixing probability is 0.3. β_2 is the coefficient on the non-intercept variable in X_n. $\lambda_0 = 0.4$, $\rho_0 = 0.2$, $\beta_{10} = 1$, $\beta_{20} = 1$ and $\sigma_0^2 = 0.25$.

NGPMLE for the spatial dependence parameters and coefficients on non-intercept exogenous variables can still be consistent when Assumption 4 is not satisfied. The SD of NGPMLE is significantly smaller than that of GPMLE when the kurtosis coefficient is 6, and it is the same as or slightly larger than that of GPMLE for smaller kurtosis coefficients.

6 More application results

In this section, we first report impact estimates for the application in the main text on the hedonic pricing data, and then report application results on two more data sets.

6.1 Impact estimates for the hedonic pricing data

Table S8 reports impact estimates for the hedonic pricing data. For some variables, we observe relatively large differences in the impact estimates computed with different parameter estimates. We also observe differences in impact significance. In particular, in the main text, the GPMLEs, BGMMEs and NGPMLEs of the coefficients on NOX^2 and AGE are observed to have different significance results. Here we also observe that the impacts for NOX^2 and AGE computed with different parameter estimates have different significance results. For example, for the variable AGE, the average total, direct and indirect impacts computed with GPMLE are all insignificant at any usual significance level, but these impacts computed with NGPMLE are significant at the 1% or 5% level.

6.2 Application to the crime data in Anselin (1988)

In this subsection, we apply our NGPMLE to the crime data for 49 neighborhoods in Columbus, Ohio in Anselin (1988). This data set has been used in, e.g., LeSage (1999a,b), LeSage and Pace (2009), and Arbia (2014).

We first estimate an SARAR model, where the dependent variable is Crime, defined as the total of residential burglaries and vehicle thefts per thousand households, and the explanatory variables include an intercept term, income in thousand dollars (Income) and house value in thousand dollars (House_value). The spatial weights matrix W_n for the spatial lag process is based on first order contiguity and is row-normalized. The spatial weights matrix M_n for the spatial error process is set to be the same as W_n. Table S9 reports the results of estimation and some diagnostic tests. We observe that the normality of innovations is rejected at the 1% level. At the 10% level, while the skewness coefficient of innovations being zero is not rejected, the excess kurtosis coefficient being zero is rejected, with a p-value equal to 0.064. The GPMLEs, BGMMEs and NGPMLEs of model parameters are

Table S6: Performance of various estimators for the SARAR model with a non-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	-0.004	0.063	0.063	-0.044	0.169	0.175	0.000	0.042	0.042
		BGMME	-0.002	0.063	0.064	-0.027	0.182	0.184	0.000	0.040	0.040
		NGPMLE$_o$	-0.007	0.050	0.051	-0.028	0.148	0.151	0.000	0.032	0.032
6	0.05	GPMLE	-0.004	0.064	0.064	-0.047	0.170	0.176	-0.001	0.042	0.042
		BGMME	-0.004	0.069	0.069	-0.023	0.185	0.187	-0.001	0.044	0.044
		NGPMLE$_o$	-0.003	0.050	0.050	-0.039	0.147	0.152	-0.001	0.033	0.033
4	0.4	GPMLE	-0.004	0.065	0.065	-0.045	0.172	0.178	-0.001	0.042	0.042
		BGMME	-0.002	0.067	0.067	-0.025	0.186	0.188	-0.001	0.042	0.042
		NGPMLE$_o$	-0.005	0.069	0.069	-0.043	0.176	0.181	-0.001	0.044	0.044
4	0.05	GPMLE	-0.004	0.064	0.065	-0.043	0.169	0.174	-0.001	0.042	0.042
		BGMME	-0.004	0.070	0.070	-0.018	0.184	0.185	-0.001	0.044	0.044
		NGPMLE$_o$	-0.005	0.065	0.065	-0.042	0.171	0.176	-0.001	0.043	0.043
3.05	0.05	GPMLE	-0.005	0.064	0.064	-0.044	0.172	0.177	-0.001	0.042	0.042
		BGMME	-0.005	0.071	0.071	-0.017	0.187	0.188	-0.001	0.043	0.043
		NGPMLE$_o$	-0.006	0.066	0.067	-0.043	0.178	0.183	-0.001	0.042	0.042
		Panel B: $n = 294$									
6	0.8	GPMLE	-0.002	0.044	0.044	-0.021	0.116	0.118	0.000	0.030	0.030
		BGMME	-0.001	0.042	0.042	-0.014	0.118	0.119	0.000	0.028	0.028
		NGPMLE$_o$	-0.006	0.035	0.036	-0.011	0.103	0.104	0.000	0.023	0.023
6	0.05	GPMLE	-0.001	0.044	0.044	-0.022	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$_o$	-0.001	0.034	0.034	-0.018	0.100	0.101	0.000	0.023	0.023
4	0.4	GPMLE	-0.002	0.044	0.044	-0.021	0.115	0.117	0.000	0.030	0.030
		BGMME	0.000	0.043	0.043	-0.013	0.118	0.119	0.000	0.029	0.029
		NGPMLE$_o$	-0.002	0.045	0.045	-0.019	0.117	0.118	0.000	0.029	0.029
4	0.05	GPMLE	-0.002	0.043	0.043	-0.020	0.116	0.118	0.000	0.029	0.029
		BGMME	-0.001	0.044	0.044	-0.008	0.119	0.119	0.000	0.030	0.030
		NGPMLE$_o$	-0.002	0.044	0.044	-0.020	0.116	0.118	0.000	0.029	0.029
3.05	0.05	GPMLE	-0.003	0.044	0.044	-0.019	0.119	0.121	0.000	0.030	0.030
		BGMME	-0.002	0.045	0.045	-0.006	0.123	0.123	0.000	0.030	0.030
		NGPMLE$_o$	-0.003	0.045	0.045	-0.019	0.120	0.122	0.000	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 non-intercept variable in X_n. $\lambda_0 = 0.4$, $\rho_0 = 0.2$, $\beta_{10} = 1$, $\beta_{20} = 1$ and $\sigma_0^2 = 0.25$.

Table S7: Performance of various estimators for the SARAR model with a row-normalized M_n, no intercept term 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	-0.003	0.053	0.053	-0.013	0.128	0.129	-0.001	0.042	0.042
		BGMME	-0.002	0.051	0.052	0.000	0.133	0.133	0.000	0.039	0.039
		NGPMLE$_o$	-0.003	0.043	0.043	-0.010	0.113	0.114	-0.001	0.033	0.033
6	0.05	GPMLE	-0.004	0.052	0.053	-0.012	0.128	0.128	-0.001	0.042	0.042
		BGMME	-0.004	0.055	0.055	0.002	0.134	0.134	-0.001	0.043	0.043
		NGPMLE$_o$	-0.003	0.041	0.041	-0.009	0.109	0.109	0.000	0.033	0.033
4	0.4	GPMLE	-0.003	0.054	0.054	-0.014	0.130	0.130	-0.001	0.042	0.042
		BGMME	-0.003	0.054	0.054	0.000	0.135	0.135	-0.001	0.042	0.042
		NGPMLE$_o$	-0.004	0.056	0.056	-0.013	0.130	0.131	-0.001	0.043	0.043
4	0.05	GPMLE	-0.002	0.053	0.053	-0.017	0.127	0.128	-0.001	0.043	0.043
		BGMME	-0.002	0.056	0.056	-0.003	0.133	0.133	0.000	0.044	0.044
		NGPMLE$_o$	-0.002	0.054	0.054	-0.017	0.129	0.130	-0.001	0.043	0.043
3.05	0.05	GPMLE	-0.003	0.053	0.053	-0.012	0.126	0.127	-0.001	0.042	0.042
		BGMME	-0.002	0.055	0.055	0.002	0.132	0.132	-0.001	0.043	0.043
		NGPMLE$_o$	-0.003	0.055	0.055	-0.012	0.127	0.128	-0.001	0.042	0.042
		Panel B: $n = 294$									
6	0.8	GPMLE	-0.001	0.036	0.036	-0.007	0.087	0.087	-0.001	0.030	0.030
		BGMME	0.000	0.034	0.034	-0.001	0.088	0.088	-0.001	0.028	0.028
		NGPMLE$_o$	-0.001	0.030	0.030	-0.005	0.076	0.076	-0.001	0.023	0.023
6	0.05	GPMLE	-0.003	0.037	0.037	-0.006	0.089	0.090	0.000	0.029	0.029
		BGMME	-0.002	0.038	0.038	0.002	0.091	0.091	0.000	0.030	0.030
		NGPMLE$_o$	-0.001	0.029	0.029	-0.005	0.075	0.075	0.000	0.022	0.022
4	0.4	GPMLE	-0.001	0.037	0.037	-0.005	0.089	0.089	0.000	0.029	0.029
		BGMME	-0.001	0.037	0.037	0.002	0.090	0.090	0.000	0.029	0.029
		NGPMLE$_o$	-0.001	0.038	0.038	-0.004	0.089	0.089	0.000	0.031	0.031
4	0.05	GPMLE	-0.002	0.037	0.037	-0.007	0.089	0.090	-0.001	0.029	0.029
		BGMME	-0.002	0.038	0.038	0.000	0.091	0.091	-0.001	0.030	0.030
		NGPMLE$_o$	-0.002	0.037	0.037	-0.007	0.089	0.089	-0.001	0.029	0.029
3.05	0.05	GPMLE	-0.002	0.037	0.037	-0.007	0.090	0.091	-0.001	0.029	0.029
		BGMME	-0.002	0.038	0.038	0.001	0.092	0.092	-0.001	0.030	0.030
		NGPMLE$_o$	-0.002	0.037	0.037	-0.007	0.091	0.092	-0.001	0.034	0.034

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 second exogenous variable in X_n. $\lambda_0 = 0.4$, $\rho_0 = 0.2$, $\beta_{10} = 1$, $\beta_{20} = 1$ and $\sigma_0^2 = 0.25$.

Table S8: Impact estimates for the hedonic pricing data

		Average total impacts		Average direct impacts		Average indirect impacts	
		estimate	SE	estimate	SE	estimate	SE
CRIM	GPMLE	−0.230***	0.032	−0.188***	0.023	−0.042**	0.017
	BGMME	−0.241***	0.035	−0.179***	0.023	−0.062***	0.018
	NGPMLE	−0.188***	0.019	−0.166***	0.015	−0.022**	0.009
ZN	GPMLE	0.080**	0.039	0.065**	0.031	0.015	0.009
	BGMME	0.086**	0.043	0.064**	0.031	0.022*	0.013
	NGPMLE	0.052**	0.024	0.046**	0.021	0.006	0.004
INDUS	GPMLE	0.020	0.056	0.016	0.046	0.004	0.011
	BGMME	−0.001	0.062	−0.001	0.046	−0.000	0.016
	NGPMLE	0.001	0.035	0.001	0.031	0.000	0.004
CHAS	GPMLE	−0.008	0.026	−0.007	0.021	−0.002	0.005
	BGMME	−0.014	0.028	−0.010	0.021	−0.004	0.007
	NGPMLE	−0.016	0.016	−0.014	0.014	−0.002	0.002
NOX^2	GPMLE	−0.235***	0.068	−0.192***	0.055	−0.043**	0.019
	BGMME	−0.423***	0.075	−0.314***	0.056	−0.109***	0.032
	NGPMLE	−0.081*	0.043	−0.071*	0.038	−0.010	0.006
RM^2	GPMLE	0.246***	0.036	0.201***	0.024	0.045**	0.018
	BGMME	0.263***	0.040	0.195***	0.025	0.068***	0.021
	NGPMLE	0.472***	0.032	0.416***	0.016	0.056**	0.023
AGE	GPMLE	−0.057	0.045	−0.046	0.037	−0.010	0.009
	BGMME	−0.101**	0.050	−0.075**	0.037	−0.026*	0.014
	NGPMLE	−0.183***	0.029	−0.162***	0.025	−0.022**	0.009
DIS	GPMLE	−0.315***	0.074	−0.258***	0.056	−0.058**	0.026
	BGMME	−0.282***	0.083	−0.210***	0.056	−0.073**	0.031
	NGPMLE	−0.196***	0.047	−0.173***	0.039	−0.023**	0.012
RAD	GPMLE	0.421***	0.081	0.344***	0.061	0.077**	0.033
	BGMME	0.528***	0.090	0.392***	0.061	0.136***	0.042
	NGPMLE	0.230***	0.049	0.202***	0.042	0.027**	0.013
TAX	GPMLE	−0.319***	0.073	−0.261***	0.057	−0.058**	0.026
	BGMME	−0.320***	0.081	−0.238***	0.058	−0.082***	0.030
	NGPMLE	−0.243***	0.045	−0.214***	0.038	−0.029**	0.013
PTRATIO	GPMLE	−0.156***	0.038	−0.128***	0.030	−0.029**	0.012
	BGMME	−0.140***	0.042	−0.104***	0.031	−0.036***	0.014
	NGPMLE	−0.089***	0.023	−0.079***	0.021	−0.011**	0.005
B	GPMLE	0.147***	0.033	0.120***	0.026	0.027**	0.011
	BGMME	0.179***	0.037	0.133***	0.026	0.046***	0.015
	NGPMLE	0.173***	0.021	0.153***	0.018	0.021**	0.008
LSTAT	GPMLE	−0.465***	0.048	−0.380***	0.035	−0.085***	0.032
	BGMME	−0.509***	0.054	−0.378***	0.035	−0.131***	0.035
	NGPMLE	−0.176***	0.026	−0.155***	0.023	−0.021**	0.008

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

Table S9: Results of estimation and diagnostic tests based on the SARAR model for the crime data

	GPMLE		BGMME		NGPMLE	
	estimate	SE	estimate	SE	estimate	SE
λ	0.368^{*}	0.189	-0.364	0.293	0.444^{***}	0.128
ρ	0.167	0.294	0.826^{***}	0.103	0.110	0.239
Constant	47.784^{***}	9.553	62.039^{***}	11.088	46.420^{***}	7.525
Income	-1.026^{***}	0.324	-0.775^{**}	0.317	-1.645^{***}	0.279
House_value	-0.282^{***}	0.090	-0.241^{***}	0.084	-0.066	0.073
Test for normality of innovations: Test statistic: 11.094; p-value: 0.004.						
Test for skewness of innovations: Test statistic: -0.764; p-value: 0.445; estimated skewness coefficient = -0.712.						
Test for excess kurtosis of innovations: Test statistic: 1.526; p-value: 0.064; estimated kurtosis coefficient = 5.846.						

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

very different. In particular, for the spatial lag dependence parameter λ, GPMLE and NGPMLE are positive, but BGMME is negative. NGPMLEs have uniformly smaller SEs than GPMLEs, but BGMMEs have larger SEs than GPMLE for some parameters. For λ, GPMLE is only significant at the 10% level, but NGPMLE is significant at the 1% level. For the spatial error dependence parameter ρ, both GPMLE and NGPMLE are insignificant.

Since the GPMLE and NGPMLE of ρ are insignificant, we estimate a SAR model and report the results in Table S10. The results for the normality test, the skewness test and the excess kurtosis test of innovations are similar to those for the SARAR model, showing some evidence of leptokurtic innovations. We still observe that NGPMLEs have uniformly smaller SEs than GPMLEs, but BGMMEs do not. All parameter estimates are significant at the 1% level, except the NGPMLE of the coefficient on House_value, which is not significant even at the 10% level.

Table S11 reports the impact estimates based on the SAR model. All impact estimates are significant at the 1% level, except that the average indirect impacts computed with GPMLE are significant at the 5% level. The impacts computed with NGPMLE have the same sign as those computed with GPMLE, although their differences can be large. The average indirect impacts computed with BGMME have different signs from those computed with GPMLE and NGPMLE.

Overall, for this data set with a small sample size, there is some evidence of non-normal

Table S10: Results of estimation and diagnostic tests based on the SAR model for the crime data

	GPMLE		BGMME		NGPMLE	
	estimate	SE	estimate	SE	estimate	SE
λ	0.431***	0.118	0.332***	0.121	0.469***	0.092
Constant	45.079***	7.163	50.960***	7.166	45.105***	6.057
Income	−1.032***	0.304	−1.264***	0.294	−1.633***	0.266
House_value	−0.266***	0.089	−0.244***	0.084	−0.060	0.073
Test for normality of innovations: Test statistic: 12.472; p-value: 0.002.						
Test for skewness of innovations: Test statistic: −0.800; p-value: 0.424; estimated skewness coefficient = −0.772.						
Test for excess kurtosis of innovations: Test statistic: 1.492; p-value: 0.068; estimated kurtosis coefficient = 5.981.						

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

Table S11: Impact estimates based on the SAR model for the crime data

		Average total impacts		Average direct impacts		Average indirect impacts	
		estimate	SE	estimate	SE	estimate	SE
Income	GPMLE	75.615***	14.322	49.525***	1.423	26.090**	12.924
	BGMME	45.493***	4.127	63.731***	1.043	−18.238***	5.128
	NGPMLE	83.533***	14.008	49.054***	1.467	34.479***	12.558
House_value	GPMLE	−1.623***	0.324	−1.063***	0.094	−0.560**	0.276
	BGMME	−0.568***	0.077	−0.796***	0.089	0.228***	0.071
	NGPMLE	−2.961***	0.505	−1.739***	0.090	−1.222***	0.445

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

and leptokurtic innovations, and different estimation methods lead to very different results.

6.3 Application to a presidential election data set

In this section, we apply our NGPMLE to the data set in Pace and Barry (1997), which is on the votes cast in the 1980 presidential election across $3,107$ U.S. counties. This data set is used in LeSage (1999b) and LeSage and Pace (2009). As in LeSage (1999b, p. 95), we estimate an SARAR model, where the dependent variable is the logged proportion of voting age population that voted in the election, and the explanatory variables include an intercept term, the logged proportion of the population with high school level education or higher (Education), the logged proportion of the population that are homeowners (Home-

Table S12: Results of estimation and diagnostic tests for the presidential election data

	GPMLE		BGMME		NGPMLE	
	estimate	SE	estimate	SE	estimate	SE
λ	0.311^{***}	0.021	0.259^{***}	0.023	0.343^{***}	0.019
ρ	0.591^{***}	0.029	0.666^{***}	0.025	0.605^{***}	0.026
constant	0.586^{***}	0.054	0.206^{***}	0.057	0.765^{***}	0.048
Education	0.245^{***}	0.020	0.129^{***}	0.021	0.302^{***}	0.018
Homeowners	0.557^{***}	0.015	0.602^{***}	0.016	0.523^{***}	0.013
Income	-0.113^{***}	0.020	0.033	0.021	-0.187^{***}	0.018
Test for normality of innovations: Test statistic: 7.336×10^3; p-value: 0.000.						
Test for skewness of innovations: Test statistic: 0.277; p-value: 0.782; estimated skewness coefficient = 0.086.						
Test for excess kurtosis of innovations: Test statistic: 2.627; p-value: 0.004; estimated kurtosis coefficient = 9.042.						

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

owners), and the logged income per capita (Income). The spatial weights matrix W_n is a first order contiguity matrix, and the spatial weights matrix M_n is on the basis of second order contiguity, i.e., two counties are treated as connected if they are contiguous or have a joint neighbor. Both W_n and M_n are row-normalized.

Table S12 reports the results of estimation and some diagnostic tests. The normality test of innovations rejects the normality of innovations. The skewness test of innovations does not show evidence of skewed innovations, while the excess kurtosis test shows some evidence of leptokurtic innovations, with a p-value smaller than 0.01 and a kurtosis coefficient estimate of 9.042. The GPMLEs, BGMMEs and NGPMLEs of model parameters can have large differences. The SEs of BGMME are slightly larger than those of GPMLE, except the SE for the spatial error dependence parameter ρ, but the SEs of NGPMLE are uniformly smaller than those of GPMLE. All coefficient estimates are significant at the 1% level, except that the BGMME of the coefficient on Income is not significant at any usual significance level.

The impact estimates computed with different parameter estimates are reported in Table S13. For a given variable and a given kind of impact, different parameter estimates can generate very different impact estimates. All the impacts are significant at the 1% level, except that the average total, direct and indirect impacts of Income computed with BGMME are all insignificant.

Table S13: Impact estimates for the presidential election data

		Average total impacts		Average direct impacts		Average indirect impacts	
		estimate	SE	estimate	SE	estimate	SE
Education	GPMLE	0.356***	0.029	0.251***	0.021	0.105***	0.012
	BGMME	0.175***	0.028	0.131***	0.021	0.043***	0.008
	NGPMLE	0.459***	0.027	0.310***	0.018	0.149***	0.013
Homeowners	GPMLE	0.808***	0.032	0.570***	0.016	0.238***	0.023
	BGMME	0.813***	0.031	0.612***	0.016	0.201***	0.023
	NGPMLE	0.796***	0.029	0.538***	0.014	0.258***	0.022
Income	GPMLE	−0.164***	0.029	−0.115***	0.021	−0.048***	0.009
	BGMME	0.045	0.028	0.034	0.021	0.011	0.007
	NGPMLE	−0.284***	0.027	−0.192***	0.018	−0.092***	0.011

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

7 Proof of Lemma B.3

Lemma B.3. *Suppose that $h(x)$ is a scalar function, v_i's in $V_n = [v_1, \cdots, v_n]'$ 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, $\mathrm{E}(|v_i|^{c_v}) < \infty$ 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. We only prove the result for c_{1n}, as that for c_{2n} is similar. The relations between the indices i, j and k include the following cases: $i = j = k$, $i = j \neq k$, $i = k \neq j$, $i \neq j = k$ and $(i \neq j, j \neq k, k \neq i)$. For simplicity, denote $h_0(v_i) = h(v_i) - \mathrm{E}[h(v_i)]$, $h_1(v_i) = h(v_i)v_i - \mathrm{E}[h(v_i)v_i]$ and $h_2(v_i) = h(v_i)v_i^2 - \mathrm{E}[h(v_i)v_i^2]$. Then,

$$\begin{aligned}
c_{1n} - \mathrm{E}(c_{1n}) &= \frac{1}{n}\sum_{i=1}^n h_2(v_i)a_{n,ii}b_{n,ii} + \frac{1}{n}\sum_{i=1}^n\sum_{j\neq i}\{h_1(v_i) + \mathrm{E}[h(v_i)v_i]\}v_j(a_{n,ij}b_{n,ii} + a_{n,ii}b_{n,ij}) \\
&\quad + \frac{1}{n}\sum_{i=1}^n\sum_{j\neq i}\{h_0(v_i)(v_j^2 - \sigma_0^2) + \sigma_0^2 h_0(v_i) + \mathrm{E}[h(v_i)](v_j^2 - \sigma_0^2)\}a_{n,ij}b_{n,ij} \\
&\quad + \frac{1}{n}\sum_{i=1}^n\sum_{j\neq i}\sum_{k:\, k\neq i, k\neq j}\{h_0(v_i) + \mathrm{E}[h(v_i)]\}v_j v_k a_{n,ij}b_{n,ik} \\
&= \frac{1}{n}\sum_{i=1}^n c_{1n,i},
\end{aligned}$$

where

$$c_{1n,i} = h_2(v_i)a_{n,ii}b_{n,ii} + h_1(v_i)\sum_{j=1}^{i-1} v_j(a_{n,ij}b_{n,ii} + a_{n,ii}b_{n,ij}) + v_i\sum_{j=1}^{i-1} h_1(v_j)(a_{n,ji}b_{n,jj} + a_{n,jj}b_{n,ji})$$
$$+ \mathrm{E}[h(v_i)v_i]v_i\sum_{j\neq i}(a_{n,ji}b_{n,jj} + a_{n,jj}b_{n,ji}) + h_0(v_i)\sum_{j=1}^{i-1}(v_j^2 - \sigma_0^2)a_{n,ij}b_{n,ij}$$
$$+ (v_i^2 - \sigma_0^2)\sum_{j=1}^{i-1} h_0(v_j)a_{n,ji}b_{n,ji} + \sigma_0^2 h_0(v_i)\sum_{j\neq i} a_{n,ij}b_{n,ij} + \mathrm{E}[h(v_i)](v_i^2 - \sigma_0^2)\sum_{j\neq i} a_{n,ji}b_{n,ji}$$
$$+ \sum_{j=1}^{i-1}\sum_{k=1}^{j-1}[h_0(v_i)v_jv_k(a_{n,ij}b_{n,ik} + a_{n,ik}b_{n,ij})$$
$$+ v_ih_0(v_j)v_k(a_{n,ji}b_{n,jk} + a_{n,jk}b_{n,ji}) + v_iv_jh_0(v_k)(a_{n,ki}b_{n,kj} + a_{n,kj}b_{n,ki})]$$
$$+ \mathrm{E}[h(v_i)]v_i\sum_{j=1}^{i-1} v_j \sum_{k:\, k\neq i, k\neq j}(a_{n,ki}b_{n,kj} + a_{n,kj}b_{n,ki}).$$

Note that each term in $c_{1n,i}$ has a zero conditional mean given $v_1, \ldots, v_{i-1}$. Then each term is a martingale difference. Furthermore, each term in $c_{1n,i}$ can be shown to be uniformly integrable (UI). We only prove that the typical terms (i) $z_{1n,i} \equiv h_2(v_i)a_{n,ii}b_{n,ii}$, (ii) $z_{2n,i} \equiv h_1(v_i)\sum_{j=1}^{i-1} v_ja_{n,ij}b_{n,ii}$ and

$$(iii)\ z_{3n,i} = \sum_{j=1}^{i-1}\sum_{k=1}^{j-1} h_0(v_i)v_jv_ka_{n,ij}b_{n,ik} = h_0(v_i)\sum_{j=1}^{i-1} a_{n,ij}v_j\sum_{k=1}^{j-1} b_{n,ik}v_k$$

are UI. Note that the three terms all have the form $\prod_{j=1}^{l}(e'_{ni}A_{jn}U_{jn})$ in Lemma B.2, where each A_{jn} is bounded in both the row and column sum norms. Let $\iota_1 = 1/(\frac{1}{c_h} + \frac{2}{c_v}) > 1$. Then by Hölder's inequality, $\mathrm{E}[|h(v_i)v_i^2|^{\iota_1}] \leq [\mathrm{E}\,|h(v_i)|^{\iota_1\cdot c_h/\iota_1}]^{\iota_1/c_h}[\mathrm{E}(|v_i|^{2\iota_1\cdot c_v/(2\iota_1)})]^{2\iota_1/c_v} < \infty$. Thus, $\mathrm{E}[|h_2(v_i)|^{\iota_1}] \leq c\{\mathrm{E}[|h(v_i)v_i^2|^{\iota_1}] + |\,\mathrm{E}[h(v_i)v_i^2]|^{\iota_1}\} < \infty$ for some constant c by the c_r-inequality and Jensen's inequality. Similarly, we have $\mathrm{E}[|h_1(v_i)|^{\iota_2}] < \infty$ for $\iota_2 = 1/(\frac{1}{c_h} + \frac{1}{c_v})$ and $\mathrm{E}[|h_0(v_i)|^{c_h}] < \infty$. Then by Lemma B.2, $\sum_{j=1}^{3}\sup_{i,n}\mathrm{E}(|z_{jn,i}|^{\iota_1}) < \infty$. Thus, $z_{jn,i}$ is UI for $j = 1, 2, 3$. Hence, by the LLN for martingale differences in Theorem 19.7 of Davidson (1994, p. 299), $c_{1n} - \mathrm{E}(c_{1n}) = o_p(1)$. □

8 NGPMLE for the SAR model

In Remark 1 of the main text, we claim that, for the SAR model with no SAR process of disturbances, a result similar to Proposition 1(i) holds when Assumption 4(i) is replaced by the conditions that (a) X_n contains an intercept term and (b) $\mathrm{E}[\ln f(\frac{\sigma_0 v_i - \alpha}{\sigma}, \eta)] - \frac{1}{2}\ln(\sigma^2)$ is uniquely maximized at $(\sigma_\infty, \alpha_\infty, \eta_\infty)$. We prove the claim in this section. Note that (b) is equivalent to that $\mathcal{Q}(\sigma, \beta_1, \eta) \equiv \mathrm{E}\big[\ln f\big(\frac{\sigma_0 v_i - (\beta_1 - \beta_{10})}{\sigma}, \eta\big)\big] - \frac{1}{2}\ln(\sigma^2)$ is uniquely maximized at $(\sigma, \beta_1, \eta) = (\sigma_\infty, \beta_{1\infty}, \eta_\infty)$ for $\beta_{1\infty} = \beta_{10} + \alpha_\infty$. Consider the SAR model:

$$Y_n = \lambda_0 W_n Y_n + X_n \beta_0 + \sigma_0 V_n, \tag{S11}$$

where the notations are similar to those for the SARAR model in the main text. For a given λ_0, the model is a linear regression model $S_n Y_n = X_n \beta_0 + \sigma_0 V_n$, with $S_n Y_n$ being the dependent variable and X_n being the exogenous variable matrix. The pseudo log likelihood function of model (S11), as if v_i had the density function $f(v_i, \eta)$, is

$$\ln L_n(\gamma) = \sum_{i=1}^n \ln f\big(v_i(\theta), \eta\big) - \frac{n}{2}\ln(\sigma^2) + \ln|S_n(\lambda)|, \tag{S12}$$

where $\theta = [\lambda, \beta', \sigma^2]'$, $\gamma = [\theta', \eta']'$ and $v_i(\theta) = \frac{1}{\sigma} e_{ni}'[S_n(\lambda)Y_n - X_n\beta]$. Denote $T_n(\lambda) = S_n(\lambda)S_n^{-1} = [t_{n,ij}(\lambda)]$, and $\Psi_{ni}(\theta) = \sigma_0 e_{ni}' T_n(\lambda) V_n - \sigma_0 v_i t_{n,ii}(\lambda) + e_{ni}'[S_n(\lambda)S_n^{-1}X_n\beta_0 - X_n\beta]$. Note that $\Psi_{ni}(\theta)$ does not depend on v_i. As $Y_n = S_n^{-1}(X_n\beta_0 + \sigma_0 V_n)$, $v_i(\theta) = \frac{1}{\sigma}\Psi_i(\theta) + \frac{1}{\sigma}\sigma_0 v_i t_{n,ii}(\lambda)$. Denote $\sigma_{ni} = \frac{\sigma}{t_{n,ii}(\lambda)}$, and $\beta_{1,ni} = \beta_{10} - \frac{1}{t_{n,ii}(\tau)}\Psi_{ni}(\theta)$. Let $\mathrm{E}_{-i}(\cdot)$ be the conditional expectation given v_1, ..., v_{i-1}, v_{i+1}, ..., v_n. Then,

$$\begin{aligned}
\mathrm{E}[\ln L_n(\gamma)] &= \sum_{i=1}^n \mathrm{E}\{\mathrm{E}_{-i}[\ln f(v_i(\theta), \eta)]\} - \frac{n}{2}\ln(\sigma^2) + \ln|S_n(\lambda)| \\
&= \sum_{i=1}^n \mathrm{E}[\mathcal{Q}(\sigma_{ni}, \beta_{1,ni}, \eta)] - \sum_{i=1}^n \ln|t_{n,ii}(\lambda)| + \ln|S_n(\lambda)| \\
&\le n\mathcal{Q}(\sigma_\infty, \beta_{1\infty}, \eta_\infty) - \sum_{i=1}^n \ln|t_{n,ii}(\lambda)| + \ln|S_n(\lambda)| \qquad \text{(S13)} \\
&= n\mathcal{Q}(\sigma_\infty, \beta_{1\infty}, \eta_\infty) - \sum_{i=1}^n \ln|t_{n,ii}(\lambda)| + \ln|T_n(\lambda)| + \ln|S_n| \\
&\le \mathrm{E}[\ln L_n(\gamma_*)], \qquad \text{(S14)}
\end{aligned}$$

where (S13) uses the assumption that $\mathcal{Q}(\sigma, \beta_1, \eta)$ is uniquely maximized at $(\sigma, \beta_1, \eta) = (\sigma_\infty, \beta_{1\infty}, \eta_\infty)$, (S14) uses the assumption that $\ln|T_n(\lambda)| \le \sum_{i=1}^n \ln|t_{n,ii}(\lambda)|$, and $\gamma_* = [\lambda_0, \beta_{1\infty}, \beta_{20}', \sigma_\infty^2, \eta_\infty']'$. The inequality in (S14) is strict if $\lambda \neq \lambda_0$. With $\lambda = \lambda_0$, we have $T_n(\lambda) = I_n$, $t_{n,ii}(\lambda) = 1$, $\sigma_{ni} = \sigma$, and $\beta_{1,ni} = \beta_{10} - e_{ni}' X_n(\beta_0 - \beta) = \beta_1 - e_{ni}' X_{2n}(\beta_{20} - \beta_2)$. Since X_n has full column rank, $\beta_{1,ni} \neq \beta_{1\infty}$ for some i if $\beta_2 \neq \beta_{20}$. Thus, with $\lambda = \lambda_0$, the inequality in (S13) is strict if $(\beta_2, \sigma, \eta) \neq (\beta_{20}, \sigma_\infty, \eta_\infty)$. Hence, $\mathrm{E}[\ln L_n(\gamma)]$ is uniquely maximized at γ_*.

9 NGPMLEs for other spatial econometric models

In this section, we show that the NGPMLEs for some other spatial econometric models can also be consistent under conditions similar to those for the SARAR model. We consider the SAR model with spatial moving average (MA) disturbances, the matrix exponential spatial specification (MESS) model, and a high order SARAR model.

9.1 The SAR model with spatial MA disturbances

Consider the following SAR model with spatial MA disturbances:

$$Y_n = \lambda_0 W_n Y_n + X_n \beta_0 + U_n, \quad U_n = \sigma_0(\rho_0 M_n V_n + V_n), \tag{S15}$$

where the notations are similar to those for the SARAR model in the main text. The pseudo log likelihood function of model (S15), as if v_i had the density function $f(v_i, \eta)$, is

$$\ln L_n(\gamma) = \sum_{i=1}^{n} \ln f\big(v_i(\theta), \eta\big) - \frac{n}{2}\ln(\sigma^2) + \ln|S_n(\lambda)| - \ln|J_n(\rho)|, \tag{S16}$$

where $\theta = [\lambda, \rho, \beta', \sigma^2]'$, $\gamma = [\theta', \eta']'$, $S_n(\lambda) = I_n - \lambda W_n$, $J_n(\rho) = I_n + \rho M_n$, and $v_i(\theta) = \frac{1}{\sigma} e_{ni}' J_n^{-1}(\rho)[S_n(\lambda)Y_n - X_n\beta]$. Denote $T_n(\tau) = J_n^{-1}(\rho) S_n(\lambda) S_n^{-1} J_n = [t_{n,ij}(\tau)]$ with $\tau = [\lambda, \rho]'$, and $\Psi_{ni}(\theta) = \sigma_0 e_{ni}' T_n(\tau) V_n - \sigma_0 v_i t_{n,ii}(\tau) + e_{ni}' J_n^{-1}(\rho)[S_n(\lambda) S_n^{-1} X_n \beta_0 - X_n \beta]$. Note that $\Psi_{ni}(\theta)$ does not depend on v_i. As $Y_n = S_n^{-1}(X_n\beta_0 + \sigma_0 J_n V_n)$, $v_i(\theta) = \frac{1}{\sigma}\Psi_i(\theta) + \frac{1}{\sigma}\sigma_0 v_i t_{n,ii}(\tau)$.

Case 1: M_n is row-normalized and X_n contains an intercept term. Since M_n is row-normalized, $J_n 1_n = (1+\rho_0)1_n$. Then the nonsingularity of J_n implies that $\rho_0 \neq -1$. Denote

$$\mathcal{Q}(\sigma, \beta_1, \eta) = \mathrm{E}\left[\ln f\left(\frac{\sigma_0 v_i}{\sigma} - \frac{1}{(1+\rho_0)\sigma}(\beta_1 - \beta_{10}), \eta\right)\right] - \frac{1}{2}\ln(\sigma^2),$$

$\sigma_{ni} = \frac{\sigma}{t_{n,ii}(\tau)}$, and $\beta_{1,ni} = \beta_{10} - \frac{1+\rho_0}{t_{n,ii}(\tau)}\Psi_{ni}(\theta)$. Assume that (a) $\mathrm{E}[\ln f(\frac{\sigma_0 v_i - \alpha}{\sigma}, \eta)] - \frac{1}{2}\ln(\sigma^2)$ is uniquely maximized at $(\sigma_\infty, \alpha_\infty, \eta_\infty)$, and (b) $\ln|T_n(\tau)| < \sum_{i=1}^n \ln|t_{n,ii}(\tau)|$ for $\tau \neq \tau_0$. Note that (a) is equivalent to that $\mathcal{Q}(\sigma, \beta_1, \eta)$ is uniquely maximized at $(\sigma, \beta_1, \eta) = (\sigma_\infty, \beta_{1\infty}, \eta_\infty)$ for $\beta_{1\infty} = \beta_{10} + (1+\rho_0)\alpha_\infty$. Then,

$$\mathrm{E}[\ln L_n(\gamma)] = \sum_{i=1}^{n} \mathrm{E}\{\mathrm{E}_{-i}[\ln f(v_i(\theta), \eta)]\} - \frac{n}{2}\ln(\sigma^2) + \ln|S_n(\lambda)| - \ln|J_n(\rho)|$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} \mathrm{E}[\mathcal{Q}(\sigma_{ni}, \beta_{1,ni}, \eta)] - \sum_{i=1}^{n} \ln|t_{n,ii}(\tau)| + \ln|S_n(\lambda)| - \ln|J_n(\rho)| \\
&\leq n\mathcal{Q}(\sigma_\infty, \beta_{1\infty}, \eta_\infty) - \sum_{i=1}^{n} \ln|t_{n,ii}(\tau)| + \ln|S_n(\lambda)| - \ln|J_n(\rho)| \qquad \text{(S17)} \\
&= n\mathcal{Q}(\sigma_\infty, \beta_{1\infty}, \eta_\infty) - \sum_{i=1}^{n} \ln|t_{n,ii}(\tau)| + \ln|T_n(\tau)| + \ln|S_n| - \ln|J_n| \\
&\leq \mathrm{E}[\ln L_n(\gamma_*)], \qquad \text{(S18)}
\end{aligned}
$$

where $\gamma_* = [\lambda_0, \rho_0, \beta_{1\infty}, \beta_{20}', \sigma_\infty^2, \eta_\infty']'$. The inequality in (S18) is strict if $\tau \neq \tau_0$. With $\tau = \tau_0$, we have $T_n(\tau) = I_n$, $t_{n,ii}(\tau) = 1$, $\sigma_{ni} = \sigma$, and $\beta_{1,ni} = \beta_{10} - (1+\rho_0)e_{ni}' J_n^{-1} X_n(\beta_0 - \beta) = \beta_1 - (1+\rho_0)e_{ni}' J_n^{-1} X_{2n}(\beta_{20} - \beta_2)$. Assume that $J_n^{-1}X_n$ has full column rank, then $\beta_{1,ni} \neq \beta_{1\infty}$ for some i if $\beta_2 \neq \beta_{20}$. Thus, with $\tau = \tau_0$, the inequality in (S17) is strict if $(\beta_2, \sigma, \eta) \neq (\beta_{20}, \sigma_\infty, \eta_\infty)$. Hence, $\mathrm{E}[\ln L_n(\gamma)]$ is uniquely maximized at γ_*. Therefore, we can obtain similar conditions for the convergence of the NGPMLE for model (S15) to γ_*.

Case 2: v_i's are symmetrically distributed around zero with unimodal density. By Lemma A in Newey and Steigerwald (1997), $\mathrm{E}[\ln f(v_i(\theta), \eta)] = \mathrm{E}\{\mathrm{E}_{-i}[\ln f(v_i(\theta), \eta)]\} \leq \mathrm{E}\{\mathrm{E}_{-i}[\ln f(\frac{\sigma_0}{\sigma} v_i t_{n,ii}(\tau), \eta)]\} = \mathrm{E}[\ln f(\frac{\sigma_0}{\sigma} v_i t_{n,ii}(\tau), \eta)]$, where the inequality is strict if $\Psi_{ni}(\theta) \neq 0$. Denote $\mathcal{Q}(\sigma, \eta) = \mathrm{E}[\ln f(\frac{\sigma_0 v_i}{\sigma}, \eta)] - \frac{1}{2}\ln(\sigma^2)$. Assume that $\mathcal{Q}(\sigma, \eta)$ is uniquely maximized at $(\sigma, \eta) = (\sigma_\infty, \eta_\infty)$, and $\ln|T_n(\tau)| < \sum_{i=1}^n \ln|t_{n,ii}(\tau)|$ for $\tau \neq \tau_0$. Then,

$$
\begin{aligned}
\mathrm{E}[\ln L_n(\gamma)] &\leq \sum_{i=1}^{n} \mathcal{Q}(\sigma_{ni}, \eta) - \sum_{i=1}^{n} \ln|t_{n,ii}(\tau)| + \ln|S_n(\lambda)| - \ln|J_n(\rho)| \\
&\leq n\mathcal{Q}(\sigma_\infty, \eta_\infty) - \sum_{i=1}^{n} \ln|t_{n,ii}(\tau)| + \ln|S_n(\lambda)| - \ln|J_n(\rho)| \qquad \text{(S19)} \\
&\leq \mathrm{E}[\ln L_n(\gamma_\#)], \qquad \text{(S20)}
\end{aligned}
$$

where $\gamma_\# = [\lambda_0, \rho_0, \beta_0', \sigma_\infty^2, \eta_\infty']'$. The inequality in (S20) is strict if $\tau \neq \tau_0$. With $\tau = \tau_0$, the inequality in (S19) is strict if $(\sigma, \eta) \neq (\sigma_\infty, \eta_\infty)$. With $(\tau, \sigma, \eta) = (\tau_0, \sigma_\infty, \eta_\infty)$, we have $T_n(\tau) = I_n$ and $\Psi_{ni}(\theta) = e_{ni}' J_n^{-1} X_n(\beta_0 - \beta)$. Assume that $J_n^{-1}X_n$ has full column rank, with $(\tau, \sigma, \eta) = (\tau_0, \sigma_\infty^2, \eta_\infty)$, then the inequality in (S19) is strict if $\beta \neq \beta_0$. Therefore, $\mathrm{E}[\ln L_n(\gamma)]$ is uniquely maximized at $\gamma_\#$.

Case 3: Non-row-normalized M_n and asymmetric innovations. Add a location parameter α to the non-Gaussian pseudo log likelihood function (S16) so that we have a modified function

$$\ln L_n(\delta) = \sum_{i=1}^{n} \ln f\Big(v_i(\theta) - \frac{1}{\sigma}\alpha, \eta\Big) - \frac{n}{2}\ln(\sigma^2) + \ln|S_n(\lambda)| - \ln|J_n(\rho)|, \tag{S21}$$

where $\delta = [\lambda, \rho, \beta', \sigma^2, \alpha, \eta']'$. 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 $\alpha_{ni} = \frac{\alpha - \Psi_{ni}(\theta)}{t_{n,ii}(\tau)}$. Assume that $J_n^{-1}X_n$ has full column rank and does not contain an intercept term, $\mathcal{Q}(\sigma, \alpha, \eta)$ is uniquely maximized at $(\sigma, \alpha, \eta) = (\sigma_\infty, \alpha_\infty, \eta_\infty)$, and $\ln|T_n(\tau)| < \sum_{i=1}^{n} \ln|t_{n,ii}(\tau)|$ for $\tau \neq \tau_0$. Then,

$$\begin{aligned}
\mathrm{E}[\ln L_n(\delta)] &= \sum_{i=1}^{n} \mathrm{E}\Big\{\mathrm{E}_{-i}\Big[\ln f\Big(v_i(\theta) - \frac{\alpha}{\sigma}, \eta\Big)\Big]\Big\} - \frac{n}{2}\ln(\sigma^2) + \ln|S_n(\lambda)| - \ln|J_n(\rho)| \\
&= \sum_{i=1}^{n} \mathrm{E}[\mathcal{Q}(\sigma_{ni}, \alpha_{ni}, \eta)] - \sum_{i=1}^{n} \ln|t_{n,ii}(\tau)| + \ln|S_n(\lambda)| - \ln|J_n(\rho)| \\
&\leq n\mathcal{Q}(\sigma_\infty, \alpha_\infty, \eta_\infty) - \sum_{i=1}^{n} \ln|t_{n,ii}(\tau)| + \ln|S_n(\lambda)| - \ln|J_n(\rho)| \qquad \text{(S22)} \\
&= n\mathcal{Q}(\sigma_\infty, \alpha_\infty, \eta_\infty) - \sum_{i=1}^{n} \ln|t_{n,ii}(\tau)| + \ln|T_n(\tau)| + \ln|S_n| - \ln|J_n| \\
&\leq \mathrm{E}[\ln L_n(\delta_\#)], \qquad \text{(S23)}
\end{aligned}$$

where $\delta_\# = [\lambda_0, \rho_0, \beta_0', \sigma_\infty^2, \alpha_\infty, \eta_\infty']'$. The inequality in (S23) is strict if $\tau \neq \tau_0$. With $\tau = \tau_0$, we have $T_n(\tau) = I_n$, $t_{n,ii}(\tau) = 1$, $\sigma_{ni} = \sigma$, and $\alpha_{ni} = \alpha - e_{ni}' J_n^{-1} X_n(\beta_0 - \beta)$. Since $J_n^{-1}X_n$ has full column rank and does not contain an intercept term, $\alpha_{ni} \neq \alpha_\infty$ for some i if $\beta \neq \beta_0$. Thus, with $\tau = \tau_0$, the inequality in (S22) 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 $\delta = \delta_\#$.

9.2 The MESS model

Consider the following MESS model:

$$e^{\lambda_0 W_n} Y_n = X_n \beta_0 + U_n, \quad e^{\rho_0 M_n} U_n = \sigma_0 V_n, \tag{S24}$$

where the matrix exponential e^A is defined as $e^A = \sum_{i=0}^{\infty} \frac{A^i}{i!}$ for any square matrix A, and other notations are the same as for the SARAR model in the main text. Since the determinant $|e^A| = e^{\mathrm{tr}(A)}$ (LeSage and Pace, 2007) and W_n and M_n have zero diagonals, $|e^{\lambda W_n}| = 1$ and $|e^{\rho M_n}| = 1$. Then the pseudo log likelihood function of model (S24), as if v_i

had the density function $f(v_i, \eta)$, is

$$\ln L_n(\gamma) = \sum_{i=1}^{n} \ln f\big(v_i(\theta), \eta\big) - \frac{n}{2}\ln(\sigma^2), \tag{S25}$$

where $\theta = [\lambda, \rho, \beta', \sigma^2]'$, $\gamma = [\theta', \eta']'$ and $v_i(\theta) = \frac{1}{\sigma} e'_{ni} e^{\rho M_n}(e^{\lambda W_n} Y_n - X_n\beta)$. Denote $T_n(\tau) = e^{\rho M_n} e^{(\lambda-\lambda_0)W_n} e^{-\rho_0 M_n} = [t_{n,ij}(\tau)]$ with $\tau = [\lambda, \rho]'$, and $\Psi_{ni}(\theta) = \sigma_0 e'_{ni} T_n(\tau) V_n - \sigma_0 v_i t_{n,ii}(\tau) + e'_{ni} e^{\rho M_n}(e^{(\lambda-\lambda_0)W_n} X_n\beta_0 - X_n\beta)$. Note that $\Psi_{ni}(\theta)$ does not depend on v_i. As $Y_n = e^{-\lambda_0 W_n}(X_n\beta_0 + \sigma_0 e^{-\rho_0 M_n} V_n)$, $v_i(\theta) = \frac{1}{\sigma}\Psi_i(\theta) + \frac{1}{\sigma}\sigma_0 v_i t_{n,ii}(\tau)$.

Case 1: M_n is row-normalized and X_n contains an intercept term. Denote

$$\mathcal{Q}(\sigma, \beta_1, \eta) = \mathrm{E}\left[\ln f\left(\frac{\sigma_0 v_i - e^{\rho_0}(\beta_1 - \beta_{10})}{\sigma}, \eta\right)\right] - \frac{1}{2}\ln(\sigma^2),$$

$\sigma_{ni} = \frac{\sigma}{t_{n,ii}(\tau)}$, and $\beta_{1,ni} = \beta_{10} - \frac{e^{-\rho_0}}{t_{n,ii}(\tau)}\Psi_{ni}(\theta)$. Assume that (*a*) $\mathrm{E}[\ln f(\frac{\sigma_0 v_i - \alpha}{\sigma}, \eta)] - \frac{1}{2}\ln(\sigma^2)$ is uniquely maximized at $(\sigma_\infty, \alpha_\infty, \eta_\infty)$, and (*b*) $\ln|T_n(\tau)| < \sum_{i=1}^{n}\ln|t_{n,ii}(\tau)|$ for $\tau \neq \tau_0$. Note that (*a*) is equivalent to that $\mathcal{Q}(\sigma, \beta_1, \eta)$ is uniquely maximized at $(\sigma, \beta_1, \eta) = (\sigma_\infty, \beta_{1\infty}, \eta_\infty)$ for $\beta_{1\infty} = \beta_{10} + e^{-\rho_0}\alpha_\infty$. Then,

$$\begin{aligned}
\mathrm{E}[\ln L_n(\gamma)] &= \sum_{i=1}^{n} \mathrm{E}\{\mathrm{E}_{-i}[\ln f(v_i(\theta), \eta)]\} - \frac{n}{2}\ln(\sigma^2) \\
&= \sum_{i=1}^{n} \mathrm{E}[\mathcal{Q}(\sigma_{ni}, \beta_{1,ni}, \eta)] - \sum_{i=1}^{n}\ln|t_{n,ii}(\tau)| \\
&\leq n\mathcal{Q}(\sigma_\infty, \beta_{1\infty}, \eta_\infty) - \sum_{i=1}^{n}\ln|t_{n,ii}(\tau)| \qquad \text{(S26)} \\
&\leq n\mathcal{Q}(\sigma_\infty, \beta_{1\infty}, \eta_\infty) - \ln|T_n(\tau)| \qquad \text{(S27)} \\
&= \mathrm{E}[\ln L_n(\gamma_*)],
\end{aligned}$$

where $\gamma_* = [\lambda_0, \rho_0, \beta_{1\infty}, \beta'_{20}, \sigma^2_\infty, \eta'_\infty]'$ and the last equality uses $\ln|T_n(\tau)| = \ln|e^{\rho M_n} e^{(\lambda-\lambda_0)W_n} e^{-\rho_0 M_n}| = \ln(e^{\rho \operatorname{tr}(M_n)} e^{(\lambda-\lambda_0)\operatorname{tr}(W_n)} e^{-\rho_0 \operatorname{tr}(M_n)}) = 0$. The inequality in (S27) is strict if $\tau \neq \tau_0$. With $\tau = \tau_0$, we have $T_n(\tau) = I_n$, $t_{n,ii}(\tau) = 1$, $\sigma_{ni} = \sigma$, and $\beta_{1,ni} = \beta_{10} - e^{-\rho_0} e'_{ni} e^{\rho_0 M_n} X_n(\beta_0 - \beta) = \beta_1 - e^{-\rho_0} e'_{ni} e^{\rho_0 M_n} X_{2n}(\beta_{20} - \beta_2)$. Assume that $e^{\rho_0 M_n} X_n$ has full column rank, then $\beta_{1,ni} \neq \beta_{1\infty}$ for some i if $\beta_2 \neq \beta_{20}$. Thus, with $\tau = \tau_0$, the inequality in (S26) is strict if $(\beta_2, \sigma, \eta) \neq (\beta_{20}, \sigma_\infty, \eta_\infty)$. Hence, $\mathrm{E}[\ln L_n(\gamma)]$ is uniquely maximized at γ_*.

Case 2: v_i's are symmetrically distributed around zero with unimodal density. By Lemma A in Newey and Steigerwald (1997), $\mathrm{E}[\ln f(v_i(\theta), \eta)] = \mathrm{E}\{\mathrm{E}_{-i}[\ln f(v_i(\theta), \eta)]\} \leq \mathrm{E}\{\mathrm{E}_{-i}[\ln f(\frac{\sigma_0}{\sigma} v_i t_{n,ii}(\tau), \eta)]\} = \mathrm{E}[\ln f(\frac{\sigma_0}{\sigma} v_i t_{n,ii}(\tau), \eta)]$, where the inequality is strict if $\Psi_{ni}(\theta) \neq$

0. Denote $\mathcal{Q}(\sigma,\eta) = \mathrm{E}[\ln f(\frac{\sigma_0 v_i}{\sigma},\eta)] - \frac{1}{2}\ln(\sigma^2)$. Assume that $\mathcal{Q}(\sigma,\eta)$ is uniquely maximized at $(\sigma,\eta) = (\sigma_\infty,\eta_\infty)$, and $\ln|T_n(\tau)| < \sum_{i=1}^n \ln|t_{n,ii}(\tau)|$ for $\tau \neq \tau_0$. Then,

$$
\begin{aligned}
\mathrm{E}[\ln L_n(\gamma)] &\leq \sum_{i=1}^n \mathcal{Q}(\sigma_{ni},\eta) - \sum_{i=1}^n \ln|t_{n,ii}(\tau)| \\
&\leq n\mathcal{Q}(\sigma_\infty,\eta_\infty) - \sum_{i=1}^n \ln|t_{n,ii}(\tau)| \qquad \text{(S28)} \\
&\leq n\mathcal{Q}(\sigma_\infty,\eta_\infty) - \ln|T_n(\tau)| \qquad \text{(S29)} \\
&= \mathrm{E}[\ln L_n(\gamma_\#)],
\end{aligned}
$$

where $\gamma_\# = [\lambda_0,\rho_0,\beta_0',\sigma_\infty^2,\eta_\infty']'$. The inequality in (S29) is strict if $\tau \neq \tau_0$. With $\tau = \tau_0$, the inequality in (S28) is strict if $(\sigma,\eta) \neq (\sigma_\infty,\eta_\infty)$. With $(\tau,\sigma,\eta) = (\tau_0,\sigma_\infty,\eta_\infty)$, we have $T_n(\tau) = I_n$ and $\Psi_{ni}(\theta) = e_{ni}' e^{\rho_0 M_n} X_n(\beta_0 - \beta)$. Assume that $e^{\rho_0 M_n} X_n$ has full column rank, with $(\tau,\sigma,\eta) = (\tau_0,\sigma_\infty^2,\eta_\infty)$, then the inequality in (S28) is strict if $\beta \neq \beta_0$. Therefore, $\mathrm{E}[\ln L_n(\gamma)]$ is uniquely maximized at $\gamma_\#$.

Case 3: Non-row-normalized M_n and asymmetric innovations. Add a location parameter α to the non-Gaussian pseudo log likelihood function (S25) so that we have a modified function

$$
\ln L_n(\delta) = \sum_{i=1}^n \ln f\Big(v_i(\theta) - \frac{1}{\sigma}\alpha, \eta\Big) - \frac{n}{2}\ln(\sigma^2), \qquad \text{(S30)}
$$

where $\delta = [\lambda,\rho,\beta',\sigma^2,\alpha,\eta']'$. 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 $\alpha_{ni} = \frac{\alpha - \Psi_{ni}(\theta)}{t_{n,ii}(\tau)}$. Assume that $e^{\rho_0 M_n} X_n$ has full column rank and does not contain an intercept term, $\mathcal{Q}(\sigma,\alpha,\eta)$ is uniquely maximized at $(\sigma,\alpha,\eta) = (\sigma_\infty,\alpha_\infty,\eta_\infty)$, and $\ln|T_n(\tau)| < \sum_{i=1}^n \ln|t_{n,ii}(\tau)|$ for $\tau \neq \tau_0$. Then,

$$
\begin{aligned}
\mathrm{E}[\ln L_n(\delta)] &= \sum_{i=1}^n \mathrm{E}\big\{\mathrm{E}_{-i}\big[\ln f\big(v_i(\theta) - \frac{\alpha}{\sigma},\eta\big)\big]\big\} - \frac{n}{2}\ln(\sigma^2) \\
&= \sum_{i=1}^n \mathrm{E}[\mathcal{Q}(\sigma_{ni},\alpha_{ni},\eta)] - \sum_{i=1}^n \ln|t_{n,ii}(\tau)| \\
&\leq n\mathcal{Q}(\sigma_\infty,\alpha_\infty,\eta_\infty) - \sum_{i=1}^n \ln|t_{n,ii}(\tau)| \qquad \text{(S31)} \\
&\leq n\mathcal{Q}(\sigma_\infty,\alpha_\infty,\eta_\infty) - \ln|T_n(\tau)| \qquad \text{(S32)} \\
&= \mathrm{E}[\ln L_n(\delta_\#)],
\end{aligned}
$$

where $\delta_\# = [\lambda_0,\rho_0,\beta_0',\sigma_\infty^2,\alpha_\infty,\eta_\infty']'$. The inequality in (S32) is strict if $\tau \neq \tau_0$. With $\tau = \tau_0$, we have $T_n(\tau) = I_n$, $t_{n,ii}(\tau) = 1$, $\sigma_{ni} = \sigma$, and $\alpha_{ni} = \alpha - e_{ni}' e^{\rho_0 M_n} X_n(\beta_0 - \beta)$. Since $e^{\rho_0 M_n} X_n$

has full column rank and does not contain an intercept term, $\alpha_{ni} \neq \alpha_\infty$ for some i if $\beta \neq \beta_0$. Thus, with $\tau = \tau_0$, the inequality in (S31) 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 $\delta = \delta_\#$.

9.3 A high order SARAR model

Consider the following high order SARAR model:

$$Y_n = \sum_{j=1}^{p} \lambda_{j0} W_{jn} Y_n + X_n \beta_0 + U_n, \quad U_n = \sum_{k=1}^{q} \rho_{k0} M_{kn} U_n + \sigma_0 V_n, \tag{S33}$$

where W_{jn}'s and M_{kn}'s are $n \times n$ spatial weights matrices and other notations are similar to those for the SARAR model in the main text. Denote $\lambda = [\lambda_1, \ldots, \lambda_p]'$, $\rho = [\rho_1, \ldots, \rho_q]'$, $\theta = [\lambda', \rho', \beta', \sigma^2]'$, $\gamma = [\theta', \eta']'$, $S_n(\lambda) = I_n - \sum_{j=1}^{p} \lambda_j W_{jn}$ and $R_n(\rho) = I_n - \sum_{k=1}^{q} \rho_k M_{kn}$. Let $\theta_0 = [\lambda_0', \rho_0', \beta_0', \sigma_0^2]'$ be the true value of θ, $R_n = R_n(\rho_0)$ and $S_n = S_n(\lambda_0)$. The pseudo log likelihood function of model (S33), as if v_i had the density function $f(v_i, \eta)$, is

$$\ln L_n(\gamma) = \sum_{i=1}^{n} \ln f\big(v_i(\theta), \eta\big) - \frac{n}{2} \ln(\sigma^2) + \ln |S_n(\lambda)| + \ln |R_n(\rho)|, \tag{S34}$$

where $v_i(\theta) = \frac{1}{\sigma} e_{ni}' R_n(\rho)[S_n(\lambda) Y_n - X_n \beta]$. Denote $T_n(\tau) = R_n(\rho) S_n(\lambda) S_n^{-1} R_n^{-1} = [t_{n,ij}(\tau)]$ with $\tau = [\lambda', \rho']'$, and $\Psi_{ni}(\theta) = \sigma_0 e_{ni}' T_n(\tau) V_n - \sigma_0 v_i t_{n,ii}(\tau) + e_{ni}' R_n(\rho)[S_n(\lambda) S_n^{-1} X_n \beta_0 - X_n \beta]$. Note that $\Psi_{ni}(\theta)$ does not depend on v_i. As $Y_n = S_n^{-1}(X_n \beta_0 + \sigma_0 R_n^{-1} V_n)$, $v_i(\theta) = \frac{1}{\sigma} \Psi_i(\theta) + \frac{1}{\sigma} \sigma_0 v_i t_{n,ii}(\tau)$.

Case 1: All M_{kn}'s are row-normalized and X_n contains an intercept term. Since M_{kn}'s are row-normalized, $R_n 1_n = c(\rho_0) 1_n$ for $c(\rho_0) = 1 - \sum_{k=1}^{q} \rho_{k0}$. Then the nonsingularity of R_n implies that $c(\rho_0) \neq 0$. Denote

$$\mathcal{Q}(\sigma, \beta_1, \eta) = \mathrm{E}\bigg[\ln f\bigg(\frac{\sigma_0 v_i}{\sigma} - \frac{c(\rho_0)}{\sigma}(\beta_1 - \beta_{10}), \eta\bigg)\bigg] - \frac{1}{2} \ln(\sigma^2),$$

$\sigma_{ni} = \frac{\sigma}{t_{n,ii}(\tau)}$, and $\beta_{1,ni} = \beta_{10} - \frac{1}{c(\rho_0) t_{n,ii}(\tau)} \Psi_{ni}(\theta)$. Assume that (a) $\mathrm{E}[\ln f(\frac{\sigma_0 v_i - \alpha}{\sigma}, \eta)] - \frac{1}{2} \ln(\sigma^2)$ is uniquely maximized at $(\sigma_\infty, \alpha_\infty, \eta_\infty)$, and (b) $\ln |T_n(\tau)| < \sum_{i=1}^{n} \ln |t_{n,ii}(\tau)|$ for $\tau \neq \tau_0$. Note that (a) is equivalent to that $\mathcal{Q}(\sigma, \beta_1, \eta)$ is uniquely maximized at $(\sigma, \beta_1, \eta) = (\sigma_\infty, \beta_{1\infty}, \eta_\infty)$ for $\beta_{1\infty} = \beta_{10} + \frac{1}{c(\rho_0)} \alpha_\infty$.Then,

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

$$
\begin{aligned}
&= \sum_{i=1}^{n} \mathrm{E}[\mathcal{Q}(\sigma_{ni}, \beta_{1,ni}, \eta)] - \sum_{i=1}^{n} \ln|t_{n,ii}(\tau)| + \ln|S_n(\lambda)| + \ln|R_n(\rho)| \\
&\leq n\mathcal{Q}(\sigma_\infty, \beta_{1\infty}, \eta_\infty) - \sum_{i=1}^{n} \ln|t_{n,ii}(\tau)| + \ln|S_n(\lambda)| + \ln|R_n(\rho)| \qquad \text{(S35)} \\
&= n\mathcal{Q}(\sigma_\infty, \beta_{1\infty}, \eta_\infty) - \sum_{i=1}^{n} \ln|t_{n,ii}(\tau)| + \ln|T_n(\tau)| + \ln|S_n| + \ln|R_n| \\
&\leq \mathrm{E}[\ln L_n(\gamma_*)], \qquad \text{(S36)}
\end{aligned}
$$

where $\gamma_* = [\lambda_0', \rho_0', \beta_{1\infty}, \beta_{20}', \sigma_\infty^2, \eta_\infty']'$. The inequality in (S36) is strict if $\tau \neq \tau_0$. With $\tau = \tau_0$, we have $T_n(\tau) = I_n$, $t_{n,ii}(\tau) = 1$, $\sigma_{ni} = \sigma$, and $\beta_{1,ni} = \beta_{10} - \frac{1}{c(\rho_0)} e_{ni}' R_n X_n(\beta_0 - \beta) = \beta_1 - \frac{1}{c(\rho_0)} e_{ni}' R_n X_{2n}(\beta_{20} - \beta_2)$. Assume that $R_n X_n$ has full column rank, then $\beta_{1,ni} \neq \beta_{1\infty}$ for some i if $\beta_2 \neq \beta_{20}$. Thus, with $\tau = \tau_0$, the inequality in (S35) is strict if $(\beta_2, \sigma, \eta) \neq (\beta_{20}, \sigma_\infty, \eta_\infty)$. Hence, $\mathrm{E}[\ln L_n(\gamma)]$ is uniquely maximized at γ_*.

Case 2: v_i's are symmetrically distributed around zero with unimodal density. By Lemma A in Newey and Steigerwald (1997), $\mathrm{E}[\ln f(v_i(\theta), \eta)] = \mathrm{E}\{\mathrm{E}_{-i}[\ln f(v_i(\theta), \eta)]\} \leq \mathrm{E}\{\mathrm{E}_{-i}[\ln f(\frac{\sigma_0}{\sigma} v_i t_{n,ii}(\tau), \eta)]\} = \mathrm{E}[\ln f(\frac{\sigma_0}{\sigma} v_i t_{n,ii}(\tau), \eta)]$, where the inequality is strict if $\Psi_{ni}(\theta) \neq 0$. Denote $\mathcal{Q}(\sigma, \eta) = \mathrm{E}[\ln f(\frac{\sigma_0 v_i}{\sigma}, \eta)] - \frac{1}{2}\ln(\sigma^2)$. Assume that $\mathcal{Q}(\sigma, \eta)$ is uniquely maximized at $(\sigma, \eta) = (\sigma_\infty, \eta_\infty)$, and $\ln|T_n(\tau)| < \sum_{i=1}^n \ln|t_{n,ii}(\tau)|$ for $\tau \neq \tau_0$. Then,

$$
\begin{aligned}
\mathrm{E}[\ln L_n(\gamma)] &\leq \sum_{i=1}^{n} \mathcal{Q}(\sigma_{ni}, \eta) - \sum_{i=1}^{n} \ln|t_{n,ii}(\tau)| + \ln|S_n(\lambda)| + \ln|R_n(\rho)| \\
&\leq n\mathcal{Q}(\sigma_\infty, \eta_\infty) - \sum_{i=1}^{n} \ln|t_{n,ii}(\tau)| + \ln|S_n(\lambda)| + \ln|R_n(\rho)| \qquad \text{(S37)} \\
&\leq \mathrm{E}[\ln L_n(\gamma_\#)], \qquad \text{(S38)}
\end{aligned}
$$

where $\gamma_\# = [\lambda_0', \rho_0', \beta_0', \sigma_\infty^2, \eta_\infty']'$. The inequality in (S38) is strict if $\tau \neq \tau_0$. With $\tau = \tau_0$, the inequality in (S37) is strict if $(\sigma, \eta) \neq (\sigma_\infty, \eta_\infty)$. With $(\tau, \sigma, \eta) = (\tau_0, \sigma_\infty, \eta_\infty)$, we have $T_n(\tau) = I_n$ and $\Psi_{ni}(\theta) = e_{ni}' R_n X_n(\beta_0 - \beta)$. Assume that $R_n X_n$ has full column rank, with $(\tau, \sigma, \eta) = (\tau_0, \sigma_\infty^2, \eta_\infty)$, then the inequality in (S37) is strict if $\beta \neq \beta_0$. Therefore, $\mathrm{E}[\ln L_n(\gamma)]$ is uniquely maximized at $\gamma_\#$.

Case 3: Non-row-normalized M_{kn}'s and asymmetric innovations. Add a location parameter α to the non-Gaussian pseudo log likelihood function (S34) so that we have a modified function

$$
\ln L_n(\delta) = \sum_{i=1}^{n} \ln f\Big(v_i(\theta) - \frac{1}{\sigma}\alpha, \eta\Big) - \frac{n}{2}\ln(\sigma^2) + \ln|S_n(\lambda)| + \ln|R_n(\rho)|, \qquad \text{(S39)}
$$

where $\delta = [\lambda', \rho', \beta', \sigma^2, \alpha, \eta']'$. 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 $\alpha_{ni} = \frac{\alpha - \Psi_{ni}(\theta)}{t_{n,ii}(\tau)}$. Assume that $R_n X_n$ has full column rank and does not contain an intercept term, $\mathcal{Q}(\sigma, \alpha, \eta)$ is uniquely maximized at $(\sigma, \alpha, \eta) = (\sigma_\infty, \alpha_\infty, \eta_\infty)$, and $\ln|T_n(\tau)| < \sum_{i=1}^n \ln|t_{n,ii}(\tau)|$ for $\tau \neq \tau_0$. Then,

$$
\begin{aligned}
\mathrm{E}[\ln L_n(\delta)] &= \sum_{i=1}^n \mathrm{E}\Big\{\mathrm{E}_{-i}\Big[\ln f\Big(v_i(\theta) - \frac{\alpha}{\sigma}, \eta\Big)\Big]\Big\} - \frac{n}{2}\ln(\sigma^2) + \ln|S_n(\lambda)| + \ln|R_n(\rho)| \\
&= \sum_{i=1}^n \mathrm{E}[\mathcal{Q}(\sigma_{ni}, \alpha_{ni}, \eta)] - \sum_{i=1}^n \ln|t_{n,ii}(\tau)| + \ln|S_n(\lambda)| + \ln|R_n(\rho)| \\
&\leq n\mathcal{Q}(\sigma_\infty, \alpha_\infty, \eta_\infty) - \sum_{i=1}^n \ln|t_{n,ii}(\tau)| + \ln|S_n(\lambda)| + \ln|R_n(\rho)| \qquad \text{(S40)} \\
&= n\mathcal{Q}(\sigma_\infty, \alpha_\infty, \eta_\infty) - \sum_{i=1}^n \ln|t_{n,ii}(\tau)| + \ln|T_n(\tau)| + \ln|S_n| + \ln|R_n| \\
&\leq \mathrm{E}[\ln L_n(\delta_\#)], \qquad \text{(S41)}
\end{aligned}
$$

where $\delta_\# = [\lambda_0', \rho_0', \beta_0', \sigma_\infty^2, \alpha_\infty, \eta_\infty']'$. The inequality in (S41) is strict if $\tau \neq \tau_0$. 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, $\alpha_{ni} \neq \alpha_\infty$ for some i if $\beta \neq \beta_0$. Thus, with $\tau = \tau_0$, the inequality in (S40) 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 $\delta = \delta_\#$.

References

Anselin, L., 1988. Spatial Econometrics: Methods and Models. Boston: Kluwer Academic Publishers.

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

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

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

Jin, F., Lee, L.F., 2018. Outer-product-of-gradients tests for spatial autoregressive models. Regional Science and Urban Economics 72, 35–57.

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

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

LeSage, J.P., 1999a. Spatial Econometrics. Department of Economics, University of Toledo.

LeSage, J.P., 1999b. The Theory and Practice of Spatial Econometrics. University of Toledo. Toledo, Ohio 28(11).

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

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

Pace, R.K., Barry, R., 1997. Sparse spatial autoregressions. Statistics and Probability Letters 33, 291–297.