

A Generalized Spatial Panel Data Model with Random Effects

Badi H. Baltagi*, Peter Egger** and Michael Pfaffermayr ***

October 6, 2007

Abstract

This paper proposes a generalized specification for the panel data model with random effects and first-order spatially autocorrelated residuals that encompasses two previously suggested specifications. The first one is described in Anselin's (1988) book and the second one by Kapoor, Kelejian, and Prucha (2007). Our encompassing specification allows us to test for these models as restricted specifications. In particular, we derive three LM and LR tests that restrict our generalized model to obtain (i) the Anselin model, (ii) the Kapoor, Kelejian,

*Badi H. Baltagi, Department of Economics and Center for Policy Research, Syracuse University, Syracuse, NY 13244-1020 U.S.A.; bbaltagi@maxwell.syr.edu;

**Peter Egger: University of Munich and CESifo, Poschingerstr. 5, 81679 Munich, Germany, E-mail: egger@ifo.de;

***Michael Pfaffermayr: Department of Economics, University of Innsbruck, Universitaetsstrasse 15, 6020 Innsbruck, Austria and Austrian Institute of Economic Research, P.O.-Box 91, A-1103 Vienna, Austria; Michael.Pfaffermayr@uibk.ac.at. Michael Pfaffermayr gratefully acknowledges financial support from the Austrian Science Foundation grant 17028. We would like to thank the editor Cheng Hsiao and two anonymous referees for their helpful comments and suggestions. Preliminary versions of this paper were presented at the 13th International conference on panel data held in Cambridge, England, and the 23rd annual Canadian econometric study group meeting in Niagara Falls, Canada.

and Prucha model, and (iii) the simple random effects model that ignores the spatial correlation in the residuals. We derive the large sample distributions of the three LM tests. For two of these three tests, we obtain closed form solutions. Our Monte Carlo results show that the suggested tests are powerful in testing for these restricted specifications even in small and medium sized samples.

JEL classification: C23; C12

Keywords: Panel data; Spatially autocorrelated residuals; maximum-likelihood estimation

1 Introduction

The recent literature on spatial panels distinguishes between two different spatial autoregressive error processes. One specification assumes that spatial correlation occurs only in the remainder error term, whereas no spatial correlation takes place in the individual effects (see Anselin, 1988, Baltagi, Song, and Koh, 2003, and Anselin, Le Gallo, and Jayet, 2006; henceforth referred to as the Anselin model). Another specification assumes that the same spatial error process applies to both the individual and remainder error components (see Kapoor, Kelejian, and Prucha, 2007; henceforth referred to as the KKP model).

While the two data generating processes look similar, they imply different spatial spillover mechanisms. For example, consider the question of firm productivity using panel data. Besides the deterministic components, firms differ also with respect to their unobserved know-how or their managerial ability to organize production processes efficiently. At least over a short time period, this managerial ability may be time-invariant. Beyond that there are innovations that vary from period to period like random firm-specific technology shocks, capacity utilization shocks, etc. Under this scenario, it seems reasonable to assume that firm productivity may be spatially correlated due to spillovers. Such spillovers can occur, e.g., through information flows (transmission of process technologies) embodied in worker flows between firms at local labor markets or through input-output channels (technology requirements and interdependence of capacity utilization). Whereas the Anselin model assumes that spillovers are inherently time-varying, the KKP process assumes the spillovers to be time-invariant as well as time-variant. For ex-

ample, firms located in the neighborhood of highly productive firms may get time-invariant permanent spillovers affecting their productivity in addition to the time-variant spillovers as in the Anselin model. While the Anselin model seems restrictive in that it does not allow permanent spillovers through the individual firm effects, the KKP approach is restrictive in the sense that it does not allow for a differential intensity of spillovers of the permanent and transitory shocks.

This paper introduces a generalized spatial panel model which encompasses these two models and allows for spatial correlation in the individual and remainder error components that may have different spatial autoregressive parameters. We derive the maximum likelihood estimator (MLE) for this more general spatial panel model when the individual effects are assumed to be random. This in turn allows us to test the restrictions on our generalized model to obtain (i) the Anselin model, (ii) the Kapoor, Kelejian, and Prucha model, and (iii) a simple random effects model that ignores the spatial correlation in the residuals. We derive the corresponding LM and LR tests for these three hypotheses and we compare their size and power performance using Monte Carlo experiments. Moreover, we derive the asymptotic distribution of the proposed LM tests.

2 A Generalized Model

Econometric models for panel data with spatial error processes have been proposed by Anselin (1988), Baltagi, Song, and Koh (2003), Kapoor, Kelejian, and Prucha (2007) and Anselin, Le Gallo, and Jayet (2006), to mention

a few. A generalized spatial panel data model that encompasses these previous specifications is given as follows:¹

$$\begin{aligned}
\mathbf{y}_t &= \mathbf{X}_t\boldsymbol{\beta} + \mathbf{u}_t, & t = 1, \dots, T \\
\mathbf{u}_t &= \mathbf{u}_1 + \mathbf{u}_{2t} \\
\mathbf{u}_1 &= \rho_1 \mathbf{W}\mathbf{u}_1 + \boldsymbol{\mu} \\
\mathbf{u}_{2t} &= \rho_2 \mathbf{W}\mathbf{u}_{2t} + \boldsymbol{\nu}_t,
\end{aligned}$$

where the $(N \times 1)$ vector \mathbf{y}_t includes the observations on the dependent variable at time t , with N denoting the number of unique cross-sectional units. The non-stochastic $(N \times K)$ matrix \mathbf{X}_t gives the observations at time t for a set of K exogenous variables, including the constant. $\boldsymbol{\beta}$ is the corresponding $(K \times 1)$ parameter vector. The disturbance term follows an error component model which involves the sum of two disturbances. The $(N \times 1)$ vector of random variables \mathbf{u}_1 captures the time-invariant unit-specific effects and therefore has no time subscript. The $(N \times 1)$ vector of the remainder disturbances \mathbf{u}_{2t} varies with time. Both \mathbf{u}_1 and \mathbf{u}_{2t} are spatially correlated with the same spatial weights matrix \mathbf{W} , but with different spatial autocorrelation parameters ρ_1 and ρ_2 , respectively. The $(N \times N)$ spatial weights matrix \mathbf{W} has zero diagonal elements and its entries are typically declining with distance. We further assume that the row and column sums of \mathbf{W} are uniformly bounded in absolute value and that ρ_r is bounded in absolute value, i.e., $|\rho_r| < \lambda_{\max}$ for $r = 1, 2$, where λ_{\max} is the largest absolute value of

¹To avoid index cluttering, we suppress the subscript indicating that the elements of the spatial weights matrix may depend on N and that the dependent variable and the disturbances form triangular arrays.

the eigenvalues of \mathbf{W} . Hence, the spatial weights matrix may be either row normalized or maximum row normalized (see Kelejian and Prucha, 2007). Further, the matrices $\mathbf{I}_N - \rho_r \mathbf{W}$ are assumed to be non-singular.

The elements of $\boldsymbol{\mu}$ are assumed to be independent across $i = 1, \dots, N$, and identically distributed as $N(0, \sigma_\mu^2)$. The elements of $\boldsymbol{\nu}_t$ are assumed to be independent across i and t and identically distributed as $N(0, \sigma_\nu^2)$. Also, the elements of $\boldsymbol{\mu}$ and $\boldsymbol{\nu}_t$ are assumed to be independent of each other. Appendix B provides a more detailed set of assumptions.

Stacking the cross-sections over time yields

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{u} & (1) \\ \mathbf{u} &= \mathbf{Z}_\mu \mathbf{u}_1 + \mathbf{u}_2 \\ \mathbf{u}_1 &= \rho_1 \mathbf{W} \mathbf{u}_1 + \boldsymbol{\mu} \\ \mathbf{u}_2 &= \rho_2 (\mathbf{I}_T \otimes \mathbf{W}) \mathbf{u}_2 + \boldsymbol{\nu}, \end{aligned}$$

where $\mathbf{y} = [\mathbf{y}'_1, \dots, \mathbf{y}'_T]'$, $\mathbf{X} = [\mathbf{X}'_1, \dots, \mathbf{X}'_T]'$, etc., so that the faster index is i and the slower index is t . The unit-specific errors \mathbf{u}_1 are repeated in all time periods using the $(NT \times N)$ selector matrix $\mathbf{Z}_\mu = \boldsymbol{\nu}_T \otimes \mathbf{I}_N$. $\boldsymbol{\nu}_T$ is a vector of ones of dimension T and \mathbf{I}_N is an identity matrix of dimension N .

This model encompasses both the KKP model, which assumes that $\rho_1 = \rho_2$, and the Anselin model, which assumes that $\rho_1 = 0$. If $\rho_1 = \rho_2 = 0$, i.e., there is no spatial correlation, this model reduces to the familiar random effects (RE) panel data model; see Baltagi (2005).

Let $\mathbf{A} = (\mathbf{I}_N - \rho_1 \mathbf{W})$ and $\mathbf{B} = (\mathbf{I}_N - \rho_2 \mathbf{W})$, then, under the present

assumptions we have

$$\begin{aligned}\mathbf{u}_1 &= \mathbf{A}^{-1}\boldsymbol{\mu} \sim N(\mathbf{0}, \sigma_\mu^2(\mathbf{A}'\mathbf{A})^{-1}) \\ \mathbf{u}_2 &= (\mathbf{I}_T \otimes \mathbf{B}^{-1})\boldsymbol{\nu} \sim N(\mathbf{0}, \sigma_\nu^2(\mathbf{I}_T \otimes (\mathbf{B}'\mathbf{B})^{-1})).\end{aligned}\quad (2)$$

The variance-covariance matrix of the spatial random effects panel data model is given by

$$\begin{aligned}\boldsymbol{\Omega}_u &= E(\mathbf{u}\mathbf{u}') = E[(\mathbf{Z}_\mu\mathbf{u}_1 + \mathbf{u}_2)(\mathbf{Z}_\mu\mathbf{u}_1 + \mathbf{u}_2)'] \\ &= \sigma_\mu^2(\mathbf{J}_T \otimes (\mathbf{A}'\mathbf{A})^{-1}) + \sigma_\nu^2(\mathbf{I}_T \otimes (\mathbf{B}'\mathbf{B})^{-1}) \\ &= \bar{\mathbf{J}}_T \otimes [T\sigma_\mu^2(\mathbf{A}'\mathbf{A})^{-1} + \sigma_\nu^2(\mathbf{B}'\mathbf{B})^{-1}] + \sigma_\nu^2(\mathbf{E}_T \otimes (\mathbf{B}'\mathbf{B})^{-1}) = \sigma_\nu^2\boldsymbol{\Sigma}_u.\end{aligned}\quad (3)$$

This uses the fact that $E[\mathbf{u}_1\mathbf{u}_2'] = \mathbf{0}$ since $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ are assumed to be independent. Note that $\mathbf{Z}_\mu\mathbf{Z}_\mu' = \mathbf{J}_T \otimes \mathbf{I}_N$, where \mathbf{J}_T is a matrix of ones of dimension T . Let $\mathbf{E}_T = \mathbf{I}_T - \bar{\mathbf{J}}_T$, where $\bar{\mathbf{J}}_T = \mathbf{J}_T/T$ is the averaging matrix, the last equality replaces \mathbf{J}_T by $T\bar{\mathbf{J}}_T$ and \mathbf{I}_T by $\mathbf{E}_T + \bar{\mathbf{J}}_T$. It is easy to show that the inverse of the $(NT \times NT)$ matrix $\boldsymbol{\Omega}_u$ can be obtained from the inverse of matrices of smaller dimension $(N \times N)$ as follows: $\boldsymbol{\Omega}_u^{-1} = (\bar{\mathbf{J}}_T \otimes (T\sigma_\mu^2(\mathbf{A}'\mathbf{A})^{-1} + \sigma_\nu^2(\mathbf{B}'\mathbf{B})^{-1})^{-1}) + \frac{1}{\sigma_\nu^2}(\mathbf{E}_T \otimes \mathbf{B}'\mathbf{B}) = \frac{1}{\sigma_\nu^2}\boldsymbol{\Sigma}_u^{-1}$, where

$$\boldsymbol{\Sigma}_u^{-1} = (\bar{\mathbf{J}}_T \otimes (T\frac{\sigma_\mu^2}{\sigma_\nu^2}(\mathbf{A}'\mathbf{A})^{-1} + (\mathbf{B}'\mathbf{B})^{-1})^{-1}) + (\mathbf{E}_T \otimes \mathbf{B}'\mathbf{B}).$$

Also, $\det[\boldsymbol{\Omega}_u] = \det[T\sigma_\mu^2(\mathbf{A}'\mathbf{A})^{-1} + \sigma_\nu^2(\mathbf{B}'\mathbf{B})^{-1}] \det[\sigma_\nu^2(\mathbf{B}'\mathbf{B})^{-1}]^{T-1}$. Under the assumption of normality of the disturbances, the log-likelihood function of the general model is given by

$$\begin{aligned}L(\boldsymbol{\beta}, \boldsymbol{\theta}) &= -\frac{NT}{2} \ln 2\pi - \frac{1}{2} \ln \det[T\sigma_\mu^2(\mathbf{A}'\mathbf{A})^{-1} + \sigma_\nu^2(\mathbf{B}'\mathbf{B})^{-1}] \\ &\quad - \frac{T-1}{2} \ln \det[\sigma_\nu^2(\mathbf{B}'\mathbf{B})^{-1}] - \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Omega}_u^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}),\end{aligned}\quad (4)$$

where $\boldsymbol{\theta} = (\sigma_\nu^2, \sigma_\mu^2, \rho_1, \rho_2)$. The maximum likelihood estimates are obtained by maximizing the log-likelihood function numerically using a constrained quasi-Newton method with the constraints as implied by our assumptions.²

The hypotheses under consideration in this paper are the following:

(1) H_0^A : $\rho_1 = \rho_2 = 0$, and the alternative H_1^A is that at least one component is not zero. The restricted model is the standard random effects (RE) panel data model with no spatial correlation, see Baltagi (2005).

(2) H_0^B : $\rho_1 = 0$, and the alternative is H_1^B : $\rho_1 \neq 0$. The restricted model is the Anselin (1988) spatial panel model with random effects. In fact, the restricted log-likelihood function reduces to the one considered by Anselin (1988, p.154).

(3) H_0^C : $\rho_1 = \rho_2 = \rho$ and the alternative is H_1^C : $\rho_1 \neq \rho_2$. The restricted model is the KKP spatial panel model with random effects.

In the next subsections, we derive the corresponding LM tests for these hypotheses and we compare their performance with the corresponding LR tests using Monte Carlo experiments.³ Appendix A describes some general results used to derive the score and information matrix for these alternative models; Appendix B proves the consistency of the ML estimates of the general model; while Appendices C-E provide the derivations of the large sample

²The numerical maximization procedure can be simplified, if one concentrates the likelihood with respect to $\boldsymbol{\beta}$ and σ_ν^2 . However, our optimization for the Monte Carlo simulation using MATLAB were quite fast using the constrained quasi-Newton method. Appendix F describes some details on the numerical optimization procedure.

³LM tests for spatial models are surveyed in Anselin (1988, 2001) and Anselin and Bera (1998), to mention a few. For a joint test for the absence of spatial correlation and random effects in a panel data model, see Baltagi, Song, and Koh (2003).

distributions of these LM tests.

2.1 LM and LR Tests for $H_0^A : \rho_1 = \rho_2 = 0$

The ML estimates under H_0^A are labeled by a tilde and the corresponding restricted parameter vector is indexed by A . The joint LM test statistic for the null hypothesis of no spatial correlation, $H_0^A : \rho_1 = \rho_2 = 0$, is derived in Appendix C and it is given by

$$LM_A = \frac{1}{2b_A\tilde{\sigma}_1^4}\tilde{G}_A^2 + \frac{1}{2b_A(T-1)\tilde{\sigma}_\nu^4}\tilde{M}_A^2, \quad (5)$$

where $\tilde{\sigma}_1^2 = T\tilde{\sigma}_\mu^2 + \tilde{\sigma}_\nu^2$, $b_A = \text{tr}[(\mathbf{W}' + \mathbf{W})^2]$, $\tilde{G}_A = \tilde{\mathbf{u}}'[\bar{\mathbf{J}}_T \otimes (\mathbf{W}' + \mathbf{W})]\tilde{\mathbf{u}}$, and $\tilde{M}_A = \tilde{\mathbf{u}}'[\mathbf{E}_T \otimes (\mathbf{W}' + \mathbf{W})]\tilde{\mathbf{u}}$. In this case, $\tilde{\mathbf{u}} = \mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}$ denotes the vector of the estimated residuals under H_0^A . The restricted model is the simple random effects (RE) panel data model without any spatial autocorrelation. In fact, $\tilde{\sigma}_\nu^2 = \frac{\tilde{\mathbf{u}}'(\mathbf{E}_T \otimes \mathbf{I}_N)\tilde{\mathbf{u}}}{N(T-1)}$ and $\tilde{\sigma}_1^2 = \frac{\tilde{\mathbf{u}}'(\bar{\mathbf{J}}_T \otimes \mathbf{I}_N)\tilde{\mathbf{u}}}{N}$. Under H_0^A , the LM_A statistic is asymptotically distributed as χ_2^2 as shown in Appendix C.

One can also derive the corresponding LR test for $H_0^A : \rho_1 = \rho_2 = 0$ as

$$LR_A = 2(L_G - L_A),$$

using the maximized log-likelihood of the general model denoted by L_G and the maximized log-likelihood under H_0^A :

$$L_A = -\frac{NT}{2} \ln 2\pi\tilde{\sigma}_\nu^2 - \frac{N}{2} \ln \frac{\tilde{\sigma}_1^2}{\tilde{\sigma}_\nu^2} - \frac{1}{2}\tilde{\mathbf{u}}'\tilde{\boldsymbol{\Omega}}_u^{-1}\tilde{\mathbf{u}}.$$

This test statistic is likewise asymptotically distributed as χ_2^2 .

2.2 LM and LR Tests for $H_0^B : \rho_1 = 0$

Under $H_0^B : \rho_1 = 0$, the restricted model is the spatial panel data model with random effects described in Anselin (1988). The corresponding LM test for H_0^B is a conditional test for zero spatial correlation in the individual effects, allowing for the possibility of spatial correlation in the remainder error term, i.e., $\rho_2 \neq 0$. Appendix D gives the formal derivation of this LM statistic. In fact, under H_0^B , the information matrix is block-diagonal with the lower block being independent of $\boldsymbol{\beta}$. Let \mathbf{d}_θ be the (4×1) score vector referring to the parameter vector $\boldsymbol{\theta} = (\sigma_\mu^2, \sigma_\nu^2, \rho_1, \rho_2)$ and denote the 4×4 lower block of the information matrix by \mathbf{J}_θ . The ML estimates under H_0^B are labeled by a hat. The corresponding estimated residuals are then $\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$. The LM test for H_0^B makes use of the estimated score $\hat{\mathbf{d}}_\theta = [0, 0, \hat{d}_{\rho_1}, 0]'$ with

$$\begin{aligned}\hat{d}_{\rho_1} &= \left. \frac{\partial L}{\partial \rho_1} \right|_{H_0^B} = -\frac{1}{2}T\hat{\sigma}_\mu^2 \text{tr}[\hat{\mathbf{C}}_1 \mathbf{C}_2] + \frac{1}{2}\hat{\sigma}_\mu^2 \hat{\mathbf{u}}'(\mathbf{J}_T \otimes \hat{\mathbf{C}}_1 \mathbf{C}_2 \hat{\mathbf{C}}_1) \hat{\mathbf{u}} \\ &= \frac{1}{2}T\hat{\sigma}_\mu^2 [(\hat{\mathbf{u}}' \hat{\mathbf{G}}_B \hat{\mathbf{u}}) - \hat{g}_B],\end{aligned}$$

where $\hat{\mathbf{C}}_1 = [T\hat{\sigma}_\mu^2 \mathbf{I}_N + \hat{\sigma}_\nu^2 (\hat{\mathbf{B}}' \hat{\mathbf{B}})^{-1}]^{-1}$ and $\mathbf{C}_2 = (\mathbf{W}' + \mathbf{W})$, $\hat{\mathbf{G}}_B = (\bar{\mathbf{J}}_T \otimes \hat{\mathbf{C}}_1 \mathbf{C}_2 \hat{\mathbf{C}}_1)$, and $\hat{g}_B = \text{tr}[\hat{\mathbf{C}}_1 \mathbf{C}_2]$. An estimate of the lower (4×4) block of the information matrix $\hat{\mathbf{J}}_\theta$ under H_0^B is given by

$$\hat{\mathbf{J}}_\theta \Big|_{H_0^B} = \begin{bmatrix} \frac{1}{2} \text{tr}[\hat{\mathbf{C}}_3^2] + \frac{N(T-1)}{2\hat{\sigma}_\nu^4} & \frac{T}{2} \text{tr}[\hat{\mathbf{C}}_3 \hat{\mathbf{C}}_1] & \frac{T\hat{\sigma}_\mu^2}{2} \text{tr}[\hat{\mathbf{C}}_3 \hat{\mathbf{C}}_1 \mathbf{C}_2] & \frac{\hat{\sigma}_\nu^2}{2} \text{tr}[\hat{\mathbf{C}}_3 \hat{\mathbf{C}}_1 \hat{\mathbf{C}}_5] + \frac{(T-1)}{2\hat{\sigma}_\nu^2} \text{tr}[\hat{\mathbf{C}}_4] \\ \frac{T}{2} \text{tr}[\hat{\mathbf{C}}_3 \hat{\mathbf{C}}_1] & \frac{T^2}{2} \text{tr}[\hat{\mathbf{C}}_1^2] & \frac{T^2 \hat{\sigma}_\mu^2}{2} \text{tr}[\hat{\mathbf{C}}_1^2 \mathbf{C}_2] & \frac{T\hat{\sigma}_\nu^2}{2} \text{tr}[\hat{\mathbf{C}}_1^2 \hat{\mathbf{C}}_5] \\ \frac{T\hat{\sigma}_\mu^2}{2} \text{tr}[\hat{\mathbf{C}}_3 \hat{\mathbf{C}}_1 \mathbf{C}_2] & \frac{T^2 \hat{\sigma}_\mu^2}{2} \text{tr}[\hat{\mathbf{C}}_1^2 \mathbf{C}_2] & \frac{T^2 \hat{\sigma}_\mu^4}{2} \text{tr}[(\hat{\mathbf{C}}_1 \mathbf{C}_2)^2] & \frac{T\hat{\sigma}_\mu^2 \hat{\sigma}_\nu^2}{2} \text{tr}[\hat{\mathbf{C}}_1 \mathbf{C}_2 \hat{\mathbf{C}}_1 \hat{\mathbf{C}}_5] \\ \frac{\hat{\sigma}_\nu^2}{2} \text{tr}[\hat{\mathbf{C}}_3 \hat{\mathbf{C}}_1 \hat{\mathbf{C}}_5] + \frac{(T-1)}{2\hat{\sigma}_\nu^2} \text{tr}[\hat{\mathbf{C}}_4] & \frac{T\hat{\sigma}_\nu^2}{2} \text{tr}[\hat{\mathbf{C}}_1^2 \hat{\mathbf{C}}_5] & \frac{T\hat{\sigma}_\mu^2 \hat{\sigma}_\nu^2}{2} \text{tr}[\hat{\mathbf{C}}_1 \mathbf{C}_2 \hat{\mathbf{C}}_1 \hat{\mathbf{C}}_5] & \frac{\hat{\sigma}_\nu^4}{2} \text{tr}[(\hat{\mathbf{C}}_1 \hat{\mathbf{C}}_5)^2] + \frac{(T-1)}{2} \text{tr}[\hat{\mathbf{C}}_4^2] \end{bmatrix},$$

where $\hat{\mathbf{C}}_3 = (\hat{\mathbf{B}}' \hat{\mathbf{B}})^{-1} \hat{\mathbf{C}}_1$, $\hat{\mathbf{C}}_4 = (\mathbf{W}' \hat{\mathbf{B}} + \hat{\mathbf{B}}' \mathbf{W})(\hat{\mathbf{B}}' \hat{\mathbf{B}})^{-1}$ and $\hat{\mathbf{C}}_5 = (\hat{\mathbf{B}}' \hat{\mathbf{B}})^{-1} \hat{\mathbf{C}}_4$.

The LM test for H_0^B has no simple closed form representation and it is cal-

culated as

$$LM_B = \widehat{\mathbf{d}}'_\theta \widehat{\mathbf{J}}_\theta^{-1} \widehat{\mathbf{d}}_\theta = \widehat{d}_{\rho_1}^2 \widehat{\mathbf{J}}_{33}^{-1}, \quad (6)$$

where $\widehat{\mathbf{J}}_{33}^{-1}$ is the (3,3) element of the inverse of the estimated information matrix $\widehat{\mathbf{J}}_\theta^{-1}$ under H_0^B . This test statistic is shown to be asymptotically distributed as χ_1^2 .

In Appendix D, we follow Moulton and Randolph (1989) in deriving the asymptotic distribution of an alternative LM test statistic based on the standardized score under H_0^B , which we denote by LM'_B .⁴ This results in an alternative closed form expression for this LM statistic, namely,

$$LM'_B = \frac{(\widehat{\mathbf{u}}' \widehat{\mathbf{G}}_B \widehat{\mathbf{u}} - \widehat{g}_B)^2}{2\widehat{b}_B}, \quad (7)$$

where $\widehat{b}_B = tr[(\widehat{\mathbf{C}}_1 \widehat{\mathbf{C}}_2)^2]$. In Appendix D, we show that this test statistic is asymptotically distributed as χ_1^2 . LM'_B is a simple and practical alternative to LM_B which performs just as well in the Monte Carlo experiments.

The corresponding LR test is based upon the maximized log-likelihood under H_0^B :

$$L_B = -\frac{NT}{2} \ln 2\pi \widehat{\sigma}_\nu^2 - \frac{1}{2} \ln \det(\widehat{\mathbf{C}}_1) + \frac{T-1}{2} \ln \det(\widehat{\mathbf{B}}' \widehat{\mathbf{B}}) - \frac{1}{2} \widehat{\mathbf{u}}' \widehat{\mathbf{\Omega}}_u^{-1} \widehat{\mathbf{u}}. \quad (8)$$

⁴Moulton and Randolph (1989, p. 687) find that "the asymptotic unit normal distribution poorly approximates the finite sample distribution of the LM statistic" in the one-way random effects model. Therefore, they propose an alternative LM statistic, which is based on the standardized score. The suggested standardization does not affect the test when exact critical values are used, but may improve the approximation of asymptotic critical values.

This restricted log-likelihood is the same as that given by Anselin (1988, p. 154).

2.3 LM and LR Tests for $H_0^C : \rho_1 = \rho_2 = \rho$

Under $H_0^C : \rho_1 = \rho_2 = \rho$, the true model is the one suggested by Kapoor, Kelejian, and Prucha (2007). In this case, $\mathbf{B} = \mathbf{A}$ and the parameter estimates under H_0^C are labeled by a bar. The corresponding estimated residuals are given by $\bar{\mathbf{u}} = \mathbf{y} - \mathbf{X}\bar{\boldsymbol{\beta}}$. The score and the information matrix needed for this test are derived in Appendix E. The joint LM test statistic for H_0^C is given by

$$LM_C = \frac{T}{2\bar{b}_C(T-1)\bar{\sigma}_1^4} \bar{G}_C^2, \quad (9)$$

with $\bar{G}_C = \bar{\mathbf{u}}'(\bar{\mathbf{J}}_T \otimes \bar{\mathbf{F}})\bar{\mathbf{u}} - \bar{\sigma}_1^2 \text{tr}[\bar{\mathbf{D}}]$, $\bar{\mathbf{F}} = \mathbf{W}'\bar{\mathbf{A}} + \bar{\mathbf{A}}'\mathbf{W}$ and $\bar{\mathbf{D}} = \bar{\mathbf{F}}(\bar{\mathbf{A}}'\bar{\mathbf{A}})^{-1}$. Also, $\bar{b}_C = \text{tr}[\bar{\mathbf{D}}^2] - (\text{tr}[\bar{\mathbf{D}}])^2/N$, $\bar{\sigma}_1^2 = \frac{\bar{\mathbf{u}}'[\bar{\mathbf{J}}_T \otimes (\bar{\mathbf{A}}'\bar{\mathbf{A}})]\bar{\mathbf{u}}}{N}$ and $\bar{\sigma}_\nu^2 = \frac{\bar{\mathbf{u}}'[\mathbf{E}_T \otimes (\bar{\mathbf{A}}'\bar{\mathbf{A}})]\bar{\mathbf{u}}}{N(T-1)}$. Under H_0^C , the LM_C statistic is supposed to be asymptotically distributed as χ_1^2 . In Appendix E, we follow Moulton and Randolph (1989) in deriving the asymptotic distribution of a slightly modified LM test statistic based on the standardized score, which we denote by LM'_C :

$$LM'_C = \frac{1}{2\bar{b}'_C\bar{\sigma}_1^4} \bar{G}_C^2$$

with $\bar{b}'_C = \text{tr}[\bar{\mathbf{D}}^2]$. The Monte Carlo simulations indicate that the normalized LM'_C performs nearly as well as LM_C in small samples.

The LR test is based on the following maximized log-likelihood under H_0^C :

$$L_C = -\frac{NT}{2} \ln 2\pi\bar{\sigma}_\nu^2 - \frac{N}{2} \ln\left(\frac{\bar{\sigma}_1^2}{\bar{\sigma}_\nu^2}\right) + \frac{T}{2} \ln \det(\bar{\mathbf{A}}'\bar{\mathbf{A}}) - \frac{1}{2} \bar{\mathbf{u}}'\bar{\boldsymbol{\Omega}}_u^{-1}\bar{\mathbf{u}}.$$

Kapoor, Kelejian, and Prucha (2007) consider a generalized method of moments estimator, rather than MLE, for their spatial random effects panel data model. Nevertheless, L_C is the maximized log-likelihood for the KKP model with normal disturbances.

3 Monte Carlo Results

In the Monte Carlo analysis, we use a simple panel data model that includes one explanatory variable and a constant ($K = 2$)

$$y_{it} = \beta_0 + \beta_1 x_{it} + u_{it}, \quad i = 1, \dots, N \text{ and } t = 1, \dots, T,$$

where $\beta_0 = 5$ and $\beta_1 = 0.5$. x_{it} is generated by $x_{it} = \zeta_i + z_{it}$, where $\zeta_i \sim i.i.d. U[-7.5, 7.5]$ and $z_{it} \sim i.i.d. U[-5, 5]$ with $U[a, b]$ denoting the uniform distribution on the interval $[a, b]$. The individual-specific effects are drawn from a normal distribution so that $\mu_i \sim i.i.d. N(0, 20\theta)$, while for the remainder error we assume $\nu_{it} \sim i.i.d. N(0, 20(1 - \theta))$ with $0 < \theta < 1$. $\theta = \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_\nu^2}$ is the proportion of the total variance due to the heterogeneity of the individual-specific effects. This implies that $\sigma_\mu^2 + \sigma_\nu^2 = 20$.

We generate the spatial weights matrix by allocating observations randomly on a grid of $2N$ squares. Consequently, as the number of observations N increases, the number of squares in the grid grows larger, too. The probability that an observation is located on a particular coordinate is equal for all coordinates on the grid. This results in an irregular lattice, where each observation possesses 3 neighbors on average. The spatial weighting scheme is based on the Queens design and the corresponding spatial weights matrix is normalized so that each row sums to one.

The parameters ρ_1 and ρ_2 vary over the set $\{-0.8, -0.5, -0.2, 0, 0.2, 0.5, 0.8\}$. The cross-sectional and time dimensions are $N = 50, 100$ and $T = 3, 5, 10$, respectively. Lastly, the proportion of the variance due to the random individual effects takes the values $\theta = 0.25, 0.50, 0.75$. In total, this gives 882 experiments. For each experiment, we calculate the three LM and LR tests as derived above, using 2000 replications.⁵

===== Tables 1-3 =====

Table 1 reports the frequency of rejections for $N = 50, T = 5$, and $\theta = 0.5$ in 2000 replications. This means that $\sigma_\mu^2 = \sigma_\nu^2 = 10$. The size of each test is denoted in bold figures and is not statistically different from the 5% nominal size. The only exception where the LM test might be undersized is for the KKP model, for high absolute values of ρ_1 and ρ_2 , both equal to 0.8. The size adjusted power⁶ of the LR and LM tests is reasonably high for all three hypotheses considered. The performance of the LM test is almost the same as that of the LR test, except for a few cases. For $H_0^A : \rho_1 = \rho_2 = 0$, when $\rho_1 = -0.5$ and $\rho_2 = 0$, the size adjusted power of the LM test is 61.4% as compared to 64.6% for LR. At $\rho_1 = 0.5$ and $\rho_2 = 0$, the size adjusted power of the LM test is 70% as compared to 66.4% for LR. Similarly, for $H_0^B : \rho_1 = 0$, when $\rho_1 = -0.5$ and $\rho_2 = 0$, the size adjusted power of the LM

⁵In a few cases, we got negative LR test statistics due to numerical imprecision. These cases occur mainly with the Anselin model at $\rho_1 = 0$. However, this happened in less than 0.5 percent of the Monte Carlo experiments. We drop the corresponding experiments in the subsequent calculations of the size and power of the tests.

⁶The size corrected critical level for the test is inferred from the empirical distribution of the test statistic in the Monte Carlo experiments, so that the rejection region under the empirical distribution has the correct nominal size.

test is 70.2% as compared to 72.9% for LR. At $\rho_1 = 0.5$ and $\rho_2 = 0$, the size adjusted power of the LM test is 76.7% as compared to 74.6% for LR. For $H_0^C : \rho_1 = \rho_2 = \rho$, when $\rho_1 = -0.5$ and $\rho_2 = 0$, the size adjusted power of the LM test is 66.1% as compared to 68.5% for LR. At $\rho_1 = 0.5$ and $\rho_2 = 0$, the size adjusted power of the LM test is 70.6% as compared to 65% for LR.

Table 1 also reports the large sample approximations of LM_B and LM_C , namely, LM'_B and LM'_C , respectively. These results indicate that the large sample approximations are accurate for small and medium absolute values of ρ_1 (Anselin model) and of $\rho_1 = \rho_2$ in the KKP model. However, the tests tend to be undersized whenever ρ_1 or ρ_2 is large in absolute value.

Tables 2 and 3 repeat the same experiments but now for $\theta = 0.25$ and 0.75 , respectively. These tables show that as we increase θ , we increase the power of these tests. In fact, the power of all three tests is higher, the higher the variance of the individual-specific effect as a proportion of the total variance. For example, for $H_0^A : \rho_1 = \rho_2 = 0$, when $\rho_1 = -0.5$ and $\rho_2 = 0$, the size adjusted power of the LM test increases from 61.4% for $\theta = 0.5$ (in Table 1) to 68% for $\theta = 0.75$ (in Table 3), while the size adjusted power of the LR test increases from 64.6% to 74.8%. Similarly, when $\rho_1 = 0.5$ and $\rho_2 = 0$, the size adjusted power of the LM test increases from 70% for $\theta = 0.5$ to 78.4% for $\theta = 0.75$, while the size adjusted power of the LR test increases from 66.4% to 77.4%. For $H_0^B : \rho_1 = 0$, when $\rho_1 = -0.5$ and $\rho_2 = 0$, the size adjusted power of the LM test increases from 70.2% for $\theta = 0.5$ to 81% for $\theta = 0.75$, while the size adjusted power of the LR test increases from 72.9% to 83.4%. At $\rho_1 = 0.5$ and $\rho_2 = 0$, the size adjusted power of the LM test increases from 76.7% for $\theta = 0.5$ to 86.6% for $\theta = 0.75$, while the size adjusted power

of the LR test increases from 74.6% to 84.9% for LR. For $H_0^C : \rho_1 = \rho_2 = \rho$, when $\rho_1 = -0.5$ and $\rho_2 = 0$, the size adjusted power of the LM test increases from 66.1% for $\theta = 0.5$ to 73% for $\theta = 0.75$, while the size adjusted power of the LR test increases from 68.5% to 74.8%. At $\rho_1 = 0.5$ and $\rho_2 = 0$, the size adjusted power of the LM test increases from 70.6% for $\theta = 0.5$ to 80.4% for $\theta = 0.75$, while the size adjusted power of the LR test increases from 65% to 77.3%.

Things also improve if the number of observations increases. The increase in power is larger when we double N from 50 to 100 as compared to doubling T from 5 to 10.⁷ We conclude that the three LM and LR tests perform reasonably well in testing the restrictions underlying the simple random effects model without spatial correlation, the Anselin model and the KKP model in small and medium sized samples.

Figures 1-4 plot the size adjusted power for the various hypotheses considered. In Figure 1, the pure random effects model is true, whereas in Figure 2, the Anselin model is true. In Figures 3 and 4, the KKP-type model is true with different values for the common ρ .

===== Figures 1-2 =====

Let us start with a comparison of the panels given in Figure 1, which assumes that the random effects model is true ($\rho_1 = \rho_2 = 0$). On the left hand side, we plot the size adjusted power of the LM test for deviations of

⁷We do not include the corresponding Tables for $(N = 50, T = 10)$ and $(N = 100, T = 5)$, for $\theta = 0.25, 0.50$, and 0.75 , in order to save space. However, these tables are available upon request from the authors. Below, we summarize the corresponding information by means of size adjusted power plots.

ρ_1 from 0, maintaining that $\rho_2 = 0$. On the right hand side it is the other way around. Observe that the power of the LM test is higher for deviations of ρ_2 from 0 as compared to deviations of ρ_1 from 0. Keep in mind that the estimates of ρ_2 are based on NT observations, while those of ρ_1 rely on only N observations. The top two panels show that the power increases for deviations in ρ_1 as θ increases. However, for deviations in ρ_2 , the power of the test is insensitive to θ . The two panels at the center of Figure 1 illustrate that both the size and the power of the LM test improve as the sample size increases, especially as N becomes larger. A comparison of the two panels at the center with those at the bottom of Figure 1 provides information on the interaction of sample size (N, T) and the relative importance of θ . It is obvious that for deviations of ρ_1 from 0 (on the left), the power improves with N , especially as θ increases.

Figure 2 assumes that the Anselin-type process of the error term is the true model ($\rho_1 = 0$). One important difference when compared to Figure 1 is that ρ_2 is now a nuisance parameter. The qualitative effects of an increase in N, T , and θ are similar to those in Figure 1 on the left hand side. The right hand side panels of Figure 2 show that the size adjusted power of the LM test is lower if ρ_2 is high (0.5 compared to 0), especially for low θ (0.25 compared to 0.75).

===== Figures 3-4 =====

Figures 3 and 4 assume that the KKP model is the true one. Note that an assessment of the performance of the LM test is different here, since the KKP model assumes that $\rho_1 = \rho_2$. The null hypothesis in Figure 3 is $\rho_1 = \rho_2 = 0.2$

and the one in Figure 4 is $\rho_1 = \rho_2 = 0.5$. The major difference between the two figures is that assuming a null that is different from $\rho_1 = \rho_2 = 0$ shifts the size adjusted power function and renders it skewed to the right. Otherwise, the conclusions regarding the impact of θ , N , and T are qualitatively similar to those of the random effects model. A major difference from the random effects model is that for the KKP model the power is lower in the ρ_2 direction, especially for small θ .

3.1 Robustness Checks

We also assess the robustness of the proposed LM tests with respect to (i) non-normal errors and (ii) the specification of the spatial weighting matrix. To compare the simulated power functions for normal vs. non-normal errors, we generated the remainder error term first as $\nu_{it} \sim t(5)$ and normalized its variance to 10. Hence, $\theta = 0.5$ holds in this case and the results are comparable to the basic Monte Carlo set-up defined above. This implies that the distribution of the remainder error exhibits heavier tails as compared to the normal distribution but it is still symmetric. Second, we analyzed a skewed error distribution assuming ν_{it} follows a log-normal distribution with variance 10, i.e., $\nu_{it} = \sqrt{10}(e^\xi - e^{0.5})/\sqrt{e^2 - e^1}$, where $\xi \sim N(0, 1)$. For $N = 50$ and $T = 5$, the Monte Carlo experiments show that there are minor changes in the size adjusted power curves under both error distributions. This holds true for all LM tests considered. The power figures are available upon request from the authors.

===== Table 4 =====

The non-normality of the remainder error, however, does affect the size of the tests. In Table 4, we focus on the size of the LM and LR tests under alternative distributional assumptions of the error term for $N = 50$, $T = 5$ and $\theta = 0.5$. In the first pair of columns we give the true parameters ρ_1 , ρ_2 , the second pair of columns summarizes the size of the tests under the assumption that $\nu_{it} \sim t(5)$, in the third pair of columns we assume that ν_{it} follows a log-normal distribution with variance 10. It turns out that both the LM tests and the LR tests are fairly insensitive to the chosen alternative assumptions about the distribution of the disturbances at intermediate levels of ρ_1 and ρ_2 . However, the LM tests tend to be somewhat more undersized than the LR tests, especially for $\rho_1 = \rho_2 = 0.8$. With the caveat of the limited experiments we performed, this finding suggests that the LM tests considered are fairly robust to deviations from the assumption of a normally distributed error term.

===== Figure 5 =====

Figure 5 investigates the extent to which the specification of the spatial weighting scheme matters for the size and power of the tests considered. We generated an alternative spatial weighting matrix allowing for a more densely populated grid. In particular, we randomly allocated the observations on the grid so that there are 5 rather than 3 neighbors per observation on average. As expected, the power of the tests is somewhat lower in this case, but still big enough to detect relevant deviations from the null.

4 Conclusions

The recent literature on first-order spatially autocorrelated residuals (SAR(1)) with panel data distinguishes between two data generating processes of the error term. One process described in Anselin (1988) and Anselin, Le Gallo and Jayet (2006) assumes that only the remainder error component is spatially correlated. In an alternative process put forward by Kapoor, Kelejian, and Prucha (2007) both the individual and remainder components of the disturbances are characterized by the same spatial autocorrelation pattern. This paper formulates a SAR(1) process of the residuals with panel data that encompasses these two processes. In particular, this paper derives three LM tests based upon the more general model, testing its restricted counterparts: the Anselin model, the Kapoor, Kelejian, and Prucha model, and the random effects model without spatial correlation. For the latter two tests, closed-form expressions for the LM statistics can be obtained.

Our Monte Carlo study assesses the small sample performance of the derived tests. We find that the tests are properly sized and powerful even in relatively small samples. The LM tests are easy to calculate and their power is reasonably high for all three tests considered. The power of these LM tests matches that of the corresponding LR tests except in few cases. In general, the power of the tests increases with the relative importance of the individual effects' variance as a proportion of the total variance, as well as with increasing N and T . They are robust to non-normality of the error term and sensitive to the specification of the weight matrix. Hence, these LM and LR tests are recommended for the applied researcher to test the restrictions imposed by the RE model with no spatial correlation, the Anselin model,

and the Kapoor, Kelejian, and Prucha model.

References

- Abadir, K.M. and Magnus, J.R., 2005. Matrix algebra. Cambridge University Press. Cambridge.
- Anselin, L., 1988. Spatial econometrics: methods and models. Kluwer Academic Publishers, Dordrecht.
- Anselin, L., 2001. Rao's score tests in spatial econometrics. *Journal of Statistical Planning and Inference* 97, 113-139.
- Anselin, L. and A.K. Bera, 1998. Spatial dependence in linear regression models with an introduction to spatial econometrics. In A. Ullah and Giles, H. (eds.), *Handbook of Applied Economic Statistics*, Marcel Dekker, New York.
- Anselin, L., J. Le Gallo, and H. Jayet, 2006. Spatial panel econometrics, in L. Mátyás and Sevestre, P. (eds.), *The econometrics of panel data: fundamentals and recent developments in theory and practice*, Springer, New York, ch. 18, forthcoming.
- Baltagi, B.H., 2005. *Econometric analysis of panel data*. Wiley, Chichester.
- Baltagi, B.H., S.H. Song, and W. Koh, 2003. Testing panel data models with spatial error correlation. *Journal of Econometrics* 117, 123-150.
- Hartley, H.O. and J.N.K. Rao, 1967. Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika* 54, 93-108.
- Harville, D.A., 1977. Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* 72, 320-338.

- Hemmerle, W.J. and H.O. Hartley, 1973. Computing maximum likelihood estimates for the mixed A.O.V. model using the W-transformation. *Technometrics* 15, 819-831.
- Horn, R. and C. Johnson, 1985. *Matrix Analysis*. Cambridge University Press, Cambridge.
- Kapoor, M., H.H. Kelejian, and I.R. Prucha, 2007. Panel data models with spatially correlated error components. *Journal of Econometrics* 140, 97-130.
- Kelejian, H.H. and I.R. Prucha, 2001. On the asymptotic distribution of the Moran I test with applications. *Journal of Econometrics* 104, 219-257.
- Kelejian, H.H. and I.R. Prucha, 2007. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. Working paper, University of Maryland.
- Lee, L.-F., 2004a. Asymptotic distributions of quasi-maximum likelihood estimators for spatial econometric models. *Econometrica* 72, 1899-1926.
- Lee, L.-F., 2004b. Supplement to: Asymptotic distributions of quasi-maximum likelihood estimators for spatial econometric models, Ohio State University.
- Moulton, B.R. and W.C. Randolph, 1989. Alternative tests of the error components model. *Econometrica* 57, 685-693.
- Rao, R.C., 1973. *Linear statistical inference and its applications*. John Wiley & Sons, New Jersey.

Appendix A: Score and Information Matrix

For convenience, we reproduce the variance-covariance matrix of the general model given in (3):

$$\begin{aligned}\boldsymbol{\Omega}_u &= \bar{\mathbf{J}}_T \otimes [T\sigma_\mu^2(\mathbf{A}'\mathbf{A})^{-1} + \sigma_\nu^2(\mathbf{B}'\mathbf{B})^{-1}] + \sigma_\nu^2[\mathbf{E}_T \otimes (\mathbf{B}'\mathbf{B})^{-1}] \\ \boldsymbol{\Omega}_u^{-1} &= \bar{\mathbf{J}}_T \otimes [T\sigma_\mu^2(\mathbf{A}'\mathbf{A})^{-1} + \sigma_\nu^2(\mathbf{B}'\mathbf{B})^{-1}]^{-1} + \frac{1}{\sigma_\nu^2}(\mathbf{E}_T \otimes \mathbf{B}'\mathbf{B}),\end{aligned}$$

where $\mathbf{A} = (\mathbf{I}_N - \rho_1 \mathbf{W})$ and $\mathbf{B} = (\mathbf{I}_N - \rho_2 \mathbf{W})$.

Denote the vector of parameters of interest by $\boldsymbol{\theta} = (\sigma_\nu^2, \sigma_\mu^2, \rho_1, \rho_2)'$. Below, we can focus on the part of the information matrix corresponding to $\boldsymbol{\theta}$. The part of the information matrix corresponding to $\boldsymbol{\beta}$ can be ignored in computing the LM test statistics, since the information matrix is block-diagonal between $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$, and the first derivative with respect to $\boldsymbol{\beta}$ evaluated at the restricted MLE is zero.

First, we derive the score and the relevant information submatrix of the general model. These results are then used to test the three hypotheses of interest below. Hartley and Rao (1971) and Hemmerle and Hartley (1973) give a general useful formula that helps in obtaining the score:

$$\frac{\partial L}{\partial \theta_r} = -\frac{1}{2}tr \left(\boldsymbol{\Omega}_u^{-1} \frac{\partial \boldsymbol{\Omega}_u}{\partial \theta_r} \right) + \frac{1}{2} \mathbf{u}' \left(\boldsymbol{\Omega}_u^{-1} \frac{\partial \boldsymbol{\Omega}_u}{\partial \theta_r} \boldsymbol{\Omega}_u^{-1} \right) \mathbf{u}, \quad r = 1, \dots, 4. \quad (10)$$

Observe, that

$$\begin{aligned}\frac{\partial \boldsymbol{\Omega}_u}{\partial \sigma_\nu^2} &= \bar{\mathbf{J}}_T \otimes (\mathbf{B}'\mathbf{B})^{-1} + (\mathbf{E}_T \otimes (\mathbf{B}'\mathbf{B})^{-1}) = \mathbf{I}_T \otimes (\mathbf{B}'\mathbf{B})^{-1} \\ \frac{\partial \boldsymbol{\Omega}_u}{\partial \sigma_\mu^2} &= \bar{\mathbf{J}}_T \otimes T(\mathbf{A}'\mathbf{A})^{-1} \\ \frac{\partial \boldsymbol{\Omega}_u}{\partial \rho_1} &= \bar{\mathbf{J}}_T \otimes T\sigma_\mu^2(\mathbf{A}'\mathbf{A})^{-1}(\mathbf{W}' + \mathbf{W} - 2\rho_1 \mathbf{W}'\mathbf{W})(\mathbf{A}'\mathbf{A})^{-1} \\ \frac{\partial \boldsymbol{\Omega}_u}{\partial \rho_2} &= \mathbf{I}_T \otimes \sigma_\nu^2(\mathbf{B}'\mathbf{B})^{-1}(\mathbf{W}' + \mathbf{W} - 2\rho_2 \mathbf{W}'\mathbf{W})(\mathbf{B}'\mathbf{B})^{-1}.\end{aligned}$$

To derive the information submatrix we use the general differentiation result given in Harville (1977):

$$J_{rs} = E \left[-\frac{\partial^2 L}{\partial \theta_r \partial \theta_s} \right] = \frac{1}{2} \text{tr} \left[\boldsymbol{\Omega}_u^{-1} \frac{\partial \boldsymbol{\Omega}_u}{\partial \theta_r} \boldsymbol{\Omega}_u^{-1} \frac{\partial \boldsymbol{\Omega}_u}{\partial \theta_s} \right] \quad r, s = 1, \dots, 4.$$

Here, $\frac{\partial L}{\partial \theta_r}$ and J_{rs} are evaluated at the MLE estimates.

Appendix B: Identification and Consistency

In the sequel, we use subscript 0 to indicate true parameter values where necessary.

Assumptions⁸

A1 (random effects model): The model comprises unit-specific random effects denoted by the $(N \times 1)$ vector $\boldsymbol{\mu}$. The elements of $\boldsymbol{\mu}$ are assumed to be *i.i.d.* $N(0, \sigma_\mu^2)$ with $0 < \underline{c}_\mu < \sigma_\mu^2 < \bar{c}_\mu < \infty$. $\boldsymbol{\nu}$ is the vector of remainder errors and its elements are assumed to be *i.i.d.* $N(0, \sigma_\nu^2)$ with $0 < \underline{c}_\nu < \sigma_\nu^2 < \bar{c}_\nu < \infty$. The elements of $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ are assumed to be independent of each other.

A2 (spatial correlation): (i) Both \mathbf{u}_1 and \mathbf{u}_{2t} are spatially correlated with the same spatial weights matrix \mathbf{W} whose elements may depend on N . The $(N \times N)$ spatial weights matrix \mathbf{W} has zero diagonal elements. (ii) The row and column sums of \mathbf{W} are uniformly bounded in absolute value. (iii) ρ_r is bounded in absolute value, i.e., $|\rho_r| < \frac{1}{\lambda_{\max}}$ for $r = 1, 2$, where λ_{\max} denotes the largest absolute value of the eigenvalues of \mathbf{W} . (iv) The matrices

⁸To avoid index cluttering, we suppress the subscript indicating that the elements of the spatial weights matrix may depend on N and that the dependent variable and the disturbances form triangular arrays.

$\mathbf{I}_N - \rho_r \mathbf{W}$ are non-singular and their inverses have bounded row and column sums.

A3 (compactness of the parameter space): the parameter space Θ with elements $(\beta, \sigma_\mu^2, \sigma_\nu^2, \rho_1, \rho_2)$ is compact. The true parameter vector (indexed by 0) lies in the interior of Θ .

We note that Assumptions A1 and A2 imply that $\Xi = \{(\phi, \rho_1, \rho_2) | (\sigma_\mu^2, \sigma_\nu^2, \rho_1, \rho_2) \in \Theta\}$ with $\phi = \sigma_\mu^2 / \sigma_\nu^2$ is also compact. In the following, the elements of Ξ are denoted by the vector ϑ .

A4 (identification of ϑ): For every $\vartheta \neq \vartheta_0$:

$$-\frac{1}{2} \ln\left(\frac{1}{NT} \text{tr}[\Sigma_u(\vartheta_0) \Sigma_u(\vartheta)^{-1}]\right) - \frac{1}{2NT} \ln[\det \Sigma_u(\vartheta) / \det \Sigma_u(\vartheta_0)] < 0.$$

A5 (identification of β): The non-random matrix \mathbf{X} has rank $K < N$ and its elements are uniformly bounded constants for all N . Further, the non-random matrix $\lim_{N \rightarrow \infty} (\frac{1}{NT} \mathbf{X}' \Sigma_u(\vartheta_0)^{-1} \mathbf{X})$ is finite and non-singular and $\lim_{N \rightarrow \infty} (\frac{1}{NT} \Sigma_u(\vartheta_0)^{-1} \mathbf{X})$ is finite and has full column rank.

Consistency of the ML estimates under the general model.

In proving the consistency of MLE, we make use of the following Lemmas.

Lemma 1 *Under the maintained assumptions (i) the row and column sums of $(\mathbf{A}'\mathbf{A})^{-1}$ and $(\mathbf{B}'\mathbf{B})^{-1}$ are bounded in absolute value. (ii) The row and column sums of $\Sigma_u(\vartheta)$ and $\Sigma_u(\vartheta)^{-1}$ are bounded in absolute value.*

Proof. (i) By Assumption A2 the row and column sums of the matrices \mathbf{W} , \mathbf{A} , \mathbf{B} , \mathbf{A}^{-1} and \mathbf{B}^{-1} are bounded in absolute value. Since this property is preserved when multiplying matrices of proper dimension that have bounded row and column sums (see Kelejian and Prucha, 2001, p. 241f), one can conclude that the row and column sums of $(\mathbf{A}'\mathbf{A})^{-1}$ and $(\mathbf{B}'\mathbf{B})^{-1}$ are also bounded in absolute value, say, by constants c_A and c_B , respectively.

(ii) The row and column sums of $\Sigma_u(\boldsymbol{\vartheta})$ are bounded in absolute value, since $|\rho_r| < \frac{1}{\lambda_{\max}}$ by Assumption A2 and Ξ is compact by Assumption A3. To see this, denote the typical element of $\Sigma_u(\boldsymbol{\vartheta})$ by σ_{ij} . Then, $\max_i \sum_j \sigma_{ij} \leq T\phi c_A + c_B < \infty$ and $\max_j \sum_i \sigma_{ij} \leq T\phi c_A + c_B < \infty$. Since $\Sigma_u(\boldsymbol{\vartheta})$ is symmetric and invertible, $\|\mathbf{I}_{NT} - \Sigma_u(\boldsymbol{\vartheta})\| < 1$ (see, Horn and Johnson, 1985, p. 301), where $\|\cdot\|$ denotes a matrix norm, e.g., the maximum column or row sum norm. Accordingly, $\|\Sigma_u(\boldsymbol{\vartheta})^{-1}\| = \|\sum_{k=0}^{\infty} (\mathbf{I}_{NT} - \Sigma_u(\boldsymbol{\vartheta}))^{-1}\| \leq \sum_{k=0}^{\infty} \|\mathbf{I}_{NT} - \Sigma_u(\boldsymbol{\vartheta})\|^k = \frac{1}{1 - \|\mathbf{I}_{NT} - \Sigma_u(\boldsymbol{\vartheta})\|} < \infty$. We conclude that the row and column sums of $\Sigma_u(\boldsymbol{\vartheta})^{-1}$ are also uniformly bounded under the present assumptions. ■

Lemma 2 *Under the maintained assumptions, (i) the matrices $\Sigma_u(\boldsymbol{\vartheta})$ and $\Sigma_u(\boldsymbol{\vartheta})^{-1}$ are positive definite. (ii) Let $\mathbf{M}(\boldsymbol{\vartheta}) = \mathbf{X}(\mathbf{X}'\Sigma_u(\boldsymbol{\vartheta})^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma_u(\boldsymbol{\vartheta})^{-1}$, then $\Sigma_u(\boldsymbol{\vartheta})^{-1}(\mathbf{I}_{NT} - \mathbf{M}(\boldsymbol{\vartheta}))$ is positive definite.*

Proof. (i) Observe that $\det[\Sigma_u(\boldsymbol{\vartheta})] = \det[T\phi(\mathbf{A}'\mathbf{A})^{-1} + (\mathbf{B}'\mathbf{B})^{-1}] \det[(\mathbf{B}'\mathbf{B})^{-1}]^{T-1}$ and that $\det[T\phi(\mathbf{A}'\mathbf{A})^{-1} + (\mathbf{B}'\mathbf{B})^{-1}] \geq \det[T\phi(\mathbf{A}'\mathbf{A})^{-1}] + \det[(\mathbf{B}'\mathbf{B})^{-1}] > 0$, since $\phi > 0$ and $(\mathbf{A}'\mathbf{A})^{-1}$ as well as $(\mathbf{B}'\mathbf{B})^{-1}$ are positive definite by Assumption A2 (see Abadir and Magnus, 2005, p. 215 and p. 325). Therefore, $\Sigma_u(\boldsymbol{\vartheta})$

and $\Sigma_u(\boldsymbol{\vartheta})^{-1}$ are positive definite.

(ii) This result holds, since $\Sigma_u(\boldsymbol{\vartheta})$ is positive definite and \mathbf{X} is of full column rank (K). ■

The proof of consistency of the maximum likelihood estimates under the general model is based on the concentrated log-likelihood. Recall that the unconcentrated log-likelihood is given by

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\vartheta}) &= -\frac{NT}{2} \ln 2\pi - \frac{1}{2} \ln \det[T\sigma_\mu^2(\mathbf{A}'\mathbf{A})^{-1} + \sigma_\nu^2(\mathbf{B}'\mathbf{B})^{-1}] \\ &\quad - \frac{T-1}{2} \ln \det(\sigma_\nu^2(\mathbf{B}'\mathbf{B})^{-1}) - \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Omega}_u^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}). \end{aligned}$$

The first order conditions for $\boldsymbol{\beta}$ and σ_ν^2 are given by

$$\begin{aligned} \frac{\partial L(\boldsymbol{\beta}, \boldsymbol{\vartheta})}{\partial \boldsymbol{\beta}} &= \frac{1}{\sigma_\nu^2} (\mathbf{X}'\Sigma_u(\boldsymbol{\vartheta})^{-1}\mathbf{y} - \mathbf{X}'\Sigma_u(\boldsymbol{\vartheta})^{-1}\mathbf{X}\boldsymbol{\beta}(\boldsymbol{\vartheta})) = \mathbf{0} \\ &\Rightarrow \widehat{\boldsymbol{\beta}}(\boldsymbol{\vartheta}) = (\mathbf{X}'\Sigma_u(\boldsymbol{\vartheta})^{-1}\mathbf{X})^{-1} \mathbf{X}'\Sigma_u^{-1}(\boldsymbol{\vartheta})^{-1}\mathbf{y} \\ \frac{\partial L(\boldsymbol{\beta}, \boldsymbol{\vartheta})}{\partial \sigma_\nu^2} &= -\frac{NT}{2\sigma_\nu^2} + \frac{1}{2\sigma_\nu^4} \mathbf{u}(\widehat{\boldsymbol{\beta}}(\boldsymbol{\vartheta}))'\Sigma_u(\boldsymbol{\vartheta})^{-1} \mathbf{u}(\widehat{\boldsymbol{\beta}}(\boldsymbol{\vartheta})) = 0 \\ &\Rightarrow \widehat{\sigma}_\nu^2(\boldsymbol{\vartheta}) = \frac{\mathbf{u}(\widehat{\boldsymbol{\beta}}(\boldsymbol{\vartheta}))'\Sigma_u(\boldsymbol{\vartheta})^{-1} \mathbf{u}(\widehat{\boldsymbol{\beta}}(\boldsymbol{\vartheta}))}{NT} = \frac{\mathbf{y}'\Sigma_u(\boldsymbol{\vartheta})^{-1}(\mathbf{I}_{NT} - \mathbf{M}(\boldsymbol{\vartheta}))\mathbf{y}}{NT}, \end{aligned}$$

where $\mathbf{M}(\boldsymbol{\vartheta}) = \mathbf{X}(\mathbf{X}'\Sigma_u(\boldsymbol{\vartheta})^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma_u(\boldsymbol{\vartheta})^{-1}$ and $\mathbf{u}(\widehat{\boldsymbol{\beta}}(\boldsymbol{\vartheta})) = \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}(\boldsymbol{\vartheta})$. This uses $\Sigma_u(\boldsymbol{\vartheta})^{-1}\mathbf{M}(\boldsymbol{\vartheta}) = \mathbf{M}(\boldsymbol{\vartheta})'\Sigma_u(\boldsymbol{\vartheta})^{-1} = \mathbf{M}(\boldsymbol{\vartheta})'\Sigma_u(\boldsymbol{\vartheta})^{-1}\mathbf{M}(\boldsymbol{\vartheta})$. Since the elements of \mathbf{X} are uniformly bounded by Assumption A5 and $\Sigma_u(\boldsymbol{\vartheta})$ is uniformly bounded in its row and column sums by Lemma 1, it follows that the row and column sums of $\mathbf{M}(\boldsymbol{\vartheta})$ are uniformly bounded. Also, $\Sigma_u(\boldsymbol{\vartheta})^{-1}(\mathbf{I}_{NT} - \mathbf{M}(\boldsymbol{\vartheta}))$ is positive definite by Lemma 2. This implies that $\widehat{\sigma}_\nu^2(\boldsymbol{\vartheta}) > 0$.

The concentrated log-likelihood function is then given by

$$L^c(\boldsymbol{\vartheta}) = -\frac{NT}{2} \ln 2\pi - \frac{NT}{2} \ln \widehat{\sigma}_\nu^2(\boldsymbol{\vartheta}) - \frac{1}{2} \ln \det \Sigma_u(\boldsymbol{\vartheta}) - \frac{NT}{2}.$$

To obtain the non-stochastic counterpart of $L^c(\boldsymbol{\vartheta})$, we use

$$E[L(\boldsymbol{\beta}_0, \boldsymbol{\theta})] = -\frac{n}{2} \ln 2\pi - \frac{NT}{2} \ln \sigma_\nu^2 - \frac{1}{2} \ln [\det \boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})] - \frac{\sigma_{\nu,0}^2}{2\sigma_\nu^2} \text{tr}[\boldsymbol{\Sigma}(\boldsymbol{\vartheta})^{-1} \boldsymbol{\Sigma}_u(\boldsymbol{\vartheta}_0)]$$

and

$$\begin{aligned} \frac{\partial E[L(\boldsymbol{\beta}_0, \boldsymbol{\theta})]}{\partial \sigma_\nu^2} &= -\frac{NT}{2\sigma_\nu^{*2}} + \frac{\sigma_{\nu,0}^2}{2\sigma_\nu^{*4}} \text{tr}[\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1} \boldsymbol{\Sigma}_u(\boldsymbol{\vartheta}_0)] = 0 \\ \Rightarrow \sigma_\nu^{*2}(\boldsymbol{\vartheta}) &= \frac{\sigma_{\nu,0}^2}{NT} \text{tr}[\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1} \boldsymbol{\Sigma}_u(\boldsymbol{\vartheta}_0)]. \end{aligned}$$

Since $\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}$ is positive definite by Lemma 2, it follows that $\sigma_\nu^{*2}(\boldsymbol{\vartheta}) > 0$

and

$$\begin{aligned} Q(\boldsymbol{\vartheta}) &= \max_{\sigma_\nu^2} E[L(\boldsymbol{\beta}_0, \boldsymbol{\theta})] \\ &= -\frac{NT}{2} \ln 2\pi - \frac{NT}{2} \ln \sigma_\nu^{*2}(\boldsymbol{\vartheta}) - \frac{1}{2} \ln \det \boldsymbol{\Sigma}_u(\boldsymbol{\vartheta}) - \frac{NT}{2}. \end{aligned}$$

Theorem 3 *Under Assumptions A1-A5, the maximum likelihood estimates are unique and consistent.*

Proof. To prove consistency, we have to show that $\frac{1}{NT}(L^c(\boldsymbol{\vartheta}) - Q(\boldsymbol{\vartheta}))$ converges uniformly to 0 in probability.

Note that $\frac{1}{NT}(L^c(\boldsymbol{\vartheta}) - Q(\boldsymbol{\vartheta})) = -\frac{1}{2}(\ln \widehat{\sigma}_\nu^2(\boldsymbol{\vartheta}) - \ln \sigma_\nu^{*2}(\boldsymbol{\vartheta}))$ and $\mathbf{u}(\widehat{\boldsymbol{\beta}}(\boldsymbol{\vartheta}))' \boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1} \mathbf{u}(\widehat{\boldsymbol{\beta}}(\boldsymbol{\vartheta})) = \mathbf{u}(\boldsymbol{\beta}_0)' \boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1} \mathbf{u}(\boldsymbol{\beta}_0) - \mathbf{u}(\boldsymbol{\beta}_0)' \boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1} \mathbf{M}(\boldsymbol{\vartheta}) \mathbf{u}(\boldsymbol{\beta}_0) = \text{tr}[\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1} (\mathbf{I}_{NT} - \mathbf{M}(\boldsymbol{\vartheta})) \mathbf{u}(\boldsymbol{\beta}_0) \mathbf{u}(\boldsymbol{\beta}_0)']$.

Now, $\lim_{N \rightarrow \infty} E[\widehat{\sigma}_\nu^2(\boldsymbol{\vartheta}) - \sigma_\nu^{*2}(\boldsymbol{\vartheta})] = -\lim_{N \rightarrow \infty} \frac{1}{NT} E[\text{tr}[\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1} \mathbf{M}(\boldsymbol{\vartheta}) \mathbf{u}(\boldsymbol{\beta}_0) \mathbf{u}(\boldsymbol{\beta}_0)']] = -\lim_{N \rightarrow \infty} \frac{\sigma_{\nu,0}^2}{NT} \text{tr}[\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1} \mathbf{M}(\boldsymbol{\vartheta}) \boldsymbol{\Sigma}_u(\boldsymbol{\vartheta}_0)]$. According to Lemma 1, the row and column sums of $\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1} \mathbf{M}(\boldsymbol{\vartheta})$ and $\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta}_0)$ are bounded in absolute value

and this property is preserved under matrix multiplication. Therefore, the elements of $\Sigma_u(\boldsymbol{\vartheta})^{-1}\mathbf{M}(\boldsymbol{\vartheta})\Sigma_u(\boldsymbol{\vartheta}_0)$ are uniformly bounded by some constant c_M (see also Lemma A.7 in Lee, 2004b) so that $\frac{\sigma_{\nu,0}^2}{NT}\text{tr}[\Sigma_u(\boldsymbol{\vartheta})^{-1}\mathbf{M}(\boldsymbol{\vartheta})\Sigma_u(\boldsymbol{\vartheta}_0)] \leq \frac{\sigma_{\nu,0}^2}{NT}Kc_M$ and $\lim_{N \rightarrow \infty} \frac{\sigma_{\nu,0}^2}{NT}Kc_M = 0$. The latter follows from the fact that $\Sigma_u(\boldsymbol{\vartheta})^{-1}\mathbf{M}(\boldsymbol{\vartheta})\Sigma_u(\boldsymbol{\vartheta}_0)$ is of rank K .

$\lim_{N \rightarrow \infty} \text{Var}[\widehat{\sigma}_\nu^2(\boldsymbol{\vartheta}) - \sigma_\nu^{*2}(\boldsymbol{\vartheta})] = \lim_{N \rightarrow \infty} \text{Var}[\frac{1}{NT}\text{tr}[\Sigma_u(\boldsymbol{\vartheta})^{-1}\mathbf{M}(\boldsymbol{\vartheta})\mathbf{u}(\boldsymbol{\beta}_0)\mathbf{u}(\boldsymbol{\beta}_0)']] = \lim_{N \rightarrow \infty} \frac{2\sigma_{\nu,0}^4}{(NT)^2}\text{tr}[(\Sigma_u(\boldsymbol{\vartheta})^{-1}\mathbf{M}(\boldsymbol{\vartheta})\Sigma_u(\boldsymbol{\vartheta}_0))^2]$ using Lemma (A1) in Kelejian and Prucha (2007, p. 29) and Assumption A1. As a result, $\lim_{N \rightarrow \infty} \frac{2\sigma_{\nu,0}^4}{(NT)^2}\text{tr}[(\Sigma_u(\boldsymbol{\vartheta})^{-1}\mathbf{M}(\boldsymbol{\vartheta})\Sigma_u(\boldsymbol{\vartheta}_0))^2] = o(1)$. By Chebyshev's inequality, we conclude that $\widehat{\sigma}_\nu^2(\boldsymbol{\vartheta}) - \sigma_\nu^{*2}(\boldsymbol{\vartheta}) = o_p(1)$. Also, $\widehat{\sigma}_\nu^2(\boldsymbol{\vartheta}) > 0$ and $\sigma_\nu^{*2}(\boldsymbol{\vartheta}) > 0$ as shown above.

Using the mean value theorem it follows that $\ln \widehat{\sigma}_\nu^2(\boldsymbol{\vartheta}) = \ln \sigma_\nu^{*2}(\boldsymbol{\vartheta}) + \frac{\widehat{\sigma}_\nu^2(\boldsymbol{\vartheta}) - \sigma_\nu^{*2}(\boldsymbol{\vartheta})}{\bar{\sigma}_\nu^2(\boldsymbol{\vartheta})}$ with the constant $\bar{\sigma}_\nu^2(\boldsymbol{\vartheta})$ lying in between $\sigma_\nu^{*2}(\boldsymbol{\vartheta})$ and $\widehat{\sigma}_\nu^2(\boldsymbol{\vartheta})$ and $\frac{1}{\bar{\sigma}_\nu^2(\boldsymbol{\vartheta})} < \frac{1}{\sigma_\nu^{*2}(\boldsymbol{\vartheta})} + \frac{1}{\widehat{\sigma}_\nu^2(\boldsymbol{\vartheta})} < \infty$. Therefore, we obtain $\sup_{\boldsymbol{\vartheta} \in \Xi} \frac{2}{NT} |L^c(\boldsymbol{\vartheta}) - Q(\boldsymbol{\vartheta})| = \sup_{\boldsymbol{\vartheta} \in \Xi} |\ln \widehat{\sigma}_\nu^2(\boldsymbol{\vartheta}) - \ln \sigma_\nu^{*2}(\boldsymbol{\vartheta})| = \sup_{\boldsymbol{\vartheta} \in \Xi} \frac{1}{\bar{\sigma}_\nu^2(\boldsymbol{\vartheta})} |\widehat{\sigma}_\nu^2(\boldsymbol{\vartheta}) - \sigma_\nu^{*2}(\boldsymbol{\vartheta})| = o_p(1)$, since $\widehat{\sigma}_\nu^2(\boldsymbol{\vartheta}) - \sigma_\nu^{*2}(\boldsymbol{\vartheta}) = o_p(1)$ and $\frac{1}{\bar{\sigma}_\nu^2(\boldsymbol{\vartheta})}$ is bounded by some positive constant.

Secondly, we have to prove the following uniqueness identification condition (see Lee, 2004a). For any $\varepsilon > 0$, $\limsup_{N \rightarrow \infty} \max_{\boldsymbol{\vartheta} \in \overline{\mathbf{N}}_\varepsilon(\boldsymbol{\vartheta}_0)} \frac{1}{NT} (Q(\boldsymbol{\vartheta}) - Q(\boldsymbol{\vartheta}_0)) < 0$, where $\overline{\mathbf{N}}_\varepsilon(\boldsymbol{\vartheta}_0)$ is the complement of an open neighborhood of $\boldsymbol{\vartheta}_0$ of diameter ε . Note, $Q(\boldsymbol{\vartheta}) - Q(\boldsymbol{\vartheta}_0) = -\frac{NT}{2} [\ln \sigma_\nu^{*2}(\boldsymbol{\vartheta}) - \ln \sigma_\nu^{*2}(\boldsymbol{\vartheta}_0)] - \frac{1}{2} \ln[\det \Sigma_u(\boldsymbol{\vartheta}) / \det \Sigma_u(\boldsymbol{\vartheta}_0)]$.

Now, $\ln \sigma_\nu^{*2}(\boldsymbol{\vartheta}) - \ln \sigma_\nu^{*2}(\boldsymbol{\vartheta}_0) = \ln \text{tr} \frac{1}{NT} [\Sigma_u(\boldsymbol{\vartheta}_0)\Sigma_u(\boldsymbol{\vartheta})^{-1}] - \ln \frac{1}{NT} \text{tr} [\mathbf{I}_{NT}] =$

$$\ln tr \frac{1}{NT} [\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta}_0) \boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}] \text{ and } \frac{1}{NT} (Q(\boldsymbol{\vartheta}) - Q(\boldsymbol{\vartheta}_0)) = -\frac{1}{2} \ln \frac{1}{NT} tr [\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta}_0) \boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}] - \frac{1}{2NT} \ln(\det \boldsymbol{\Sigma}_u(\boldsymbol{\vartheta}) / \det \boldsymbol{\Sigma}_u(\boldsymbol{\vartheta}_0)) < 0$$

for every $\boldsymbol{\vartheta} \neq \boldsymbol{\vartheta}_0 \in \Xi$ by Assumption A4. Accordingly, we conclude that the maximum likelihood estimator $\widehat{\boldsymbol{\vartheta}}$ of $\boldsymbol{\vartheta}_0$ under the general model is unique and consistent, since $Q(\boldsymbol{\vartheta})$ is continuous and the parameter space is compact.

Lastly, $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\vartheta}})$ is identified by Assumption A5. Specifically, we have $\left(\frac{1}{NT} \mathbf{X}' \boldsymbol{\Sigma}_u(\widehat{\boldsymbol{\vartheta}})^{-1} \mathbf{X} \right)^{-1} \xrightarrow{p} \left(\frac{1}{NT} \mathbf{X}' \boldsymbol{\Sigma}_u(\boldsymbol{\vartheta}_0)^{-1} \mathbf{X} \right)^{-1}$ and $\frac{1}{NT} \mathbf{X}' \boldsymbol{\Sigma}_u^{-1}(\widehat{\boldsymbol{\vartheta}})^{-1} \mathbf{u}(\widehat{\boldsymbol{\vartheta}}) \xrightarrow{p} \mathbf{0}$, since $\widehat{\boldsymbol{\vartheta}} \xrightarrow{p} \boldsymbol{\vartheta}_0$. Hence, $plim_{N \rightarrow \infty} \widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\vartheta}}) = \boldsymbol{\beta}_0 + plim_{N \rightarrow \infty} \left[\left(\frac{1}{NT} \mathbf{X}' \boldsymbol{\Sigma}_u(\widehat{\boldsymbol{\vartheta}})^{-1} \mathbf{X} \right)^{-1} \frac{1}{NT} \mathbf{X}' \boldsymbol{\Sigma}_u^{-1}(\widehat{\boldsymbol{\vartheta}})^{-1} \mathbf{u}(\widehat{\boldsymbol{\vartheta}}) \right] = \boldsymbol{\beta}_0$ under Assumption A5. ■

Appendix C: LM Test for random effects

Below, Theorems 7-9 derive the asymptotic distribution of the LM tests.

The following three Lemmas are useful in proving these theorems.

Lemma 4 *Let $\boldsymbol{\mu} \sim N(\mathbf{0}, \sigma_\mu^2 \mathbf{I}_N)$ and $\boldsymbol{\nu} \sim N(\mathbf{0}, \sigma_\nu^2 (\mathbf{I}_T \otimes \mathbf{I}_N))$ and assume that Assumptions A1-A2 hold. Consider the quadratic form $Q_b = (\mathbf{Z}_\mu \mathbf{A}^{-1} \boldsymbol{\mu} + (\mathbf{I}_T \otimes \mathbf{B}^{-1}) \boldsymbol{\nu})' (\bar{\mathbf{J}}_T \otimes \mathbf{H}) (\mathbf{Z}_\mu \mathbf{A}^{-1} \boldsymbol{\mu} + (\mathbf{I}_T \otimes \mathbf{B}^{-1}) \boldsymbol{\nu})$, where \mathbf{H} is a non-stochastic symmetric $N \times N$ matrix with uniformly bounded row and column sums. Then $E[Q_b] = T \sigma_\mu^2 tr[\mathbf{H}(\mathbf{A}'\mathbf{A})^{-1}] + \sigma_\nu^2 tr[\mathbf{H}(\mathbf{B}'\mathbf{B})^{-1}]$, $Var[Q_b] = 2\{T^2 \sigma_\mu^4 tr[(\mathbf{H}(\mathbf{A}'\mathbf{A})^{-1})^2] + 2T \sigma_\nu^2 \sigma_\mu^2 tr[\mathbf{H}(\mathbf{A}'\mathbf{A})^{-1} \mathbf{H}(\mathbf{B}'\mathbf{B})^{-1}] + \sigma_\nu^4 tr[(\mathbf{H}(\mathbf{B}'\mathbf{B})^{-1})^2]\}$ and $\frac{Q_b - T \sigma_\mu^2 tr[\mathbf{H}(\mathbf{A}'\mathbf{A})^{-1}] - \sigma_\nu^2 tr[\mathbf{H}(\mathbf{B}'\mathbf{B})^{-1}]}{\sqrt{2\{T^2 \sigma_\mu^4 tr[(\mathbf{H}(\mathbf{A}'\mathbf{A})^{-1})^2] + 2T \sigma_\nu^2 \sigma_\mu^2 tr[\mathbf{H}(\mathbf{A}'\mathbf{A})^{-1} \mathbf{H}(\mathbf{B}'\mathbf{B})^{-1}] + \sigma_\nu^4 tr[(\mathbf{H}(\mathbf{B}'\mathbf{B})^{-1})^2]\}} \xrightarrow{d} N(0, 1)$.*

Proof. Inserting $\mathbf{Z}_\mu = (\iota_T \otimes \mathbf{I}_N)$ gives

$$\begin{aligned} \mathbf{A}'^{-1} \mathbf{Z}'_\mu (\bar{\mathbf{J}}_T \otimes \mathbf{H}) \mathbf{Z}_\mu \mathbf{A}^{-1} &= \mathbf{A}'^{-1} (\iota'_T \otimes \mathbf{I}_N) (\bar{\mathbf{J}}_T \otimes \mathbf{H}) (\iota_T \otimes \mathbf{I}_N) \mathbf{A}^{-1} \\ &= T \mathbf{A}'^{-1} \mathbf{H} \mathbf{A}^{-1} \end{aligned}$$

$$\mathbf{A}'^{-1} \mathbf{Z}'_\mu (\bar{\mathbf{J}}_T \otimes \mathbf{H}) (\mathbf{I}_T \otimes \mathbf{B}^{-1}) = \mathbf{A}'^{-1} (\iota'_T \otimes \mathbf{H} \mathbf{B}^{-1})$$

$$(\mathbf{I}_T \otimes \mathbf{B}^{-1})' (\bar{\mathbf{J}}_T \otimes \mathbf{H}) (\mathbf{I}_T \otimes \mathbf{B}^{-1}) = (\bar{\mathbf{J}}_T \otimes \mathbf{B}'^{-1} \mathbf{H} \mathbf{B}^{-1})$$

Let $\mathbf{L}_1 = \mathbf{A}'^{-1} \mathbf{H} \mathbf{A}^{-1}$, $\mathbf{L}_2 = \mathbf{A}'^{-1} \mathbf{H} \mathbf{B}^{-1}$, $\mathbf{L}_3 = \mathbf{B}'^{-1} \mathbf{H} \mathbf{B}^{-1}$, and $\boldsymbol{\xi} = (\boldsymbol{\mu}', \boldsymbol{\nu}'_1, \dots, \boldsymbol{\nu}'_T)'$.

Then, we obtain

$$\begin{aligned} Q_b &= \boldsymbol{\xi}' \mathbf{L}_b \boldsymbol{\xi} = \boldsymbol{\xi}' \begin{bmatrix} T \mathbf{L}_1 & \mathbf{L}_2 & \dots & \mathbf{L}_2 \\ \mathbf{L}'_2 & \frac{1}{T} \mathbf{L}_3 & \dots & \frac{1}{T} \mathbf{L}_3 \\ \dots & \dots & \dots & \dots \\ \mathbf{L}'_2 & \frac{1}{T} \mathbf{L}_3 & \dots & \frac{1}{T} \mathbf{L}_3 \end{bmatrix} \boldsymbol{\xi} \\ &= T \boldsymbol{\mu}' \mathbf{L}_1 \boldsymbol{\mu} + \sum_{t=1}^T \boldsymbol{\nu}'_t \mathbf{L}'_2 \boldsymbol{\mu} + \sum_{t=1}^T \boldsymbol{\mu}' \mathbf{L}_2 \boldsymbol{\nu}_t + \frac{1}{T} \left(\sum_{t=1}^T \boldsymbol{\nu}'_t \right) \mathbf{L}_3 \left(\sum_{t=1}^T \boldsymbol{\nu}_t \right). \end{aligned}$$

Now

$$\boldsymbol{\Omega}_\xi = E[\boldsymbol{\xi} \boldsymbol{\xi}'] = \begin{bmatrix} \sigma_\mu^2 \mathbf{I}_N & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \sigma_\nu^2 \mathbf{I}_N & \dots & \mathbf{0} \\ \dots & \dots & \dots & \dots \\ \mathbf{0} & \mathbf{0} & \dots & \sigma_\nu^2 \mathbf{I}_N \end{bmatrix},$$

since $E[\boldsymbol{\mu} \boldsymbol{\mu}'] = \sigma_\mu^2 \mathbf{I}_N$, $E[\boldsymbol{\mu} \boldsymbol{\nu}'_t] = E[\boldsymbol{\nu}_\tau \boldsymbol{\nu}'_t] = 0$ for $\tau \neq t$ and $E[\boldsymbol{\nu}_t \boldsymbol{\nu}'_t] = \sigma_\nu^2 \mathbf{I}_N$.

Using Lemma A1 of Kelejian and Prucha (2007, p. 29), we have

$$\begin{aligned}
E[Q_b] &= tr[\mathbf{L}_b \boldsymbol{\Omega}_\xi] \\
&= tr \begin{bmatrix} T\sigma_\mu^2 \mathbf{L}_1 & \sigma_\nu^2 \mathbf{L}_2 & \dots & \sigma_\nu^2 \mathbf{L}_2 \\ \sigma_\mu^2 \mathbf{L}'_2 & \frac{1}{T} \sigma_\nu^2 \mathbf{L}_3 & \dots & \frac{1}{T} \sigma_\nu^2 \mathbf{L}_3 \\ \dots & \dots & \dots & \dots \\ \sigma_\mu^2 \mathbf{L}'_2 & \frac{1}{T} \sigma_\nu^2 \mathbf{L}_3 & \dots & \frac{1}{T} \sigma_\nu^2 \mathbf{L}_3 \end{bmatrix} \\
&= T\sigma_\mu^2 tr[\mathbf{L}_1] + \sigma_\nu^2 tr[\mathbf{L}_3] = T\sigma_\mu^2 tr[\mathbf{H}(\mathbf{A}'\mathbf{A})^{-1}] + \sigma_\nu^2 tr[\mathbf{H}(\mathbf{B}'\mathbf{B})^{-1}],
\end{aligned}$$

since $tr[\mathbf{A}'^{-1}\mathbf{H}\mathbf{A}^{-1}] = tr[\mathbf{H}(\mathbf{A}'\mathbf{A})^{-1}]$, $tr[\mathbf{B}'^{-1}\mathbf{H}\mathbf{B}^{-1}] = tr[\mathbf{H}(\mathbf{B}'\mathbf{B})^{-1}]$, and

$$\begin{aligned}
Var[Q_b] &= 2tr[\mathbf{L}_b \boldsymbol{\Omega}_\xi \mathbf{L}_b \boldsymbol{\Omega}_\xi] = \\
&= 2tr \begin{bmatrix} T\sigma_\mu^2 \mathbf{L}_1 & \sigma_\nu^2 \mathbf{L}_2 & \dots & \sigma_\nu^2 \mathbf{L}_2 \\ \sigma_\mu^2 \mathbf{L}'_2 & \frac{1}{T} \sigma_\nu^2 \mathbf{L}_3 & \dots & \frac{1}{T} \sigma_\nu^2 \mathbf{L}_3 \\ \dots & \dots & \dots & \dots \\ \sigma_\mu^2 \mathbf{L}'_2 & \frac{1}{T} \sigma_\nu^2 \mathbf{L}_3 & \dots & \frac{1}{T} \sigma_\nu^2 \mathbf{L}_3 \end{bmatrix}^2 = \\
&= 2tr \begin{bmatrix} T^2 \sigma_\mu^4 \mathbf{L}_1^2 + T \sigma_\nu^2 \sigma_\mu^2 \mathbf{L}_2 \mathbf{L}'_2 & T \sigma_\mu^2 \sigma_\nu^2 \mathbf{L}_1 \mathbf{L}_2 + \sigma_\nu^4 \mathbf{L}_2 \mathbf{L}_3 & \dots & T \sigma_\mu^2 \sigma_\nu^2 \mathbf{L}_1 \mathbf{L}_2 + \sigma_\nu^4 \mathbf{L}_2 \mathbf{L}_3 \\ T \sigma_\mu^4 \mathbf{L}'_2 \mathbf{L}_1 + \sigma_\nu^2 \sigma_\mu^2 \mathbf{L}_3 \mathbf{L}'_2 & \sigma_\nu^2 \sigma_\mu^2 \mathbf{L}'_2 \mathbf{L}_2 + \frac{1}{T} \sigma_\nu^4 \mathbf{L}_3^2 & \dots & \sigma_\nu^2 \sigma_\mu^2 \mathbf{L}'_2 \mathbf{L}_2 + \frac{1}{T} \sigma_\nu^4 \mathbf{L}_3^2 \\ \dots & \dots & \dots & \dots \\ T \sigma_\mu^4 \mathbf{L}'_2 \mathbf{L}_1 + \sigma_\nu^2 \sigma_\mu^2 \mathbf{L}_3 \mathbf{L}'_2 & \sigma_\nu^2 \sigma_\mu^2 \mathbf{L}'_2 \mathbf{L}_2 + \frac{1}{T} \sigma_\nu^4 \mathbf{L}_3^2 & \dots & \sigma_\nu^2 \sigma_\mu^2 \mathbf{L}'_2 \mathbf{L}_2 + \frac{1}{T} \sigma_\nu^4 \mathbf{L}_3^2 \end{bmatrix},
\end{aligned}$$

which reduces to

$$\begin{aligned}
Var[Q_b] &= 2(T^2 \sigma_\mu^4 tr[\mathbf{L}_1^2] + 2T \sigma_\nu^2 \sigma_\mu^2 tr[\mathbf{L}'_2 \mathbf{L}_2] + \sigma_\nu^4 tr[\mathbf{L}_3^2]) \\
&= 2(T^2 \sigma_\mu^4 tr[(\mathbf{H}(\mathbf{A}'\mathbf{A})^{-1})^2] + 2T \sigma_\nu^2 \sigma_\mu^2 tr[\mathbf{H}(\mathbf{A}'\mathbf{A})^{-1} \mathbf{H}(\mathbf{B}'\mathbf{B})^{-1}] \\
&\quad + \sigma_\nu^4 tr[(\mathbf{H}(\mathbf{B}'\mathbf{B})^{-1})^2]).
\end{aligned}$$

Since the row and column sums of \mathbf{A} , \mathbf{B} , $(\mathbf{A}'\mathbf{A})^{-1}$, $(\mathbf{B}'\mathbf{B})^{-1}$ and \mathbf{H} are uniformly bounded, so are those of \mathbf{L}_1 , \mathbf{L}_2 , \mathbf{L}_3 and \mathbf{L}_b . Furthermore, since the elements of $\boldsymbol{\xi}$ are independent and normally distributed by Assumption A1 and $Q_b = \boldsymbol{\xi}'\mathbf{L}_b\boldsymbol{\xi} = \frac{1}{2}\boldsymbol{\xi}'(\mathbf{L}'_b + \mathbf{L}_b)\boldsymbol{\xi}$, the assumptions of the central limit theorem for linear quadratic forms given as Theorem 1 in Kelejian and Prucha (2001, p. 227) are fulfilled and

$$\frac{Q_b - T\sigma_\mu^2 \text{tr}[\mathbf{H}(\mathbf{A}'\mathbf{A})^{-1}] - \sigma_\nu^2 \text{tr}[\mathbf{H}(\mathbf{B}'\mathbf{B})^{-1}]}{\sqrt{2\{T^2\sigma_\mu^4 \text{tr}[(\mathbf{H}(\mathbf{A}'\mathbf{A})^{-1})^2] + 2T\sigma_\nu^2\sigma_\mu^2 \text{tr}[\mathbf{H}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{H}(\mathbf{B}'\mathbf{B})^{-1}] + \sigma_\nu^4 \text{tr}[(\mathbf{H}(\mathbf{B}'\mathbf{B})^{-1})^2]\}} \xrightarrow{d} N(0, 1). \blacksquare$$

Lemma 5 *Let $\boldsymbol{\mu} \sim N(0, \sigma_\mu^2 \mathbf{I}_N)$ and $\boldsymbol{\nu} \sim N(0, \sigma_\nu^2 (\mathbf{I}_T \otimes \mathbf{I}_N))$ and assume that Assumptions A1-A2 hold. Consider the quadratic form $Q_w = (\mathbf{Z}_\mu \boldsymbol{\mu} + (\mathbf{I}_T \otimes \mathbf{I}_N) \boldsymbol{\nu})' (\mathbf{E}_T \otimes \mathbf{H}) (\mathbf{Z}_\mu \boldsymbol{\mu} + (\mathbf{I}_T \otimes \mathbf{I}_N) \boldsymbol{\nu})$, where \mathbf{H} is a non-stochastic symmetric $N \times N$ matrix with uniformly bounded row and column sums. Then, $E[Q_w] = \sigma_\nu^2 (T-1) \text{tr}[\mathbf{H}]$, $\text{Var}[Q_w] = 2\sigma_\nu^4 (T-1) \text{tr}[\mathbf{H}^2]$ and $\frac{Q_w - \sigma_\nu^2 (T-1) \text{tr}[\mathbf{H}]}{\sigma_\nu^2 \sqrt{2(T-1) \text{tr}[\mathbf{H}^2]}} \xrightarrow{d} N(0, 1)$.*

Proof. Inserting $\mathbf{Z}_\mu = (\iota_T \otimes \mathbf{I}_N)$ and using $\iota_T' \mathbf{E}_T = \mathbf{E}_T \iota_T = 0$ gives

$$\mathbf{Z}'_\mu (\mathbf{E}_T \otimes \mathbf{H}) \mathbf{Z}_\mu = \mathbf{0}$$

$$\mathbf{Z}'_\mu (\mathbf{E}_T \otimes \mathbf{H}) (\mathbf{I}_T \otimes \mathbf{I}_N) = \mathbf{0}$$

$$(\mathbf{I}_T \otimes \mathbf{I}_N)' (\mathbf{E}_T \otimes \mathbf{H}) (\mathbf{I}_T \otimes \mathbf{I}_N) = \mathbf{E}_T \otimes \mathbf{H}$$

and

$$\begin{aligned} Q_w &= \boldsymbol{\nu}' (\mathbf{E}_T \otimes \mathbf{H}) \boldsymbol{\nu} \\ &= \sum_{t=1}^T \boldsymbol{\nu}'_t \mathbf{H} \boldsymbol{\nu}_t - \frac{1}{T} \left(\sum_{t=1}^T \boldsymbol{\nu}'_t \right) \mathbf{H} \left(\sum_{t=1}^T \boldsymbol{\nu}_t \right). \end{aligned}$$

Next, $E[\boldsymbol{\nu}' (\mathbf{E}_T \otimes \mathbf{H}) \boldsymbol{\nu}] = \text{tr}[(\mathbf{E}_T \otimes \mathbf{H}) \sigma_\nu^2 \mathbf{I}_{NT}] = \sigma_\nu^2 (T-1) \text{tr}[\mathbf{H}]$, and $\text{Var}[\boldsymbol{\nu}' (\mathbf{E}_T \otimes \mathbf{H}) \boldsymbol{\nu}] = 2\sigma_\nu^4 (T-1) \text{tr}[\mathbf{H}^2]$. Theorem 1 in Kelejian and Prucha (2001) can be

directly applied, since the elements of $\boldsymbol{\nu}$ are independent and the row and column sums of \mathbf{H} are bounded in absolute value: $\frac{Q_w - \sigma_\nu^2(T-1)\text{tr}[\mathbf{H}]}{\sigma_\nu^2 \sqrt{2(T-1)\text{tr}[\mathbf{H}^2]}} \xrightarrow{d} N(0, 1)$.

■

Lemma 6 *Assume $\mathbf{u} \sim N(\mathbf{0}, \boldsymbol{\Omega}_u)$, where $\boldsymbol{\Omega}_u = \sigma_1^2(\bar{\mathbf{J}}_T \otimes \mathbf{I}_N) + \sigma_\nu^2(\mathbf{E}_T \otimes \mathbf{I}_N)$. Then $G_A = \mathbf{u}'(\bar{\mathbf{J}}_T \otimes (\mathbf{W}' + \mathbf{W}))\mathbf{u}$ and $M_A = \mathbf{u}'(\mathbf{E}_T \otimes (\mathbf{W}' + \mathbf{W}))\mathbf{u}$ are independent.*

Proof. By Theorem (viii) in Rao (1973, p. 188), a necessary and sufficient condition for the independence of G_A and M_A is $\boldsymbol{\Omega}_u(\bar{\mathbf{J}}_T \otimes (\mathbf{W}' + \mathbf{W}))\boldsymbol{\Omega}_u(\mathbf{E}_T \otimes (\mathbf{W}' + \mathbf{W}))\boldsymbol{\Omega}_u = 0$. Now $\boldsymbol{\Omega}_u(\bar{\mathbf{J}}_T \otimes (\mathbf{W}' + \mathbf{W})) = \sigma_1^2(\bar{\mathbf{J}}_T \otimes (\mathbf{W}' + \mathbf{W}))$ and $(\mathbf{E}_T \otimes (\mathbf{W}' + \mathbf{W}))\boldsymbol{\Omega}_u = \sigma_\nu^2(\mathbf{E}_T \otimes (\mathbf{W}' + \mathbf{W}))$. Therefore,

$$\begin{aligned} & \boldsymbol{\Omega}_u(\bar{\mathbf{J}}_T \otimes (\mathbf{W}' + \mathbf{W}))\boldsymbol{\Omega}_u(\mathbf{E}_T \otimes (\mathbf{W}' + \mathbf{W}))\boldsymbol{\Omega}_u \\ &= \sigma_1^2(\bar{\mathbf{J}}_T \otimes (\mathbf{W}' + \mathbf{W}))[\sigma_1^2(\bar{\mathbf{J}}_T \otimes \mathbf{I}_N) + \sigma_\nu^2(\mathbf{E}_T \otimes \mathbf{I}_N)]\sigma_\nu^2(\mathbf{E}_T \otimes (\mathbf{W}' + \mathbf{W})) \\ &= \sigma_1^4(\bar{\mathbf{J}}_T \otimes (\mathbf{W}' + \mathbf{W}))\sigma_\nu^2(\mathbf{E}_T \otimes (\mathbf{W}' + \mathbf{W})) = \mathbf{0}, \end{aligned}$$

since $\bar{\mathbf{J}}_T\mathbf{E}_T = 0$. ■

Next, this Appendix derives the LM test for the null hypothesis $H_0^A : \rho_1 = \rho_2 = 0$, i.e., that there is no spatial correlation in the error term. The joint LM test for the null hypothesis of no spatial correlation in model (1) tests $H_0^A : \rho_1 = \rho_2 = 0$. The LM statistic is given by

$$LM_A = \tilde{\mathbf{D}}_\theta' \tilde{\mathbf{J}}_\theta^{-1} \tilde{\mathbf{D}}_\theta, \quad (11)$$

where $\tilde{\mathbf{D}}_\theta = (\partial L / \partial \boldsymbol{\theta})(\tilde{\boldsymbol{\theta}})$ is a 4×1 vector of partial derivatives of the log-likelihood function with respect to the elements of $\boldsymbol{\theta}$, evaluated at the restricted MLE, $\tilde{\boldsymbol{\theta}}$. $\tilde{\mathbf{J}}_\theta = E[-\partial^2 L / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'](\tilde{\boldsymbol{\theta}})$ is the part of the information matrix corresponding to $\boldsymbol{\theta}$, also evaluated at the restricted MLE, $\tilde{\boldsymbol{\theta}}$.

Under $H_0^A : \rho_1 = \rho_2 = 0$, $\mathbf{B} = \mathbf{A} = \mathbf{I}_N$. Using the general formulas given above, the score under H_0^A is determined as

$$\begin{aligned}\frac{\partial L}{\partial \sigma_\nu^2} \Big|_{H_0^A} &= -\frac{N}{2\sigma_1^2} - \frac{N(T-1)}{2\sigma_\nu^2} + \frac{1}{2} \mathbf{u}' \left[\left(\frac{1}{\sigma_1^4} \bar{\mathbf{J}}_T + \frac{1}{\sigma_\nu^4} \mathbf{E}_T \right) \otimes \mathbf{I}_N \right] \mathbf{u} \\ \frac{\partial L}{\partial \sigma_\mu^2} \Big|_{H_0^A} &= -\frac{NT}{2\sigma_1^2} + \frac{1}{2\sigma_1^4} \mathbf{u}' (\mathbf{J}_T \otimes \mathbf{I}_N) \mathbf{u} \\ \frac{\partial L}{\partial \rho_1} \Big|_{H_0^A} &= \frac{\sigma_\mu^2}{2\sigma_1^4} \mathbf{u}' [\mathbf{J}_T \otimes (\mathbf{W}' + \mathbf{W})] \mathbf{u} \\ \frac{\partial L}{\partial \rho_2} \Big|_{H_0^A} &= \frac{1}{2} \mathbf{u}' \left[\left(\frac{\sigma_\nu^2}{\sigma_1^4} \bar{\mathbf{J}}_T + \frac{1}{\sigma_\nu^2} \mathbf{E}_T \right) \otimes (\mathbf{W}' + \mathbf{W}) \right] \mathbf{u}\end{aligned}$$

and

$$\mathbf{J}_\theta \Big|_{H_0^A} = \begin{bmatrix} \frac{N}{2\sigma_1^4} + \frac{N(T-1)}{2\sigma_\nu^4} & \frac{NT}{2\sigma_1^4} & 0 & 0 \\ \frac{NT}{2\sigma_1^4} & \frac{NT^2}{2\sigma_1^4} & 0 & 0 \\ 0 & 0 & \frac{T^2 \sigma_\mu^4}{2\sigma_1^4} b_A & \frac{T \sigma_\mu^2 \sigma_\nu^2}{2\sigma_1^4} b_A \\ 0 & 0 & \frac{T \sigma_\mu^2 \sigma_\nu^2}{2\sigma_1^4} b_A & \left(\frac{\sigma_\nu^4}{2\sigma_1^4} + \frac{(T-1)}{2} \right) b_A \end{bmatrix},$$

where $b_A = \text{tr} [(\mathbf{W}' + \mathbf{W})^2]$. The score with respect to each element of $\boldsymbol{\theta}$ evaluated at the restricted MLE $\tilde{\boldsymbol{\theta}}$ under H_0^A with $\tilde{\mathbf{u}} = \mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}$ is given by

$$\tilde{\mathbf{D}}_\theta = \begin{bmatrix} 0 \\ 0 \\ \frac{T \tilde{\sigma}_\mu^2}{2 \tilde{\sigma}_1^4} \tilde{\mathbf{u}}' [\bar{\mathbf{J}}_T \otimes (\mathbf{W}' + \mathbf{W})] \tilde{\mathbf{u}} \\ \frac{1}{2} \tilde{\mathbf{u}}' \left[\left(\frac{\tilde{\sigma}_\nu^2}{\tilde{\sigma}_1^4} \bar{\mathbf{J}}_T + \frac{1}{\tilde{\sigma}_\nu^2} \mathbf{E}_T \right) \otimes (\mathbf{W}' + \mathbf{W}) \right] \tilde{\mathbf{u}} \end{bmatrix}.$$

The determinant of the submatrix $\tilde{\mathbf{J}}_{\rho_1, \rho_2}$ is determined as

$$\det \left[\tilde{\mathbf{J}}_{\rho_1, \rho_2} \Big|_{H_0^A} \right] = \left(\frac{b_A}{2} \right)^2 \frac{T^2 (T-1) \tilde{\sigma}_\mu^4}{\tilde{\sigma}_1^4}$$

and its inverse is

$$\tilde{\mathbf{J}}_{\rho_1, \rho_2}^{-1} \Big|_{H_0^A} = \frac{2}{b_A} \frac{1}{T^2 (T-1) \tilde{\sigma}_\mu^4} \begin{bmatrix} (T-1) \tilde{\sigma}_1^4 + \tilde{\sigma}_\nu^4 & -T \tilde{\sigma}_\mu^2 \tilde{\sigma}_\nu^2 \\ -T \tilde{\sigma}_\mu^2 \tilde{\sigma}_\nu^2 & T^2 \tilde{\sigma}_\mu^4 \end{bmatrix}.$$

Defining

$$\begin{aligned}\tilde{G}_A &= \tilde{\mathbf{u}}' [\bar{\mathbf{J}}_T \otimes (\mathbf{W}' + \mathbf{W})] \tilde{\mathbf{u}} \\ \tilde{M}_A &= \tilde{\mathbf{u}}' [\mathbf{E}_T \otimes (\mathbf{W}' + \mathbf{W})] \tilde{\mathbf{u}},\end{aligned}$$

we have

$$LM_A = \tilde{\mathbf{D}}_\theta' \tilde{\mathbf{J}}_\theta'^{-1} \tilde{\mathbf{D}}_\theta = \frac{1}{2b_A \tilde{\sigma}_1^4} \tilde{G}_A^2 + \frac{1}{2b_A(T-1)\tilde{\sigma}_\nu^4} \tilde{M}_A^2.$$

Theorem 7 (LM_A) *Suppose Assumptions A1 - A5 hold and $H_0^A : \rho_1 = \rho_2 = 0$ is true. Then, $LM_A = \frac{1}{2b_A \tilde{\sigma}_1^4} \tilde{G}_A^2 + \frac{1}{2b_A(T-1)\tilde{\sigma}_\nu^4} \tilde{M}_A^2$ is asymptotically distributed as χ_2^2 .*

Proof. First, use the residuals of the true model $\mathbf{u} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}_0$ and define

$$G_A = \mathbf{u}' \mathbf{G}_A \mathbf{u}$$

$$M_A = \mathbf{u}' \mathbf{M}_A \mathbf{u},$$

where $\mathbf{G}_A = \bar{\mathbf{J}}_T \otimes (\mathbf{W}' + \mathbf{W})$, and $\mathbf{M}_A = \mathbf{E}_T \otimes (\mathbf{W}' + \mathbf{W})$. Under H_0^A (random effects model) and Assumption A1, $\mathbf{u} \sim \mathbf{N}(0, \boldsymbol{\Omega}_u)$ with $\boldsymbol{\Omega}_u = \sigma_1^2 (\bar{\mathbf{J}}_T \otimes \mathbf{I}_N) + \sigma_\nu^2 (\mathbf{E}_T \otimes \mathbf{I}_N)$.

(i) We can apply Lemma 4 by setting $\mathbf{A} = \mathbf{B} = \mathbf{I}_N$ so that $\mathbf{H} = (\mathbf{W}' + \mathbf{W})$ with $tr[\mathbf{H}] = 0$, because $tr[\mathbf{W}] = 0$. Hence, $E[G_A] = 0$ and $Var[G_A] = 2\sigma_1^4 b_A$ with $b_A = tr[\mathbf{H}^2]$. By Assumption A2 the row and column sums of \mathbf{H} are uniformly bounded, so $\frac{G_A}{\sigma_1^2 \sqrt{2b_A}}$ converges in distribution to the standard normal.

(ii) Using Lemma 5 with $\mathbf{H} = (\mathbf{W}' + \mathbf{W})$ implies that $\frac{M_A}{\sigma_\nu^2 \sqrt{2(T-1)b_A}} \xrightarrow{d} N(0, 1)$.

(iii) Lemma 6 establishes the independence of G_A and M_A .

(iv) Given a consistent estimator of β_0 , say $\hat{\beta}$, and $\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\beta}$, we have $\frac{1}{NT}\hat{\mathbf{u}}'\mathbf{G}_A\hat{\mathbf{u}} = \frac{1}{NT}\mathbf{u}'\mathbf{G}_A\mathbf{u} + \frac{2}{NT}\mathbf{u}'\mathbf{G}_A\mathbf{X}(\hat{\beta}-\beta_0) + \frac{1}{NT}(\hat{\beta}-\beta_0)'\mathbf{X}'\mathbf{G}_A\mathbf{X}(\hat{\beta}-\beta_0) = \frac{1}{NT}\mathbf{u}'\mathbf{G}_A\mathbf{u} + o_p(1)$, since \mathbf{X} and \mathbf{G}_A are non-stochastic matrices (see Lemma 1 in Kelejian and Prucha, 2001, p. 229) and $\hat{\beta} = \beta_0 + o_p(1)$. Similarly, $\hat{\mathbf{u}}'\mathbf{M}_A\hat{\mathbf{u}} = \mathbf{u}'\mathbf{M}_A\mathbf{u} + o_p(1)$. $\sqrt{2\sigma_1^4 b_A} > 0$ and $\sqrt{2\sigma_\nu^4 (T-1)b_A} > 0$, since $\sigma_1^2, \sigma_\nu^2 > 0$ and $b_A > 0$ by Assumptions A1 and A2. As shown in Appendix B, $\hat{\sigma}_1^2 = \sigma_1^2 + o_p(1)$ and $\hat{\sigma}_\nu^2 = \sigma_\nu^2 + o_p(1)$. Then, Theorem 2 of Kelejian and Prucha (2001, p. 230) implies that $\frac{\hat{G}_A}{\sqrt{2\hat{\sigma}_1^4 b_A^2}} = \frac{G_A}{\sqrt{2\sigma_1^4 b_A^2}} + o_p(1)$ and $\frac{\hat{M}_A}{\sqrt{2\hat{\sigma}_\nu^4 (T-1)b_A}} = \frac{M_A}{\sqrt{2\sigma_\nu^4 (T-1)b_A}} + o_p(1)$. Furthermore, this theorem establishes that $\frac{\hat{G}_A}{\sqrt{2\hat{\sigma}_1^4 b_A^2}}$ and $\frac{\hat{M}_A}{\sqrt{2\hat{\sigma}_\nu^4 (T-1)b_A}}$ converge in distribution to a standard normal.

Combining these results, LM_A is a sum of squares of two independent standardized quadratic forms of random variables, which are both asymptotically standard normal under the adopted assumptions. Hence, LM_A is asymptotically distributed as χ_2^2 under H_0^A . ■

Appendix D: LM Test for the Anselin Model

This Appendix derives the LM test for the null hypothesis that the spatial correlation follows the specification described in Anselin (1988). This is given by $H_0^B : \rho_1 = 0$.

Under $H_0^B : \rho_1 = 0$; $\mathbf{A} = \mathbf{I}_N$ and $\mathbf{\Omega}_u = \bar{\mathbf{J}}_T \otimes (T\sigma_\mu^2 \mathbf{I}_N + \sigma_\nu^2 (\mathbf{B}'\mathbf{B})^{-1}) + \sigma_\nu^2 (\mathbf{E}_T \otimes (\mathbf{B}'\mathbf{B})^{-1})$, $\mathbf{\Omega}_u^{-1} = \bar{\mathbf{J}}_T \otimes (T\sigma_\mu^2 \mathbf{I}_N + \sigma_\nu^2 (\mathbf{B}'\mathbf{B})^{-1})^{-1} + \frac{1}{\sigma_\nu^2} (\mathbf{E}_T \otimes (\mathbf{B}'\mathbf{B}))$. Using the general formulas for the score and the information submatrix given

above, we get

$$\begin{aligned} \left. \frac{\partial L}{\partial \sigma_\nu^2} \right|_{H_0^B} &= -\frac{1}{2} \text{tr}[(T\sigma_\mu^2 \mathbf{B}'\mathbf{B} + \sigma_\nu^2 \mathbf{I}_N)^{-1}] - \frac{N(T-1)}{2\sigma_\nu^2} \\ &\quad + \frac{1}{2} \mathbf{u}' [\bar{\mathbf{J}}_T \otimes (T\sigma_\mu^2 \mathbf{B}'\mathbf{B} + \sigma_\nu^2 \mathbf{I}_N)^{-1} (T\sigma_\mu^2 \mathbf{I}_N + \sigma_\nu^2 \mathbf{B}'\mathbf{B})^{-1} \\ &\quad + \frac{1}{\sigma_\nu^4} (\mathbf{E}_T \otimes \mathbf{B}'\mathbf{B})] \mathbf{u} \end{aligned}$$

$$\begin{aligned} \left. \frac{\partial L}{\partial \sigma_\mu^2} \right|_{H_0^B} &= -\frac{1}{2} T \text{tr}[(T\sigma_\mu^2 \mathbf{I}_N + \sigma_\nu^2 \mathbf{B}'\mathbf{B})^{-1}] \\ &\quad + \frac{1}{2} \mathbf{u}' [\mathbf{J}_T \otimes (T\sigma_\mu^2 \mathbf{I}_N + \sigma_\nu^2 \mathbf{B}'\mathbf{B})^{-2}] \mathbf{u} \end{aligned}$$

$$\begin{aligned} \left. \frac{\partial L}{\partial \rho_1} \right|_{H_0^B} &= -\frac{1}{2} T \sigma_\mu^2 \text{tr}[(T\sigma_\mu^2 \mathbf{I}_N + \sigma_\nu^2 (\mathbf{B}'\mathbf{B})^{-1})^{-1} (\mathbf{W}' + \mathbf{W})] \\ &\quad + \frac{1}{2} \sigma_\mu^2 \mathbf{u}' [\mathbf{J}_T \otimes (T\sigma_\mu^2 \mathbf{I}_N + \sigma_\nu^2 (\mathbf{B}'\mathbf{B})^{-1})^{-1} (\mathbf{W}' + \mathbf{W}) (T\sigma_\mu^2 \mathbf{I}_N + \sigma_\nu^2 (\mathbf{B}'\mathbf{B})^{-1})^{-1}] \mathbf{u} \end{aligned}$$

$$\begin{aligned} \left. \frac{\partial L}{\partial \rho_2} \right|_{H_0^B} &= -\frac{1}{2} \sigma_\nu^2 \text{tr}[(T\sigma_\mu^2 \mathbf{I}_N + \sigma_\nu^2 \mathbf{B}'\mathbf{B})^{-1} (\mathbf{B}'\mathbf{B})^{-1} (\mathbf{W}'\mathbf{B} + \mathbf{B}'\mathbf{W}) (\mathbf{B}'\mathbf{B})^{-1}] \\ &\quad - \frac{(T-1)}{2} \text{tr}[(\mathbf{W}'\mathbf{B} + \mathbf{B}'\mathbf{W}) (\mathbf{B}'\mathbf{B})^{-1}] \\ &\quad + \frac{1}{2} \mathbf{u}' [\sigma_\nu^2 \bar{\mathbf{J}}_T \otimes (T\sigma_\mu^2 \mathbf{I}_N + \sigma_\nu^2 (\mathbf{B}'\mathbf{B})^{-1})^{-1} (\mathbf{B}'\mathbf{B})^{-1} (\mathbf{W}'\mathbf{B} + \mathbf{B}'\mathbf{W}) (\mathbf{B}'\mathbf{B})^{-1} \\ &\quad (T\sigma_\mu^2 \mathbf{I}_N + \sigma_\nu^2 (\mathbf{B}'\mathbf{B})^{-1})^{-1} + \frac{1}{\sigma_\nu^2} (\mathbf{E}_T \otimes (\mathbf{W}'\mathbf{B} + \mathbf{B}'\mathbf{W}))] \mathbf{u} \end{aligned}$$

and the elements of the information matrix are determined as

$$\begin{aligned} J_{11}|_{H_0^B} &= \frac{1}{2} \text{tr} [(T\sigma_\mu^2 \mathbf{B}'\mathbf{B} + \sigma_\nu^2 \mathbf{I}_N)^{-1}]^2 + \frac{N(T-1)}{2\sigma_\nu^4} \\ J_{12}|_{H_0^B} &= \frac{T}{2} \text{tr} [(T\sigma_\mu^2 \mathbf{B}'\mathbf{B} + \sigma_\nu^2 \mathbf{I}_N)^{-1} (T\sigma_\mu^2 \mathbf{I}_N + \sigma_\nu^2 (\mathbf{B}'\mathbf{B})^{-1})^{-1}] \\ J_{13}|_{H_0^B} &= \frac{T\sigma_\mu^2}{2} \text{tr} [(T\sigma_\mu^2 \mathbf{B}'\mathbf{B} + \sigma_\nu^2 \mathbf{I}_N)^{-1} (T\sigma_\mu^2 + \sigma_\nu^2 (\mathbf{B}'\mathbf{B})^{-1})^{-1} (\mathbf{W}' + \mathbf{W})] \\ J_{14}|_{H_0^B} &= \frac{\sigma_\nu^2}{2} \text{tr} [(T\sigma_\mu^2 \mathbf{B}'\mathbf{B} + \sigma_\nu^2 \mathbf{I}_N)^{-1} (T\sigma_\mu^2 + \sigma_\nu^2 (\mathbf{B}'\mathbf{B})^{-1})^{-1} \\ &\quad (\mathbf{B}'\mathbf{B})^{-1} (\mathbf{W}'\mathbf{B} + \mathbf{B}'\mathbf{W}) (\mathbf{B}'\mathbf{B})^{-1}] + \frac{(T-1)}{2\sigma_\nu^2} \text{tr} [(\mathbf{W}'\mathbf{B} + \mathbf{B}'\mathbf{W}) (\mathbf{B}'\mathbf{B})^{-1}] \end{aligned}$$

$$\begin{aligned}
J_{22}|_{H_0^B} &= \frac{T^2}{2} \text{tr} [(T\sigma_\mu^2 \mathbf{I}_N + \sigma_\nu^2 (\mathbf{B}'\mathbf{B})^{-1})^{-1}]^2 \\
J_{23}|_{H_0^B} &= \frac{T^2 \sigma_\mu^2}{2} \text{tr} [(T\sigma_\mu^2 \mathbf{I}_N + \sigma_\nu^2 (\mathbf{B}'\mathbf{B})^{-1})^{-2} (\mathbf{W}' + \mathbf{W})] \\
J_{24}|_{H_0^B} &= \frac{T \sigma_\nu^2}{2} \text{tr} [(T\sigma_\mu^2 \mathbf{I}_N + \sigma_\nu^2 (\mathbf{B}'\mathbf{B})^{-1})^{-2} (\mathbf{B}'\mathbf{B})^{-1} (\mathbf{W}'\mathbf{B} + \mathbf{B}'\mathbf{W}) (\mathbf{B}'\mathbf{B})^{-1}] \\
J_{33}|_{H_0^B} &= \frac{T^2 \sigma_\mu^4}{2} \text{tr} [((T\sigma_\mu^2 \mathbf{I}_N + \sigma_\nu^2 (\mathbf{B}'\mathbf{B})^{-1})^{-1} (\mathbf{W}' + \mathbf{W}))^2] \\
J_{34}|_{H_0^B} &= \frac{T \sigma_\mu^2 \sigma_\nu^2}{2} \text{tr} [(T\sigma_\mu^2 \mathbf{I}_N + \sigma_\nu^2 (\mathbf{B}'\mathbf{B})^{-1})^{-1} (\mathbf{W}' + \mathbf{W}) \\
&\quad (T\sigma_\mu^2 \mathbf{I}_N + \sigma_\nu^2 (\mathbf{B}'\mathbf{B})^{-1})^{-1} (\mathbf{B}'\mathbf{B})^{-1} (\mathbf{W}'\mathbf{B} + \mathbf{B}'\mathbf{W}) (\mathbf{B}'\mathbf{B})^{-1}] \\
J_{44}|_{H_0^B} &= \frac{\sigma_\nu^4}{2} \text{tr} [((T\sigma_\mu^2 \mathbf{I}_N + \sigma_\nu^2 (\mathbf{B}'\mathbf{B})^{-1})^{-1} (\mathbf{B}'\mathbf{B})^{-1} (\mathbf{W}'\mathbf{B} + \mathbf{B}'\mathbf{W}) (\mathbf{B}'\mathbf{B})^{-1})^2] \\
&\quad + \frac{(T-1)}{2} \text{tr} [((\mathbf{W}'\mathbf{B} + \mathbf{B}'\mathbf{W}) (\mathbf{B}'\mathbf{B})^{-1})^2].
\end{aligned}$$

The LM test for H_0^B makes use of the estimated score $\widehat{\mathbf{D}}_\theta = [0, 0, \widehat{d}_{\rho_1}, 0]'$ with

$$\widehat{d}_{\rho_1} = \left. \frac{\partial L}{\partial \rho_1} \right|_{H_0^B} = -\frac{T \widehat{\sigma}_\mu^2}{2} \text{tr} [\widehat{\mathbf{C}}_1 \mathbf{C}_2] + \frac{\widehat{\sigma}_\nu^2}{2} \widehat{\mathbf{u}}' (\mathbf{J}_T \otimes \widehat{\mathbf{C}}_1 \mathbf{C}_2 \widehat{\mathbf{C}}_1) \widehat{\mathbf{u}},$$

where $\widehat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}$, $\widehat{\mathbf{C}}_1 = (T\widehat{\sigma}_\mu^2 \mathbf{I}_N + \widehat{\sigma}_\nu^2 (\widehat{\mathbf{B}}'\widehat{\mathbf{B}})^{-1})^{-1}$ and $\mathbf{C}_2 = (\mathbf{W}' + \mathbf{W})$. An estimate of the lower (4×4) block of the information matrix $\widehat{\mathbf{J}}_\theta$ under H_0^B is given by

$$\widehat{\mathbf{J}}_\theta \Big|_{H_0^B} = \begin{bmatrix} \frac{1}{2} \text{tr} [\widehat{\mathbf{C}}_3^2] + \frac{N(T-1)}{2\widehat{\sigma}_\nu^4} & \frac{T}{2} \text{tr} [\widehat{\mathbf{C}}_3 \widehat{\mathbf{C}}_1] & \frac{T \widehat{\sigma}_\mu^2}{2} \text{tr} [\widehat{\mathbf{C}}_3 \widehat{\mathbf{C}}_1 \mathbf{C}_2] & \frac{\widehat{\sigma}_\nu^2}{2} \text{tr} [\widehat{\mathbf{C}}_3 \widehat{\mathbf{C}}_1 \widehat{\mathbf{C}}_5] + \frac{(T-1)}{2\widehat{\sigma}_\nu^2} \text{tr} [\widehat{\mathbf{C}}_4] \\ \frac{T}{2} \text{tr} [\widehat{\mathbf{C}}_3 \widehat{\mathbf{C}}_1] & \frac{T^2}{2} \text{tr} [\widehat{\mathbf{C}}_1^2] & \frac{T^2 \widehat{\sigma}_\mu^2}{2} \text{tr} [\widehat{\mathbf{C}}_1^2 \mathbf{C}_2] & \frac{T \widehat{\sigma}_\nu^2}{2} \text{tr} [\widehat{\mathbf{C}}_1^2 \widehat{\mathbf{C}}_5] \\ \frac{T \widehat{\sigma}_\mu^2}{2} \text{tr} [\widehat{\mathbf{C}}_3 \widehat{\mathbf{C}}_1 \mathbf{C}_2] & \frac{T^2 \widehat{\sigma}_\mu^2}{2} \text{tr} [\widehat{\mathbf{C}}_1^2 \mathbf{C}_2] & \frac{T^2 \widehat{\sigma}_\mu^4}{2} \text{tr} [(\widehat{\mathbf{C}}_1 \mathbf{C}_2)^2] & \frac{T \widehat{\sigma}_\mu^2 \widehat{\sigma}_\nu^2}{2} \text{tr} [\widehat{\mathbf{C}}_1 \mathbf{C}_2 \widehat{\mathbf{C}}_1 \widehat{\mathbf{C}}_5] \\ \frac{\widehat{\sigma}_\nu^2}{2} \text{tr} [\widehat{\mathbf{C}}_3 \widehat{\mathbf{C}}_1 \widehat{\mathbf{C}}_5] + \frac{(T-1)}{2\widehat{\sigma}_\nu^2} \text{tr} [\widehat{\mathbf{C}}_4] & \frac{T \widehat{\sigma}_\nu^2}{2} \text{tr} [\widehat{\mathbf{C}}_1^2 \widehat{\mathbf{C}}_5] & \frac{T \widehat{\sigma}_\mu^2 \widehat{\sigma}_\nu^2}{2} \text{tr} [\widehat{\mathbf{C}}_1 \mathbf{C}_2 \widehat{\mathbf{C}}_1 \widehat{\mathbf{C}}_5] & \frac{\widehat{\sigma}_\nu^4}{2} \text{tr} [(\widehat{\mathbf{C}}_1 \widehat{\mathbf{C}}_5)^2] + \frac{(T-1)}{2} \text{tr} [\widehat{\mathbf{C}}_4^2] \end{bmatrix},$$

where $\widehat{\mathbf{C}}_3 = (\widehat{\mathbf{B}}'\widehat{\mathbf{B}})^{-1} \widehat{\mathbf{C}}_1$, $\widehat{\mathbf{C}}_4 = (\mathbf{W}'\widehat{\mathbf{B}} + \widehat{\mathbf{B}}'\mathbf{W}) (\widehat{\mathbf{B}}'\widehat{\mathbf{B}})^{-1}$ and $\widehat{\mathbf{C}}_5 = (\widehat{\mathbf{B}}'\widehat{\mathbf{B}})^{-1} \widehat{\mathbf{C}}_4$.

The LM test for H_0^B does not have a simple closed form. However, an alternative closed form expression for this LM statistic based on the square of the standardized score is given by $LM'_B = \frac{(\widehat{\mathbf{u}}'\widehat{\mathbf{G}}_B \widehat{\mathbf{u}} - \widehat{g}_B)^2}{2\widehat{b}_B}$, where $\widehat{\mathbf{G}}_B = \mathbf{J}_T \otimes \widehat{\mathbf{C}}_1 \mathbf{C}_2 \widehat{\mathbf{C}}_1$, $\widehat{g}_B = \frac{T \widehat{\sigma}_\mu^2}{2} \text{tr} [\widehat{\mathbf{C}}_1 \mathbf{C}_2]$ and $\widehat{b}_B = \text{tr} [(\widehat{\mathbf{C}}_1 \mathbf{C}_2)^2]$.

Theorem 8 (LM_B) Suppose Assumptions A1 - A5 hold and $H_0^B : \rho_1 = 0$ is true. Then, $LM'_B = \frac{(\widehat{\mathbf{u}}'\widehat{\mathbf{G}}_B\widehat{\mathbf{u}} - \widehat{g}_B)^2}{2\widehat{b}_B}$ is asymptotically distributed as χ_1^2 .

Proof. Using the residuals and the parameters of the true model under H_0^B , the score is given by

$$d_{\rho_1} = \frac{1}{2}T\sigma_\mu^2 G_B,$$

where $G_B = (\mathbf{u}'\mathbf{G}_B\mathbf{u} - g_B)$ and $\mathbf{G}_B = (\bar{\mathbf{J}}_T \otimes \mathbf{C}_1\mathbf{C}_2\mathbf{C}_1)$, using $\mathbf{C}_1 = (T\sigma_\mu^2\mathbf{I}_N + \sigma_\nu^2(\mathbf{B}'\mathbf{B})^{-1})^{-1}$, $\mathbf{C}_2 = (\mathbf{W}' + \mathbf{W})$ and $g_B = \text{tr}[\mathbf{C}_1\mathbf{C}_2]$.

(i) Under H_0^B $\mathbf{A} = \mathbf{I}_N$ and we can apply Lemma 4 setting $\mathbf{H} = \mathbf{C}_1\mathbf{C}_2\mathbf{C}_1$ so that $E[\mathbf{u}'\mathbf{G}_B\mathbf{u}] = T\sigma_\mu^2\text{tr}[\mathbf{H}] + \sigma_\nu^2\text{tr}[\mathbf{H}(\mathbf{B}'\mathbf{B})^{-1}] = T\sigma_\mu^2\text{tr}[\mathbf{C}_1\mathbf{C}_2\mathbf{C}_1] + \sigma_\nu^2\text{tr}[\mathbf{C}_1\mathbf{C}_2\mathbf{C}_1(\mathbf{B}'\mathbf{B})^{-1}] = \text{tr}[\mathbf{C}_1\mathbf{C}_2\mathbf{C}_1(T\sigma_\mu^2\mathbf{I}_N + \sigma_\nu^2(\mathbf{B}'\mathbf{B})^{-1})] = \text{tr}[\mathbf{C}_1\mathbf{C}_2] = g_B$. Observe, $T\sigma_\mu^2\mathbf{H} + \sigma_\nu^2\mathbf{H}(\mathbf{B}'\mathbf{B})^{-1} = \mathbf{C}_1\mathbf{C}_2$.

$\text{Var}[\mathbf{u}'\mathbf{G}_B\mathbf{u}] = 2(T^2\sigma_\mu^4\text{tr}[\mathbf{H}^2] + 2T\sigma_\nu^2\sigma_\mu^2\text{tr}[\mathbf{H}^2(\mathbf{B}'\mathbf{B})^{-1}] + \sigma_\nu^4\text{tr}[(\mathbf{H}(\mathbf{B}'\mathbf{B})^{-1})^2])$, and $\text{tr}[(T\sigma_\mu^2\mathbf{H} + \sigma_\nu^2\mathbf{H}(\mathbf{B}'\mathbf{B})^{-1})^2] = \text{tr}[T^2\sigma_\mu^4\mathbf{H}^2 + 2T\sigma_\nu^2\sigma_\mu^2\mathbf{H}^2(\mathbf{B}'\mathbf{B})^{-1} + \sigma_\nu^4(\mathbf{H}(\mathbf{B}'\mathbf{B})^{-1})^2] = \text{tr}[(\mathbf{C}_1\mathbf{C}_2)^2]$. Hence, by Lemma 4 $\frac{G_B}{\sigma_1^2\sqrt{2b_B}}$ with $b_B = \text{tr}[(\mathbf{C}_1\mathbf{C}_2)^2]$ converges in distribution to the standard normal.

(ii) As a result, we obtain $\text{var}[d_{\rho_1}] = \frac{1}{4}T^2\sigma_\mu^4\text{Var}[\mathbf{u}'\mathbf{G}_B\mathbf{u}] = \frac{1}{2}T^2\sigma_\mu^4b_B$. Hence, the square of the standardized score statistic is given by $LM'_B = \frac{\frac{1}{2}T^2\sigma_\mu^4G_B^2}{\frac{1}{2}T^2\sigma_\mu^4b_B} = \frac{G_B^2}{2b_B}$. By analogy to the arguments in the proof of Theorem 7, we conclude that the quadratic form $\frac{\widehat{G}_B}{\sqrt{2\widehat{b}_B}}$ converges in distribution towards a standard normal under H_0^B . Hence, LM'_B is asymptotically distributed as χ_1^2 under H_0^B . ■

Appendix E: LM Test for the KKP Model

To derive the asymptotic distribution of the LM test for H_0^C , it proves useful to re-parameterize the model so that $\rho_1 = \rho_2 + \Delta$ and to test $H_0^B : \Delta = 0$ vs. $H_1^B : \Delta \neq 0$, i.e., that the spatial panel correlation follows the specification proposed by KKP.

Under H_0^C , $\mathbf{B} = \mathbf{A}$, $\boldsymbol{\Omega}_u = (\sigma_1^2 \bar{\mathbf{J}}_T + \sigma_\nu^2 \mathbf{E}_T) \otimes (\mathbf{A}'\mathbf{A})^{-1}$ and $\boldsymbol{\Omega}_u^{-1} = (\frac{1}{\sigma_1^2} \bar{\mathbf{J}}_T + \frac{1}{\sigma_\nu^2} \mathbf{E}_T) \otimes (\mathbf{A}'\mathbf{A})$. Using the general formulas for the score and for the information matrix given above, we get

$$\begin{aligned}
\left. \frac{\partial L}{\partial \sigma_\nu^2} \right|_{H_0^C} &= -\frac{N}{2\sigma_1^2} - \frac{N(T-1)}{2\sigma_\nu^2} + \frac{1}{2} \mathbf{u}' [(\frac{1}{\sigma_1^4} \bar{\mathbf{J}}_T + \frac{1}{\sigma_\nu^2} \mathbf{E}_T) \otimes \mathbf{A}'\mathbf{A}] \mathbf{u} \\
\left. \frac{\partial L}{\partial \sigma_\mu^2} \right|_{H_0^C} &= -\frac{NT}{2\sigma_1^2} + \frac{1}{2} \mathbf{u}' [\frac{T}{\sigma_1^4} (\bar{\mathbf{J}}_T \otimes \mathbf{A}'\mathbf{A})] \mathbf{u} \\
\left. \frac{\partial L}{\partial \Delta} \right|_{H_0^C} &= -\frac{T\sigma_\mu^2}{2\sigma_1^2} \text{tr}[\mathbf{D}] + \frac{1}{2} \mathbf{u}' (\frac{T\sigma_\mu^2}{\sigma_1^4} \bar{\mathbf{J}}_T \otimes \mathbf{F}) \mathbf{u} \\
\left. \frac{\partial L}{\partial \rho_2} \right|_{H_0^C} &= -\frac{T\sigma_\mu^2}{2\sigma_1^2} \text{tr}[\mathbf{D}] + \frac{1}{2} \mathbf{u}' (\frac{T\sigma_\mu^2}{\sigma_1^4} \bar{\mathbf{J}}_T \otimes \mathbf{F}) \mathbf{u} \\
&\quad - \frac{1}{2} [\frac{\sigma_\nu^2}{\sigma_1^2} + (T-1)] \text{tr}[\mathbf{D}] + \frac{1}{2} \mathbf{u}' [(\frac{\sigma_\nu^2}{\sigma_1^2} \bar{\mathbf{J}}_T + \frac{1}{\sigma_\nu^2} \mathbf{E}_T) \otimes \mathbf{F}] \mathbf{u} \\
&= -\frac{T}{2} \text{tr}[\mathbf{D}] + \frac{1}{2} \mathbf{u}' [(\frac{1}{\sigma_1^2} \bar{\mathbf{J}}_T + \frac{1}{\sigma_\nu^2} \mathbf{E}_T) \otimes \mathbf{F}] \mathbf{u},
\end{aligned}$$

where $\mathbf{F} = \mathbf{W}'\mathbf{A} + \mathbf{A}'\mathbf{W}$ and $\mathbf{D} = \mathbf{F}(\mathbf{A}'\mathbf{A})^{-1}$. The elements of the relevant part of the information matrix are

$$\mathbf{J}_\theta|_{H_0^C} = \begin{bmatrix} \frac{N}{2\sigma_1^4} + \frac{N(T-1)}{2\sigma_\nu^4} & \frac{NT}{2\sigma_1^4} & \frac{T\sigma_\mu^2}{2\sigma_1^4} \text{tr}[\mathbf{D}] & \left(\frac{\sigma_\nu^2}{2\sigma_1^4} + \frac{(T-1)}{2\sigma_\nu^2} \right) \text{tr}[\mathbf{D}] \\ \frac{NT}{2\sigma_1^4} & \frac{NT^2}{2\sigma_1^4} & \frac{T^2\sigma_\mu^2}{2\sigma_1^4} \text{tr}[\mathbf{D}] & \frac{T\sigma_\nu^2}{2\sigma_1^4} \text{tr}[\mathbf{D}] \\ \frac{T\sigma_\mu^2}{2\sigma_1^4} \text{tr}[\mathbf{D}] & \frac{T^2\sigma_\mu^2}{2\sigma_1^4} \text{tr}[\mathbf{D}] & \frac{T^2\sigma_\mu^4}{2\sigma_1^4} \text{tr}[\mathbf{D}^2] & \frac{T\sigma_\mu^2\sigma_\nu^2}{2\sigma_1^4} \text{tr}[\mathbf{D}^2] \\ \left(\frac{\sigma_\nu^2}{2\sigma_1^4} + \frac{(T-1)}{2\sigma_\nu^2} \right) \text{tr}[\mathbf{D}] & \frac{T\sigma_\nu^2}{2\sigma_1^4} \text{tr}[\mathbf{D}] & \frac{T\sigma_\mu^2\sigma_\nu^2}{2\sigma_1^4} \text{tr}[\mathbf{D}^2] & \left(\frac{\sigma_\nu^4}{2\sigma_1^4} + \frac{(T-1)}{2} \right) \text{tr}[\mathbf{D}^2] \end{bmatrix}.$$

The restricted MLE estimates under H_0^C are labeled by a bar. In fact, this gives the MLE version of the KKP model and $\bar{\mathbf{u}} = \mathbf{y} - \mathbf{X}\bar{\boldsymbol{\beta}}$. The score with

respect to each element of $\boldsymbol{\theta}$ evaluated at the restricted MLE $\bar{\boldsymbol{\theta}}$ is given by

$$\bar{\mathbf{D}}_{\boldsymbol{\theta}} = \begin{bmatrix} 0 \\ 0 \\ \frac{T\bar{\sigma}_{\mu}^2}{2\bar{\sigma}_1^4} [-\bar{\sigma}_1^2 \text{tr}[\bar{\mathbf{D}}] + \bar{\mathbf{u}}'(\bar{\mathbf{J}}_T \otimes \bar{\mathbf{F}})\bar{\mathbf{u}}] \\ 0 \end{bmatrix}.$$

Using $\bar{d}_C = \text{tr}[\bar{\mathbf{D}}]$ and $\bar{e}_C = \text{tr}[\bar{\mathbf{D}}^2]$, the lower (4×4) block of the estimated information matrix evaluated at the restricted MLE $\bar{\boldsymbol{\theta}}$ is given by

$$\begin{aligned} \bar{\mathbf{J}}_{\boldsymbol{\theta}} &= \frac{1}{2\bar{\sigma}_1^4} \begin{bmatrix} NT \begin{bmatrix} \frac{(T-1)\bar{\sigma}_1^4 + \bar{\sigma}_{\nu}^4}{T\bar{\sigma}_{\nu}^4} & 1 \\ 1 & T \end{bmatrix} & T\bar{d}_C \begin{bmatrix} \bar{\sigma}_{\mu}^2 & \frac{(T-1)\bar{\sigma}_1^4 + \bar{\sigma}_{\nu}^2\bar{\sigma}_1^2}{T\bar{\sigma}_{\nu}^2} \\ T\bar{\sigma}_{\mu}^2 & \bar{\sigma}_1^2 \end{bmatrix} \\ T\bar{d}_C \begin{bmatrix} \bar{\sigma}_{\mu}^2 & T\bar{\sigma}_{\mu}^2 \\ \frac{(T-1)\bar{\sigma}_1^4 + \bar{\sigma}_{\nu}^2\bar{\sigma}_1^2}{T\bar{\sigma}_{\nu}^2} & \bar{\sigma}_1^2 \end{bmatrix} & T\bar{e}_C \begin{bmatrix} T\bar{\sigma}_{\mu}^4 & \bar{\sigma}_1^2\bar{\sigma}_{\mu}^2 \\ \bar{\sigma}_1^2\bar{\sigma}_{\mu}^2 & \bar{\sigma}_1^4 \end{bmatrix} \end{bmatrix} \\ &= \frac{1}{2\bar{\sigma}_1^4} \begin{bmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{21} & \mathbf{M}_{22} \end{bmatrix}. \end{aligned}$$

To derive the lower right block of the inverse $\bar{\mathbf{J}}_{\boldsymbol{\theta}}^{-1}$, we employ the formula for the partitioned inverse so that $\bar{\mathbf{J}}_{\Delta, \rho_2}^{-1} = 2\bar{\sigma}_1^4(\mathbf{M}_{22} - \mathbf{M}_{21}\mathbf{M}_{11}^{-1}\mathbf{M}_{12})^{-1}$:

$$\begin{aligned} \mathbf{M}_{21}\mathbf{M}_{11}^{-1}\mathbf{M}_{12} &= \frac{\bar{\sigma}_{\nu}^4 T^2 \bar{d}_C^2}{NT(T-1)\bar{\sigma}_1^4} \begin{bmatrix} \bar{\sigma}_{\mu}^2 & T\bar{\sigma}_{\mu}^2 \\ \frac{(T-1)\bar{\sigma}_1^4 + \bar{\sigma}_{\nu}^2\bar{\sigma}_1^2}{T\bar{\sigma}_{\nu}^2} & \bar{\sigma}_1^2 \end{bmatrix} \begin{bmatrix} T & -1 \\ -1 & \frac{(T-1)\bar{\sigma}_1^4 + \bar{\sigma}_{\nu}^4}{T\bar{\sigma}_{\nu}^4} \end{bmatrix} \begin{bmatrix} \bar{\sigma}_{\mu}^2 & \frac{(T-1)\bar{\sigma}_1^4 + \bar{\sigma}_{\nu}^2\bar{\sigma}_1^2}{T\bar{\sigma}_{\nu}^2} \\ T\bar{\sigma}_{\mu}^2 & \bar{\sigma}_1^2 \end{bmatrix} \\ &= \frac{\bar{\sigma}_{\nu}^4 T \bar{d}_C^2}{N(T-1)\bar{\sigma}_1^4} \begin{bmatrix} \bar{\sigma}_{\mu}^2 & T\bar{\sigma}_{\mu}^2 \\ \frac{(T-1)\bar{\sigma}_1^4 + \bar{\sigma}_{\nu}^2\bar{\sigma}_1^2}{T\bar{\sigma}_{\nu}^2} & \bar{\sigma}_1^2 \end{bmatrix} \begin{bmatrix} 0 & \frac{(T-1)\bar{\sigma}_1^4}{\bar{\sigma}_{\nu}^2} \\ \frac{(T-1)\bar{\sigma}_1^4\bar{\sigma}_{\mu}^2}{\bar{\sigma}_{\nu}^4} & \frac{(T-1)\bar{\sigma}_1^4\bar{\sigma}_{\mu}^2}{\bar{\sigma}_{\nu}^4} \end{bmatrix} \\ &= \frac{T\bar{d}_C^2}{N} \begin{bmatrix} \bar{\sigma}_{\mu}^2 & T\bar{\sigma}_{\mu}^2 \\ \frac{(T-1)\bar{\sigma}_1^4 + \bar{\sigma}_{\nu}^2\bar{\sigma}_1^2}{T\bar{\sigma}_{\nu}^2} & \bar{\sigma}_1^2 \end{bmatrix} \begin{bmatrix} 0 & \bar{\sigma}_{\nu}^2 \\ \bar{\sigma}_{\mu}^2 & \bar{\sigma}_{\mu}^2 \end{bmatrix} \\ &= \frac{T\bar{d}_C^2}{N} \begin{bmatrix} T\bar{\sigma}_{\mu}^4 & \bar{\sigma}_1^2\bar{\sigma}_{\mu}^2 \\ \bar{\sigma}_1^2\bar{\sigma}_{\mu}^2 & \bar{\sigma}_1^4 \end{bmatrix}, \end{aligned}$$

using $|\mathbf{M}_{11}| = NT \frac{(T-1)\bar{\sigma}_1^4}{\bar{\sigma}_\nu^4}$.

$$\begin{aligned} \mathbf{M}_{22} - \mathbf{M}_{21}\mathbf{M}_{11}^{-1}\mathbf{M}_{12} &= T e_C \begin{bmatrix} T\bar{\sigma}_\mu^4 & \bar{\sigma}_1^2\bar{\sigma}_\mu^2 \\ \bar{\sigma}_1^2\bar{\sigma}_\nu^2 & \bar{\sigma}_1^4 \end{bmatrix} - \frac{Td_C^2}{N} \begin{bmatrix} T\bar{\sigma}_\mu^4 & \bar{\sigma}_1^2\bar{\sigma}_\mu^2 \\ \bar{\sigma}_1^2\bar{\sigma}_\mu^2 & \bar{\sigma}_1^4 \end{bmatrix} \\ &= T \left(e_C - \frac{d_C^2}{N} \right) \begin{bmatrix} T\bar{\sigma}_\mu^4 & \bar{\sigma}_1^2\bar{\sigma}_\mu^2 \\ \bar{\sigma}_1^2\bar{\sigma}_\mu^2 & \bar{\sigma}_1^4 \end{bmatrix} \end{aligned}$$

$$\bar{\mathbf{J}}_{\Delta, \rho_2}^{-1} = \frac{2\bar{\sigma}_1^4}{T\bar{b}_C} \begin{bmatrix} T\bar{\sigma}_\mu^4 & \bar{\sigma}_1^2\bar{\sigma}_\mu^2 \\ \bar{\sigma}_1^2\bar{\sigma}_\mu^2 & \bar{\sigma}_1^4 \end{bmatrix}^{-1} = \frac{2}{T(T-1)\bar{b}_C\bar{\sigma}_\mu^4} \begin{bmatrix} \bar{\sigma}_1^4 & -\bar{\sigma}_1^2\bar{\sigma}_\mu^2 \\ -\bar{\sigma}_1^2\bar{\sigma}_\mu^2 & T\bar{\sigma}_\mu^4 \end{bmatrix},$$

where $\bar{b}_C = \bar{e}_C - \bar{d}_C^2/N$. Defining $\bar{G}_C = \bar{\mathbf{u}}'(\bar{\mathbf{J}}_T \otimes \bar{\mathbf{F}})\bar{\mathbf{u}} - \bar{\sigma}_1^2 tr[\bar{\mathbf{D}}]$ the resulting LM statistic for H_0^C is given by

$$LM_C = \bar{\mathbf{D}}_\theta' \bar{\mathbf{J}}_\theta^{-1} \bar{\mathbf{D}}_\theta = \frac{T}{2\bar{b}_C(T-1)\bar{\sigma}_1^4} \bar{G}_C^2.$$

Theorem 9 (LM_C) *Suppose Assumptions A1 - A5 hold and $H_0^c: \rho_1 = \rho_2 = \rho$ is true. Let $\bar{\mathbf{F}} = (\mathbf{W}'\bar{\mathbf{A}} + \bar{\mathbf{A}}'\mathbf{W})$, $\bar{\mathbf{D}} = \bar{\mathbf{F}}(\bar{\mathbf{A}}'\bar{\mathbf{A}})^{-1}$ and $\bar{G}_C = \bar{\mathbf{u}}'(\bar{\mathbf{J}}_T \otimes \bar{\mathbf{F}})\bar{\mathbf{u}} - \bar{\sigma}_1^2 tr[\bar{\mathbf{D}}]$. Then, $LM'_C = \frac{1}{2\bar{b}_C\bar{\sigma}_1^4} \bar{G}_C^2$ with $\bar{b}'_C = \bar{e}_C = tr[\bar{\mathbf{D}}^2]$ is asymptotically distributed as χ_1^2 .*

Proof. Define $\mathbf{G}_C = (\bar{\mathbf{J}}_T \otimes \mathbf{F})$ and $g_C = \sigma_1^2 tr[\mathbf{D}]$ so that the score under H_0^c is given by $d_\Delta = \frac{1}{2}\mathbf{u}'(\frac{T\sigma_\mu^2}{\sigma_1^4}\bar{\mathbf{J}}_T \otimes \mathbf{F})\mathbf{u} - \frac{T\sigma_\mu^2}{2\sigma_1^4} tr[\mathbf{D}] = \frac{T\sigma_\mu^2}{2\sigma_1^4} (\mathbf{u}'\mathbf{G}_C\mathbf{u} - g_C) = \frac{T\sigma_\mu^2}{2\sigma_1^4} G_C$. Note that first, we use the true residuals and the true parameters under H_0^c .

(i) We can apply Lemma 4 with $\mathbf{H} = \mathbf{F}$ and $\mathbf{A} = \mathbf{B}$ to obtain:

$$\begin{aligned} E[\mathbf{u}'(\bar{\mathbf{J}}_T \otimes \mathbf{F})\mathbf{u}] &= T\sigma_\mu^2 tr[\mathbf{F}(\mathbf{A}'\mathbf{A})^{-1}] + \sigma_\nu^2 tr[\mathbf{F}(\mathbf{A}'\mathbf{A})^{-1}] = \sigma_1^2 tr[\mathbf{D}] = g_C \text{ and} \\ Var[\mathbf{u}'(\bar{\mathbf{J}}_T \otimes \mathbf{F})\mathbf{u}] &= 2(T^2\sigma_\mu^4 tr[(\mathbf{F}(\mathbf{A}'\mathbf{A})^{-1})^2] + 2T\sigma_\nu^2\sigma_\mu^2 tr[\mathbf{F}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{F}(\mathbf{A}'\mathbf{A})^{-1}] + \end{aligned}$$

$\sigma_\nu^4 \text{tr}[(\mathbf{F}(\mathbf{A}'\mathbf{A})^{-1})^2] = 2\sigma_1^4 \text{tr}[\mathbf{D}^2] = 2\sigma_1^4 b'_C$. Since \mathbf{W} is uniformly bounded and $|\rho| \lambda_{\max} < 1$, we conclude that the elements of \mathbf{D} are also uniformly bounded in absolute value and that $\frac{G_C - \sigma_1^2 \text{tr}[\mathbf{D}]}{\sqrt{2\sigma_1^4 b'_C}}$ converges in distribution to the standard normal.

(ii) By analogy to the arguments in the proof of Theorem 7, we conclude that the quadratic form $\frac{\bar{G}_C}{\bar{\sigma}_1^2 \sqrt{2\bar{b}'_C}}$ converges in distribution to $N(0, 1)$ and LM'_C is asymptotically distributed as χ_1^2 under H_0^C . ■

Note that the standardized LM'_C in the proof differs from the formula for LM_C in the text which relies on the normalization $\frac{T-1}{T} \bar{b}_C \bar{\sigma}_1^4$, with $\bar{b}_C = \text{tr}[\bar{\mathbf{D}}^2] - \text{tr}[\bar{\mathbf{D}}]^2/N$.

Appendix F: Numerical optimization

We use the constrained quasi-Newton method involving the constraints $\sigma_\mu^2 > 0$, $\sigma_\nu^2 > 0$, $-1 < \rho_1 < 1$ and $-1 < \rho_2 < 1$ to estimate the parameters of the four models (the unrestricted model and the three restricted ones: random effects, Anselin, and KKP). The quasi-Newton method calculates the gradient of the log-likelihood numerically. We use the optimization routine *fmincon* available from Matlab which uses the sequential quadratic programming method. This method guarantees super-linear convergence by accumulating second order information regarding the Kuhn-Tucker equations using a quasi-Newton updating procedure. An estimate of the Hessian of the Lagrangian is updated at each iteration using the BFGS formula. All tests are based on the analytically derived formulas for both the gradient and the information matrix, using the estimated parameters.

Table 1: Monte carlo simulations for size and power of LM and LR tests of the random effects, the Anselin and the Kapoor-Kelejian-Prucha models; share of rejections in 2000 replications

(N=50, T=5, $\sigma^2_\mu=10$, $\sigma^2_\nu=10$)

		Random effects model		Anselin model			Kelejian-Prucha model		
		$H_0^A: \rho_1=0, \rho_2=0$		$H_0^B: \rho_1=0$			$H_0^C: \rho_1=\rho_2$		
ρ_1	ρ_2	LM _A	LR _A	LM _B	LM _B '	LR _B	LM _C	LM _C '	LR _C
-0.80	-0.80	1.000	1.000	0.938	0.924	0.964	0.039	0.028	0.041
-0.80	-0.50	1.000	1.000	0.985	0.973	0.992	0.590	0.580	0.565
-0.80	-0.20	0.997	0.998	0.989	0.988	0.991	0.919	0.927	0.922
-0.80	0.00	0.979	0.982	0.989	0.992	0.991	0.982	0.993	0.985
-0.80	0.20	0.997	0.997	0.989	0.989	0.993	0.999	0.999	0.999
-0.80	0.50	1.000	1.000	0.972	0.975	0.977	1.000	1.000	1.000
-0.80	0.80	1.000	1.000	0.925	0.936	0.938	1.000	1.000	1.000
-0.50	-0.80	1.000	1.000	0.562	0.570	0.595	0.172	0.135	0.307
-0.50	-0.50	1.000	1.000	0.692	0.687	0.711	0.046	0.046	0.046
-0.50	-0.20	0.913	0.925	0.727	0.752	0.742	0.318	0.315	0.324
-0.50	0.00	0.614	0.646	0.702	0.698	0.729	0.661	0.645	0.685
-0.50	0.20	0.888	0.886	0.690	0.710	0.724	0.868	0.867	0.894
-0.50	0.50	1.000	1.000	0.613	0.610	0.632	0.985	0.980	0.992
-0.50	0.80	1.000	1.000	0.430	0.445	0.450	0.999	0.998	1.000
-0.20	-0.80	1.000	1.000	0.144	0.138	0.153	0.643	0.615	0.755
-0.20	-0.50	1.000	1.000	0.175	0.157	0.183	0.209	0.228	0.231
-0.20	-0.20	0.663	0.669	0.164	0.200	0.167	0.042	0.035	0.045
-0.20	0.00	0.130	0.139	0.158	0.166	0.169	0.157	0.183	0.171
-0.20	0.20	0.696	0.660	0.186	0.187	0.203	0.453	0.466	0.499
-0.20	0.50	1.000	1.000	0.131	0.158	0.142	0.863	0.864	0.910
-0.20	0.80	1.000	1.000	0.095	0.114	0.097	0.976	0.965	0.996
0.00	-0.80	1.000	1.000	0.043	0.025	0.058	0.822	0.814	0.899
0.00	-0.50	1.000	1.000	0.043	0.037	0.055	0.501	0.485	0.509
0.00	-0.20	0.582	0.574	0.045	0.038	0.059	0.106	0.109	0.099
0.00	0.00	0.043	0.053	0.049	0.048	0.058	0.054	0.045	0.059
0.00	0.20	0.646	0.602	0.042	0.041	0.047	0.133	0.154	0.154
0.00	0.50	1.000	1.000	0.049	0.028	0.051	0.595	0.583	0.672
0.00	0.80	1.000	1.000	0.050	0.017	0.053	0.898	0.881	0.962
0.20	-0.80	1.000	1.000	0.117	0.116	0.092	0.962	0.952	0.983
0.20	-0.50	1.000	1.000	0.147	0.139	0.126	0.818	0.830	0.827
0.20	-0.20	0.605	0.593	0.174	0.163	0.142	0.402	0.407	0.382
0.20	0.00	0.130	0.110	0.148	0.141	0.125	0.131	0.154	0.111
0.20	0.20	0.686	0.649	0.171	0.160	0.140	0.048	0.039	0.053
0.20	0.50	1.000	1.000	0.134	0.141	0.116	0.283	0.307	0.348
0.20	0.80	1.000	1.000	0.093	0.087	0.082	0.798	0.752	0.909
0.50	-0.80	1.000	1.000	0.667	0.673	0.632	0.999	0.998	0.999
0.50	-0.50	1.000	1.000	0.761	0.743	0.728	0.989	0.992	0.988
0.50	-0.20	0.901	0.889	0.781	0.790	0.739	0.903	0.904	0.886
0.50	0.00	0.700	0.664	0.767	0.778	0.746	0.706	0.722	0.650
0.50	0.20	0.934	0.923	0.771	0.795	0.750	0.372	0.365	0.302
0.50	0.50	1.000	1.000	0.683	0.685	0.662	0.044	0.040	0.054
0.50	0.80	1.000	1.000	0.397	0.378	0.402	0.434	0.350	0.590
0.80	-0.80	1.000	1.000	0.994	0.999	0.995	1.000	1.000	1.000
0.80	-0.50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.80	-0.20	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.80	0.00	0.999	0.998	0.999	0.999	0.999	0.997	1.000	0.996
0.80	0.20	1.000	1.000	1.000	0.998	1.000	0.988	0.984	0.977
0.80	0.50	1.000	1.000	0.990	0.986	0.997	0.781	0.779	0.699
0.80	0.80	1.000	1.000	0.847	0.820	0.947	0.033	0.021	0.062

Note: Bold figures refer to the size of the test at nominal size of 5%. All other figures refer to the size adjusted power of the tests.

Table 2: Monte carlo simulations for size and power of LM and LR tests of the random effects, the Anselin and the Kapoor-Kelejian-Prucha models; share of rejections in 2000 replications

(N=50, T=5, $\sigma^2_\mu=5$, $\sigma^2_\nu=15$)

		Random effects model		Anselin model		Kelejian-Prucha model	
		$H_0^A: \rho_1=0, \rho_2=0$		$H_0^B: \rho_1=0$		$H_0^C: \rho_1=\rho_2$	
ρ_1	ρ_2	LM	LR	LM	LR	LM	LR
-0.80	-0.80	1.000	1.000	0.660	0.757	0.039	0.033
-0.80	-0.50	1.000	1.000	0.824	0.896	0.443	0.401
-0.80	-0.20	0.987	0.991	0.935	0.952	0.804	0.812
-0.80	0.00	0.896	0.923	0.950	0.963	0.940	0.953
-0.80	0.20	0.956	0.961	0.935	0.947	0.974	0.981
-0.80	0.50	1.000	1.000	0.875	0.902	0.993	0.999
-0.80	0.80	1.000	1.000	0.804	0.838	0.993	0.999
-0.50	-0.80	1.000	1.000	0.301	0.320	0.093	0.175
-0.50	-0.50	1.000	1.000	0.422	0.431	0.047	0.038
-0.50	-0.20	0.853	0.878	0.496	0.532	0.248	0.262
-0.50	0.00	0.389	0.425	0.489	0.502	0.448	0.484
-0.50	0.20	0.767	0.756	0.504	0.548	0.684	0.743
-0.50	0.50	1.000	1.000	0.378	0.419	0.865	0.920
-0.50	0.80	1.000	1.000	0.306	0.328	0.923	0.989
-0.20	-0.80	1.000	1.000	0.097	0.098	0.316	0.455
-0.20	-0.50	1.000	1.000	0.119	0.112	0.120	0.131
-0.20	-0.20	0.641	0.668	0.108	0.123	0.044	0.042
-0.20	0.00	0.100	0.111	0.126	0.129	0.123	0.125
-0.20	0.20	0.638	0.605	0.129	0.148	0.291	0.324
-0.20	0.50	1.000	1.000	0.084	0.097	0.588	0.674
-0.20	0.80	1.000	1.000	0.066	0.080	0.733	0.909
0.00	-0.80	1.000	1.000	0.049	0.057	0.457	0.659
0.00	-0.50	1.000	1.000	0.046	0.058	0.265	0.304
0.00	-0.20	0.570	0.586	0.050	0.053	0.076	0.071
0.00	0.00	0.050	0.055	0.048	0.052	0.053	0.049
0.00	0.20	0.627	0.596	0.039	0.039	0.096	0.119
0.00	0.50	1.000	1.000	0.050	0.047	0.310	0.413
0.00	0.80	1.000	1.000	0.050	0.045	0.521	0.753
0.20	-0.80	1.000	1.000	0.073	0.069	0.755	0.866
0.20	-0.50	1.000	1.000	0.104	0.081	0.585	0.613
0.20	-0.20	0.552	0.564	0.091	0.083	0.269	0.257
0.20	0.00	0.084	0.070	0.108	0.082	0.107	0.091
0.20	0.20	0.691	0.660	0.109	0.097	0.041	0.045
0.20	0.50	1.000	1.000	0.075	0.068	0.199	0.245
0.20	0.80	1.000	1.000	0.071	0.072	0.435	0.629
0.50	-0.80	1.000	1.000	0.468	0.438	0.971	0.989
0.50	-0.50	1.000	1.000	0.565	0.520	0.929	0.936
0.50	-0.20	0.772	0.765	0.586	0.571	0.790	0.754
0.50	0.00	0.505	0.482	0.579	0.557	0.535	0.492
0.50	0.20	0.886	0.873	0.541	0.524	0.252	0.197
0.50	0.50	1.000	1.000	0.325	0.351	0.039	0.053
0.50	0.80	1.000	1.000	0.182	0.193	0.236	0.322
0.80	-0.80	1.000	1.000	0.984	0.987	1.000	1.000
0.80	-0.50	1.000	1.000	0.993	0.993	1.000	1.000
0.80	-0.20	0.993	0.993	0.992	0.991	0.998	0.997
0.80	0.00	0.988	0.987	0.993	0.993	0.989	0.984
0.80	0.20	0.999	0.999	0.990	0.993	0.959	0.930
0.80	0.50	1.000	1.000	0.846	0.960	0.630	0.525
0.80	0.80	1.000	1.000	0.430	0.644	0.034	0.059

Note: Bold figures refer to the size of the test at nominal size of 5%. All other figures refer to the size adjusted power of the tests.

Table 3: Monte carlo simulations for size and power of LM and LR tests of the random effects, the Anselin and the Kapoor-Kelejian-Prucha models; share of rejections in 2000 replications

($N=50, T=5, \sigma^2_\mu=15, \sigma^2_\nu=5$)

		Random effects model		Anselin model		Kelejian-Prucha model	
		$H_0^A: \rho_1=0, \rho_2=0$		$H_0^B: \rho_1=0$		$H_0^C: \rho_1=\rho_2$	
ρ_1	ρ_2	LM	LR	LM	LR	LM	LR
-0.80	-0.80	1.000	1.000	0.985	0.994	0.039	0.032
-0.80	-0.50	1.000	1.000	0.997	0.999	0.642	0.610
-0.80	-0.20	0.999	1.000	0.998	0.999	0.964	0.965
-0.80	0.00	0.986	0.995	0.997	0.998	0.995	0.996
-0.80	0.20	0.998	1.000	0.996	0.998	1.000	1.000
-0.80	0.50	1.000	1.000	0.993	0.997	1.000	1.000
-0.80	0.80	1.000	1.000	0.969	0.975	1.000	1.000
-0.50	-0.80	1.000	1.000	0.727	0.769	0.271	0.408
-0.50	-0.50	1.000	1.000	0.815	0.836	0.046	0.046
-0.50	-0.20	0.927	0.945	0.814	0.831	0.384	0.370
-0.50	0.00	0.680	0.748	0.810	0.834	0.730	0.748
-0.50	0.20	0.935	0.942	0.811	0.820	0.937	0.952
-0.50	0.50	1.000	1.000	0.755	0.777	0.999	1.000
-0.50	0.80	1.000	1.000	0.589	0.619	1.000	1.000
-0.20	-0.80	1.000	1.000	0.174	0.198	0.788	0.885
-0.20	-0.50	1.000	1.000	0.210	0.235	0.241	0.267
-0.20	-0.20	0.671	0.704	0.231	0.249	0.049	0.051
-0.20	0.00	0.163	0.189	0.236	0.256	0.176	0.192
-0.20	0.20	0.735	0.732	0.230	0.237	0.509	0.555
-0.20	0.50	1.000	1.000	0.178	0.188	0.934	0.965
-0.20	0.80	1.000	1.000	0.136	0.142	1.000	1.000
0.00	-0.80	1.000	1.000	0.042	0.053	0.951	0.978
0.00	-0.50	1.000	1.000	0.035	0.042	0.632	0.652
0.00	-0.20	0.579	0.594	0.039	0.050	0.129	0.117
0.00	0.00	0.040	0.047	0.036	0.045	0.041	0.049
0.00	0.20	0.645	0.625	0.039	0.048	0.193	0.222
0.00	0.50	1.000	1.000	0.048	0.053	0.751	0.804
0.00	0.80	1.000	1.000	0.049	0.053	0.992	0.998
0.20	-0.80	1.000	1.000	0.178	0.153	0.995	0.998
0.20	-0.50	1.000	1.000	0.182	0.170	0.915	0.921
0.20	-0.20	0.644	0.655	0.196	0.166	0.514	0.480
0.20	0.00	0.153	0.136	0.214	0.189	0.176	0.142
0.20	0.20	0.699	0.673	0.206	0.165	0.038	0.045
0.20	0.50	1.000	1.000	0.178	0.148	0.414	0.476
0.20	0.80	1.000	1.000	0.120	0.102	0.969	0.990
0.50	-0.80	1.000	1.000	0.794	0.775	1.000	1.000
0.50	-0.50	1.000	1.000	0.850	0.832	0.997	0.997
0.50	-0.20	0.938	0.937	0.860	0.845	0.950	0.944
0.50	0.00	0.784	0.774	0.866	0.849	0.804	0.773
0.50	0.20	0.955	0.950	0.860	0.839	0.452	0.386
0.50	0.50	1.000	1.000	0.828	0.811	0.040	0.056
0.50	0.80	1.000	1.000	0.635	0.639	0.660	0.786
0.80	-0.80	1.000	1.000	1.000	1.000	1.000	1.000
0.80	-0.50	1.000	1.000	1.000	1.000	1.000	1.000
0.80	-0.20	1.000	1.000	1.000	1.000	1.000	1.000
0.80	0.00	0.999	1.000	1.000	1.000	0.999	0.999
0.80	0.20	1.000	1.000	1.000	1.000	0.991	0.981
0.80	0.50	1.000	1.000	0.999	0.999	0.805	0.728
0.80	0.80	1.000	1.000	0.988	0.994	0.032	0.063

Note: Bold figures refer to the size of the test at nominal size of 5%. All other figures refer to the size adjusted power of the tests.

Table 4: Monte carlo simulations for the robustness of the LM and LR tests of the random effects, the Anselin and the Kapoor-Kelejjan-Prucha models; share of rejections in 2000 replications

(N=50, T=5, $\sigma^2_\mu=10$, $\sigma^2_\nu=10$)

			$v_{it} \sim t(5)$		$v_{it} \sim \text{lognormal}(0,10)$	
	ρ_1	ρ_2	LM	LR	LM	LR
Random effects model, $H_0^A: \rho_1=0, \rho_2=0$	0.00	0.00	0.042	0.053	0.041	0.047
Anselin model, $H_0^B: \rho_1=0$	0.00	-0.80	0.055	0.066	0.045	0.055
	0.00	-0.50	0.052	0.065	0.042	0.049
	0.00	-0.20	0.045	0.053	0.043	0.047
	0.00	0.00	0.045	0.055	0.032	0.038
	0.00	0.20	0.047	0.055	0.038	0.043
	0.00	0.50	0.045	0.047	0.048	0.050
	0.00	0.80	0.050	0.049	0.039	0.040
Kapoor-Kelejjan-Prucha model, $H_0^C: \rho_1=\rho_2$	-0.80	-0.80	0.036	0.035	0.039	0.040
	-0.50	-0.50	0.049	0.046	0.048	0.048
	-0.20	-0.20	0.048	0.044	0.045	0.048
	0.00	0.00	0.043	0.048	0.035	0.039
	0.20	0.20	0.045	0.051	0.035	0.047
	0.50	0.50	0.038	0.054	0.034	0.051
	0.80	0.80	0.029	0.054	0.029	0.059

Figure 1: The power of the LM test, random effects model

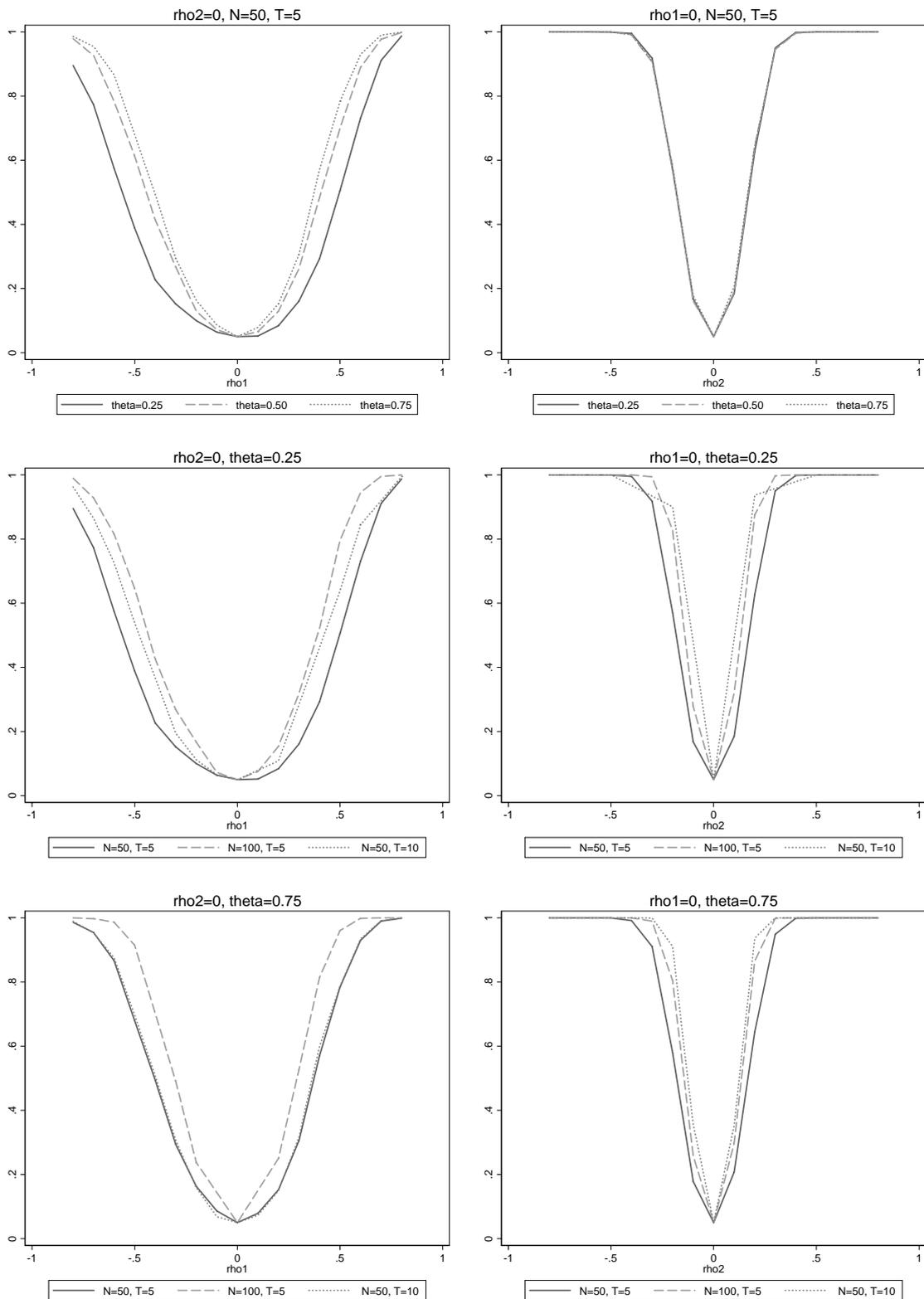


Figure 2: The power of the LM test, Anselin model

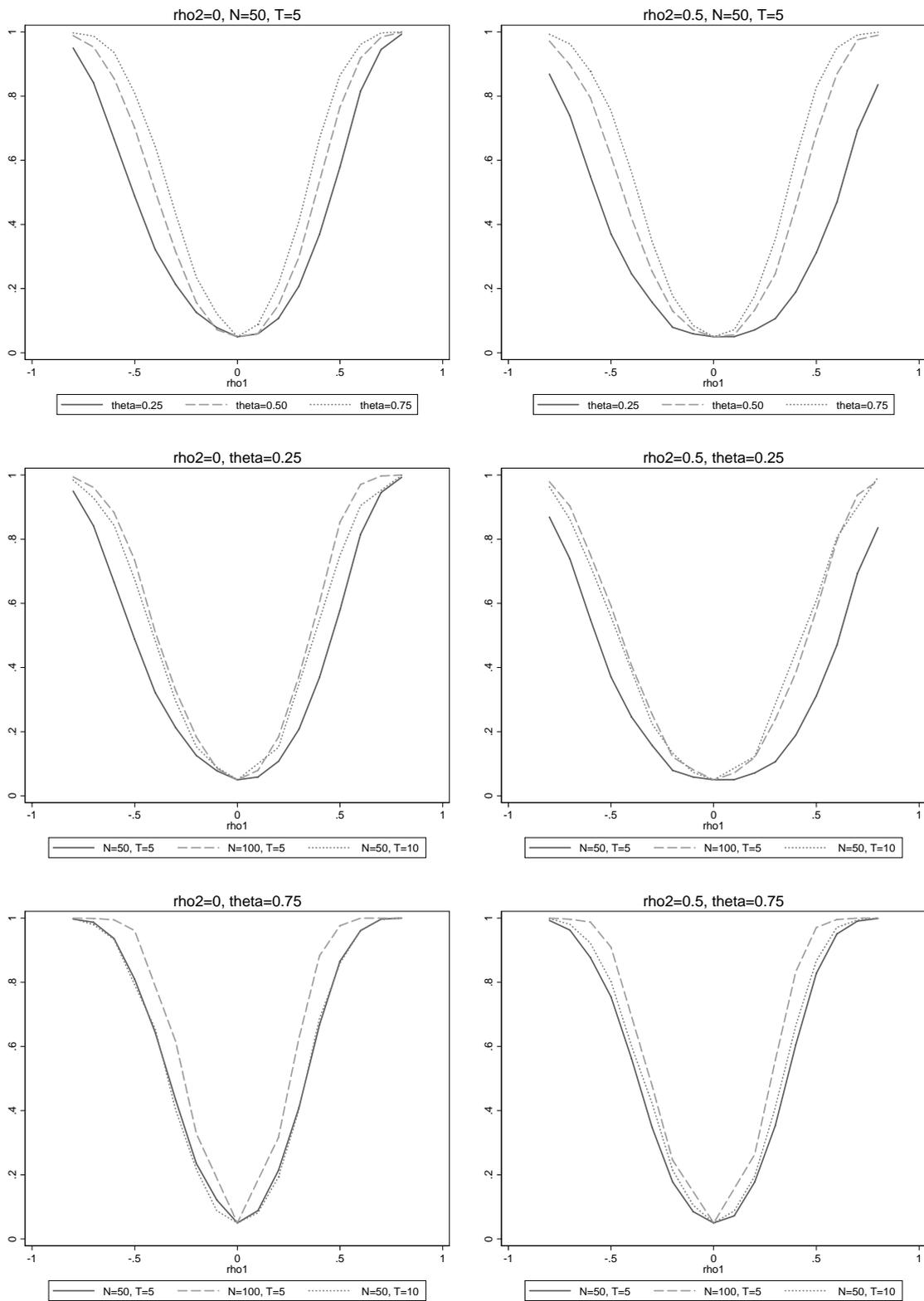


Figure 3: The power of the LM test, KKP model - part I

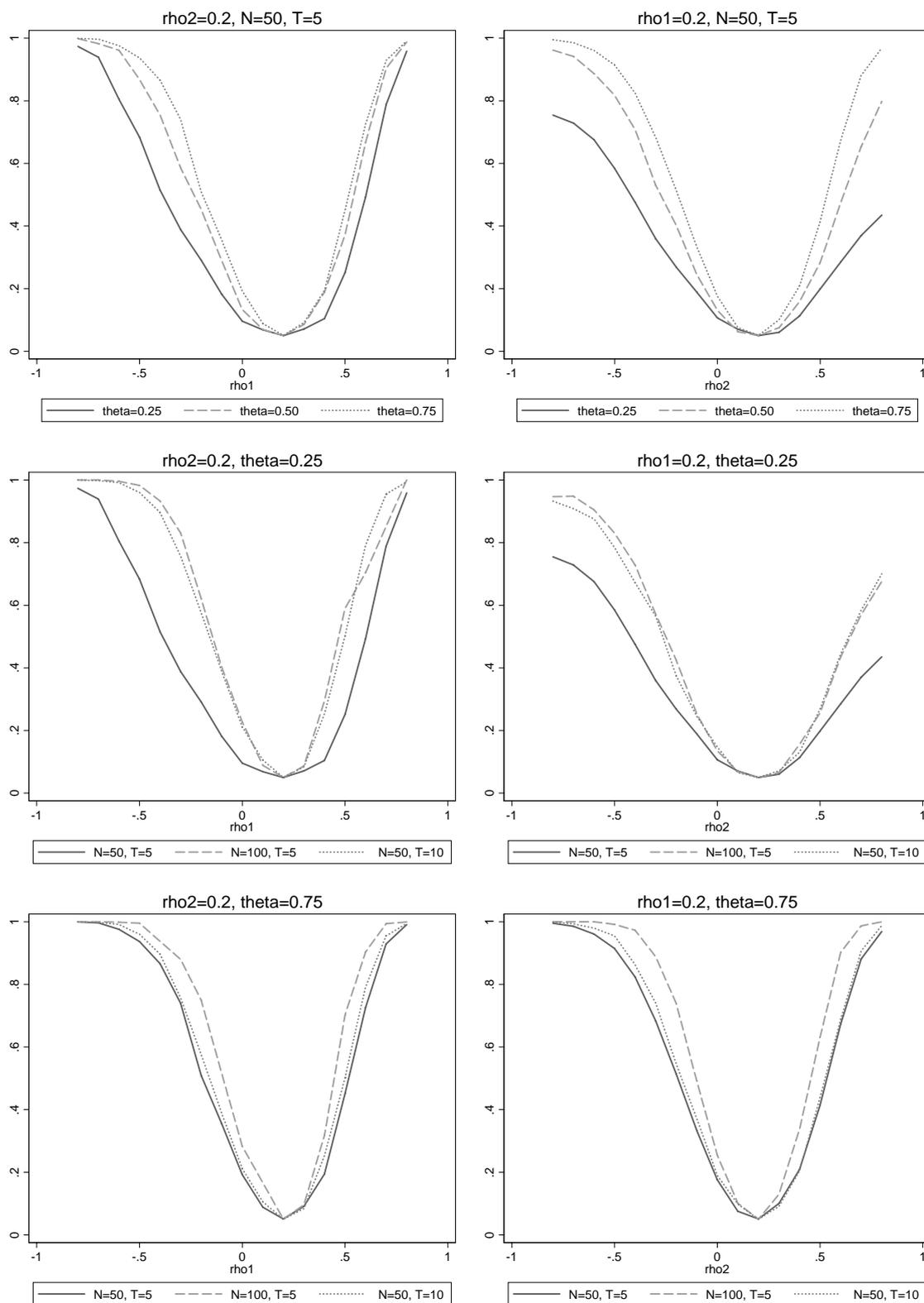


Figure 4: The power of the LM test, KKP model - part II

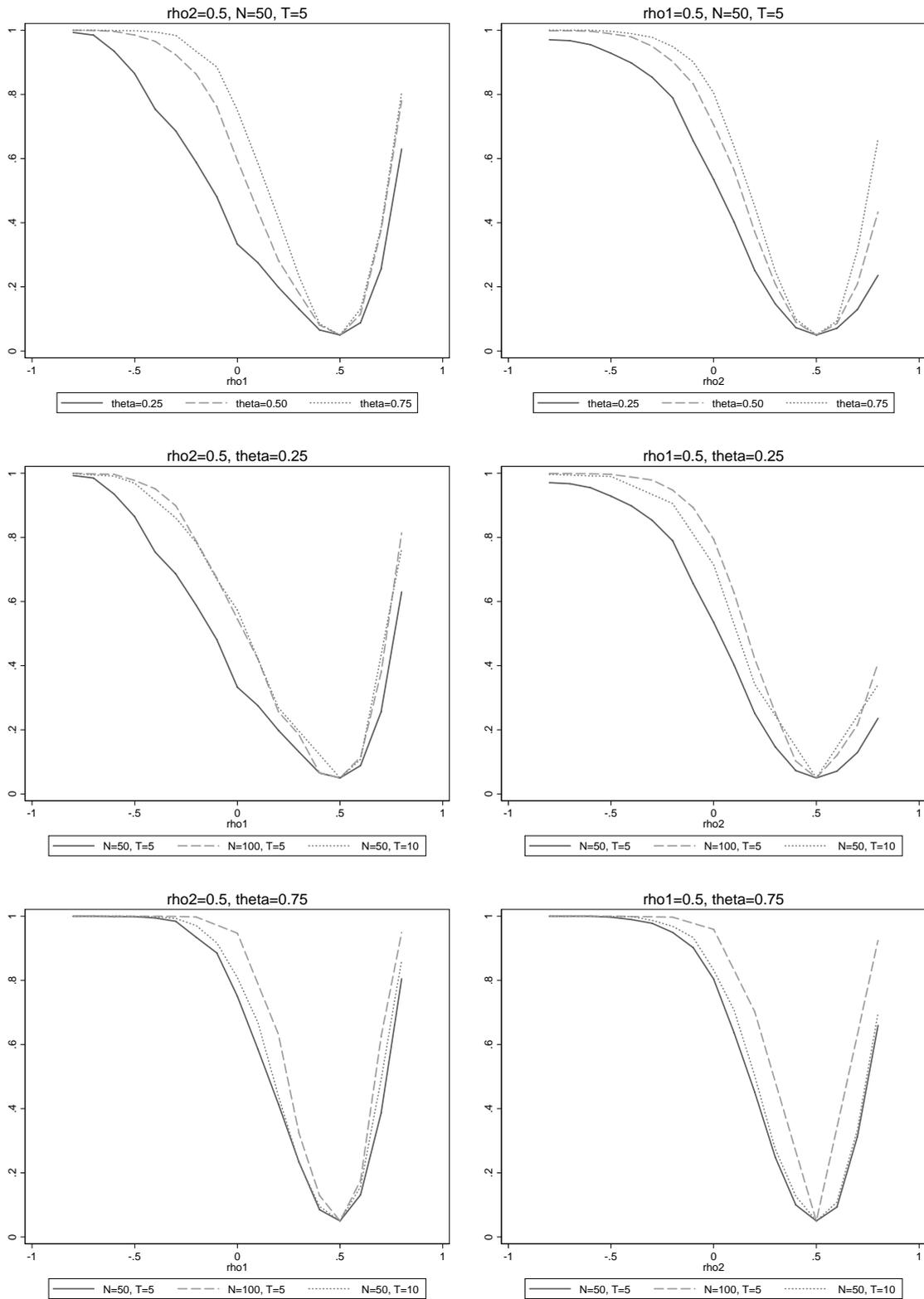


Figure 5: The power of the LM test, 3 vs 5 neighbors

