CESifo Working Papers
www.cesifo.org/wp

A Generalized Spatial Panel Data Model with Random Effects

Badi H. Baltagi
Peter Egger
Michael Pfaffermayr

CESIFO WORKING PAPER NO. 3930
CATEGORY 12: EMPIRICAL AND THEORETICAL METHODS
SEPTEMBER 2012

CESifo *Working Paper No. 3930*

A Generalized Spatial Panel Data Model with Random Effects

Abstract

This paper proposes a generalized 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, and Prucha model, and (iii) the simple random effects model that ignores the spatial correlation in the residuals. For two of these three tests, we obtain closed form solutions and we derive their large sample distributions. 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-Code: C230, C120.

Keywords: panel data, spatially autocorrelated residuals, maximum-likelihood estimation, Lagrange multiplier, likelihood ratio.

Badi H. Baltagi
Department of Economics and
Center for Policy Research
Syracuse University
USA – Syracuse, NY 13244-1020
bbaltagi@maxwell.syr.edu

Peter Egger
Department of Management, Technology, and Economics
ETH Zurich / WEH E6
Weinbergstrasse 35
Switzerland – 8092 Zurich
egger@kof.ethz.ch

Michael Pfaffermayr
Department of Economics
University of Innsbruck
Universitaetsstrasse 15
Austria – 6020 Innsbruck
Michael.Pfaffermayr@uibk.ac.at

January 27, 2012
Michael Pfaffermayr gratefully acknowledges financial support from the Austrian Science Foundation grant 17028.

1 Introduction[1]

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, 2008; 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).[2]

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

[1] We would like to thank Matthias Koch, Ingmar Prucha, two anonymous referees and the editor Esfandiar Maasoumi for their helpful comments and suggestions. Prelimimary 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.

[2] There has been a lot of attention to cross-sectional dependence in panel data models, modeled through factor models. A rapidly growing reasearch topic within this general field has been the reconciliation of factor models and spatial models, with attempts to express weak and strong cross-sectional dependence, see Chudik, Pesaran, and Tosetti (2011), Pesaran and Tosetti (2011), and Sarafidis and Wansbeek (2011), to mention a few.

inherently time-varying, the KKP process assumes the spillovers to be time-invariant as well as time-variant. For example, 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 consider a (quasi-)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.

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), Anselin, Le Gallo, and Jayet (2008), Lee and Yu (2010a, 2010b) to mention a few. A generalized spatial panel data model that encompasses these previous specifications is given as follows:[3]

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

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 and independent of N. In case $\mathbf{W}$ is row normalized, the parameter space for ρ_r is a closed interval contained in $(-1, 1)$. For the case where $\mathbf{W}$ is not normalized, we assume that the parameter space for ρ_r is contained in the closed interval $-1/\lambda_{\max} < \rho_r < 1/\lambda_{\max}$ for all N and $r = 1, 2$, where $\lambda_{\max}$ is the largest absolute value of the eigenvalues of $\mathbf{W}$. Hence, the spatial weights matrix may be either row normalized or maximum row normalized (see Kelejian and Prucha, 2010). Further, let $\mathbf{A} = \mathbf{I}_N - \rho_1 \mathbf{W}$ and $\mathbf{B} = \mathbf{I}_N - \rho_2 \mathbf{W}$. The matrices $\mathbf{A}$ and $\mathbf{B}$ are non-singular for all ρ_r, $r = 1$, 2 in the parameter space and all N.

The elements of $\boldsymbol{\mu}$ are assumed to be independently and identically distrib-

[3] To 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.

uted as $N(0,\sigma_\mu^2)$ across i. The elements of $\boldsymbol{\nu}_t$ are assumed to be independently and identically distributed as $N(0,\sigma_\nu^2)$ across i and t. 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

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

where $\mathbf{y} = [\mathbf{y}_1', ..., \mathbf{y}_T']'$, $\mathbf{X} = [\mathbf{X}_1', ..., \mathbf{X}_T']'$, 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 $\mathbf{Z}_\mu = \boldsymbol{\iota}_T \otimes \mathbf{I}_N$. $\boldsymbol{\iota}_T$ 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 (2008).

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

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

Let $\mathbf{E}_T = \mathbf{I}_T - \overline{\mathbf{J}}_T$, where $\overline{\mathbf{J}}_T = \mathbf{J}_T/T$ is the averaging matrix with $\mathbf{J}_T$ being a matrix of ones of dimension T. The variance-covariance matrix of the spatial random effects panel data model is given by

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

where $\boldsymbol{\Sigma}_u$ is defined as $\boldsymbol{\Sigma}_u = (\bar{\mathbf{J}}_T \otimes (\frac{T\sigma_\mu^2}{\sigma_\nu^2}(\mathbf{A}'\mathbf{A})^{-1} + (\mathbf{B}'\mathbf{B})^{-1})) + (\mathbf{E}_T \otimes (\mathbf{B}'\mathbf{B})^{-1})$. 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. The last equality in (3) replaces $\mathbf{J}_T$ by $T\bar{\mathbf{J}}_T$ and $\mathbf{I}_T$ by $\mathbf{E}_T + \bar{\mathbf{J}}_T$. Note that $\mathbf{Z}_\mu \mathbf{Z}_\mu' = \mathbf{J}_T \otimes \mathbf{I}_N$. 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

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

Also, $\det[\boldsymbol{\Omega}_u] = \det[T\sigma_\mu^2(\mathbf{A}'\mathbf{A})^{-1} + \sigma_\nu^2(\mathbf{B}'\mathbf{B})^{-1}]\det[\sigma_\nu^2(\mathbf{B}'\mathbf{B})^{-1}]^{T-1}$. We also assume that the inverses $\mathbf{A}^{-1}$, $\mathbf{B}^{-1}$ and $[T\sigma_\mu^2(\mathbf{A}'\mathbf{A})^{-1} + \sigma_\nu^2(\mathbf{B}'\mathbf{B})^{-1}]^{-1}$ have bounded row and column sums, uniformly in N and in the parameter space (see Assumption A2 in the Appendix for further details). Under the present assumptions, the log-likelihood function of the general model is given by

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

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.[4]

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 (2008).

(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).

[4] The 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.

(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.[5] 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 (quasi-)ML estimates of the general model; while Appendices C and E provide the derivations of the large sample distributions of the LM tests for H_0^A and H_0^C. Appendix D gives details on the the LM test for H_0^B.

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

The (quasi-)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

$$\widetilde{LM}_A = \tfrac{1}{2b_A\widetilde{\sigma}_1^4}\widetilde{G}_A^2 + \tfrac{1}{2b_A(T-1)\widetilde{\sigma}_\nu^4}\widetilde{M}_A^2, \tag{5}$$

where $\widetilde{\sigma}_1^2 = T\widetilde{\sigma}_\mu^2 + \widetilde{\sigma}_\nu^2$, $b_A = tr[(\mathbf{W}' + \mathbf{W})^2]$, $\widetilde{G}_A = \widetilde{\mathbf{u}}'[\bar{\mathbf{J}}_T \otimes (\mathbf{W}' + \mathbf{W})]\widetilde{\mathbf{u}}$, and $\widetilde{M}_A = \widetilde{\mathbf{u}}'[\mathbf{E}_T \otimes (\mathbf{W}' + \mathbf{W})]\widetilde{\mathbf{u}}$. In this case, $\widetilde{\mathbf{u}} = \mathbf{y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}$ 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, $\widetilde{\sigma}_\nu^2 = \frac{\widetilde{\mathbf{u}}'(\mathbf{E}_T\otimes\mathbf{I}_N)\widetilde{\mathbf{u}}}{N(T-1)}$ and $\widetilde{\sigma}_1^2 = \frac{\widetilde{\mathbf{u}}'(\bar{\mathbf{J}}_T\otimes\mathbf{I}_N)\widetilde{\mathbf{u}}}{N}$. Under H_0^A, the $\widetilde{LM}_A$ statistic is asymptotically distributed as χ_2^2 as shown in Appendix C. Note this test does not require the assumption of normally distributed disturbances.

Under normal disturbances one can also derive the corresponding LR test for H_0^A: $\rho_1 = \rho_2 = 0$ as

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

[5] LM 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).

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

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

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., $\rho_2 \neq 0$. In fact, under H_0^B, the information matrix is block-diagonal with the lower block being independent of $\boldsymbol{\beta}$. Let $\mathbf{d}_\theta$ 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 $\mathbf{J}_\theta$. The (quasi-)ML estimates under H_0^B are labeled by a hat. The corresponding estimated residuals are then $\widehat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}$, where $\widehat{\boldsymbol{\beta}}$ is the (quasi-)ML-estimator of $\boldsymbol{\beta}$ under H_0^B. 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

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

where $\widehat{\mathbf{C}}_1 = [T\widehat{\sigma}_\mu^2\mathbf{I}_N + \widehat{\sigma}_\nu^2(\widehat{\mathbf{B}}'\widehat{\mathbf{B}})^{-1}]^{-1}$ and $\mathbf{C}_2 = (\mathbf{W}' + \mathbf{W})$, $\widehat{\mathbf{G}}_B = (\bar{\mathbf{J}}_T \otimes \widehat{\mathbf{C}}_1\mathbf{C}_2\widehat{\mathbf{C}}_1)$, and $\widehat{g}_B = tr[\widehat{\mathbf{C}}_1\mathbf{C}_2]$. Under normal disturbances an estimate of the lower (4×4) block of the information matrix $\widehat{\mathbf{J}}_\theta$ under H_0^B is given by[6]

$$\left.\widehat{\mathbf{J}}_\theta\right|_{H_0^B} =$$

$$\begin{bmatrix} \frac{1}{2}tr[\widehat{\mathbf{C}}_3{}^2] + \frac{N(T-1)}{2\widehat{\sigma}_\nu^4} & \frac{T}{2}tr\left[\widehat{\mathbf{C}}_3\widehat{\mathbf{C}}_1\right] & \frac{T\widehat{\sigma}_\mu^2}{2}tr[\widehat{\mathbf{C}}_3\widehat{\mathbf{C}}_1\mathbf{C}_2] & \frac{\widehat{\sigma}_\nu^2}{2}tr[\widehat{\mathbf{C}}_3\widehat{\mathbf{C}}_1\widehat{\mathbf{C}}_5] + \frac{(T-1)}{2\widehat{\sigma}_\nu^2}tr[\widehat{\mathbf{C}}_4] \\ \frac{T}{2}tr\left[\widehat{\mathbf{C}}_3\widehat{\mathbf{C}}_1\right] & \frac{T^2}{2}tr\left[\widehat{\mathbf{C}}_1^2\right] & \frac{T^2\widehat{\sigma}_\mu^2}{2}tr[\widehat{\mathbf{C}}_1^2\mathbf{C}_2] & \frac{T\widehat{\sigma}_\nu^2}{2}tr[\widehat{\mathbf{C}}_1^2\widehat{\mathbf{C}}_5] \\ \frac{T\widehat{\sigma}_\mu^2}{2}tr[\widehat{\mathbf{C}}_3\widehat{\mathbf{C}}_1\mathbf{C}_2] & \frac{T^2\widehat{\sigma}_\mu^2}{2}tr[\widehat{\mathbf{C}}_1^2\mathbf{C}_2] & \frac{T^2\widehat{\sigma}_\mu^4}{2}tr[(\widehat{\mathbf{C}}_1\mathbf{C}_2)^2] & \frac{T\widehat{\sigma}_\mu^2\widehat{\sigma}_\nu^2}{2}tr[\widehat{\mathbf{C}}_1\mathbf{C}_2\widehat{\mathbf{C}}_1\widehat{\mathbf{C}}_5] \\ \frac{\widehat{\sigma}_\nu^2}{2}tr[\widehat{\mathbf{C}}_3\widehat{\mathbf{C}}_1\widehat{\mathbf{C}}_5] + \frac{(T-1)}{2\widehat{\sigma}_\nu^2}tr[\widehat{\mathbf{C}}_4] & \frac{T\widehat{\sigma}_\nu^2}{2}tr[\widehat{\mathbf{C}}_1^2\widehat{\mathbf{C}}_5] & \frac{T\widehat{\sigma}_\mu^2\widehat{\sigma}_\nu^2}{2}tr[\widehat{\mathbf{C}}_1\mathbf{C}_2\widehat{\mathbf{C}}_1\widehat{\mathbf{C}}_5] & \frac{\widehat{\sigma}_\nu^4}{2}tr[(\widehat{\mathbf{C}}_1\widehat{\mathbf{C}}_5)^2] + \frac{(T-1)}{2}tr[\widehat{\mathbf{C}}_4^2] \end{bmatrix},$$

[6] Detailed derivations are available form the authors upon request.

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 is calculated as

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

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 has no closed form representation, but using similar assumptions and proofs as in the Appendices, this test statistic should be asymptotically distributed as χ_1^2. Under non-normal disturbances the LM-test can be derived following White (1982) and in the Appendix D it is derived as

$$LM_{B,robust} = \widehat{\mathbf{d}}_\theta'\widehat{\mathbf{J}}_\theta^{-1}\mathbf{R}'\left(\mathbf{R}\left(\widehat{\mathbf{J}}_\theta^{-1} + \widehat{\mathbf{J}}_\theta^{-1}\widehat{\Sigma}_\theta\widehat{\mathbf{J}}_\theta^{-1}\right)\mathbf{R}'\right)^{-1}\mathbf{R}\widehat{\mathbf{J}}_\theta^{-1}\widehat{\mathbf{d}}_\theta', \tag{7}$$

where we define the 4×4 matrix $\mathbf{\Sigma}_\theta$ with klth element $[\frac{1}{2}\sum_{i=1}^{NT} a_{k,ii}a_{l,ii}(\mu_\eta^{(4)} - 3)]$ and $\mathbf{R} = [0,0,1,0]$. The elements $a_{k,ii}$ are defined in the Appendix D, while $\mu_\eta^{(4)} = E[\left(\mathbf{S}^{-1}\mathbf{u}\right)^4]$ with $\mathbf{\Omega}_u = \mathbf{SS}'$. This robust LM-test statistic is asymptotically distributed as $\chi^2(1)$.

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

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

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 $\overline{\mathbf{u}} = \mathbf{y} - \mathbf{X}\overline{\boldsymbol{\beta}}$. The score and the information matrix needed for this test are derived in Appendix E. With normal disturbances the joint LM test statistic for H_0^C is given by

$$LM_C = \tfrac{T}{2\bar{b}_C(T-1)\overline{\sigma}_1^4}\left(\overline{G}_{Cb} - \overline{\sigma}_1^2 tr[\overline{\mathbf{D}}]\right)^2, \tag{8}$$

with $\overline{G}_C = \overline{\mathbf{u}}'(\overline{\mathbf{J}}_T \otimes \overline{\mathbf{F}})\overline{\mathbf{u}}$, $\overline{\mathbf{F}} = \mathbf{W}'\overline{\mathbf{A}} + \overline{\mathbf{A}}'\mathbf{W}$ and $\overline{\mathbf{D}} = \overline{\mathbf{F}}(\overline{\mathbf{A}}'\overline{\mathbf{A}})^{-1}$. Also, $\bar{b}_C = tr[\overline{\mathbf{D}}^2] - (tr[\overline{\mathbf{D}}])^2/N$, $\overline{\sigma}_1^2 = \frac{\overline{\mathbf{u}}'[\overline{\mathbf{J}}_T \otimes (\overline{\mathbf{A}}'\overline{\mathbf{A}})]\overline{\mathbf{u}}}{N}$ and $\overline{\sigma}_\nu^2 = \frac{\overline{\mathbf{u}}'[\mathbf{E}_T \otimes (\overline{\mathbf{A}}'\overline{\mathbf{A}})]\overline{\mathbf{u}}}{N(T-1)}$. Under H_0^C, the LM_C statistic is asymptotically distributed as χ_1^2 as shown in Appendix F. If the disturbances are not normally distributed one may use the robust version of this LM test, which is derived in Appendix E as

$$\overline{LM}_{C,robust} = \overline{LM}_C \frac{1}{1 + \left(\overline{d}_b + \overline{d}_w'\right)\left(\frac{T-1}{T}\right)}. \tag{9}$$

The true correction factors are defined as $d_b = \frac{\sum_{i=1}^N l_{ii}^2 \left(\sigma_\mu^4 T^2 \left(\mu_\mu^{(4)} - 3\right) + \frac{1}{T}\sigma_\nu^4 \left(\mu_\nu^{(4)} - 3\right)\right)}{2\sigma_1^4 tr[\mathbf{D}^2]}$ and $d_w' = \frac{\sum_{i=1}^N l_{ii}^2 \frac{1}{T}\left(\mu_\nu^{(4)} - 3\right)}{(T-1)^2 2tr[\mathbf{D}^2]}$, respectively (see Appendix E for details).

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

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

Kapoor, Kelejian, and Prucha (2007) consider a generalized method of moments estimator, rather than (quasi-)MLE, for their spatial random effects panel data model. 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 ($K = 2$)

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

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 processes ζ_i and z_{it} are assumed to be independent and held fixed in repeated samples. We conduct an extensive analysis for the case of normally distributed disturbances as summarized in Tables 1-3 and dispense with the assumption of normality in Table 4. In the former case, 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, where each observation (except that in the first and last row and column) has four neighbors situated in the north, south, east and west neighboring cells. 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.[7]

===== 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

[7] In 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.

KKP model, for high absolute values of ρ_1 and ρ_2, both equal to 0.8. The size adjusted power[8] 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 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.

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

[8] The 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.

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.[9] 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 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

[9] We 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.

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 performance of the proposed LM tests with respect to (i) non-normal errors (using the derived robust vs. the non-robust LM test statistics) and (ii) the specification of the spatial weighting matrix. To compare the simulated power functions for normal vs. non-normal errors, we generate the remainder error term first as $\nu_{it} \sim t(5)$ and normalize 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 analyze a skewed error distribution assuming that 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 on average there are relatively small effects on the size of the (non-robust) LM tests under either error distribution in comparison to the tests under normality.

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

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 (non-robust) 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 (non-robust) LM tests considered are fairly robust to deviations from the assumption of a normally distributed error term.

Interestingly, with small samples as the ones considered and a relatively small signal-to-noise ratio as assumed here, there is no gain from using robust LM test statistics rather than non-robust ones. In many cells of Table 4, the robust test size is more off the nominal size than this is the case for the non-robust test size. The reason for this result is the following. The correction factors of the LM statistics deflate the non-robust test statistics. Hence, with oversized LM tests, the corresponding correction factors would adjust the test size towards the nominal size (see Yang, 2010, for an example with cross-section data). In our case, there is no systematic over-rejection in the samples considered so that the correction factors lead to even more pronouncedly undersized tests. In broader terms, problems with such correction factors in small samples also accrue to the use of higher moments of the disturbances which can not be estimated without bias in small samples (see Teuscher, Herrendörfer, and Guiard, 1994).[10]

Furthermore, we repeated the LM and LR tests for the same model configuration as in Table 1 for an alternative model which assumes the vector of explanatory variables, $\mathbf{x}$, to be generated as a spatial moving average of the form

$$\mathbf{x} = [\mathbf{I}_T \otimes (\mathbf{I}_N + 0.5\mathbf{W}_N)]\mathbf{x}_{old}$$

where $\mathbf{x}_{old}$ is the specification of $\mathbf{x}$ as defined above. Our original conclusions are not sensitive to this alternative specification of $\mathbf{x}$. We also investigated 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.[11]

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

[10] With robust LM tests, we estimate the kurtosis from the realized (true) disturbances for every draw. In applications, one would have to rely on the estimated kurtosis which can be biased substantially in small samples.

[11] All results on the mentioned sensitivity checks are available from the authors upon request. They are suppressed here for the sake of brevity.

Jayet (2008) 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. In addition, we derive robust LM tests that do not rely on the assumption of normally distributed disturbances.

Our Monte Carlo study assesses the small sample performance of the derived tests. We find that under normal disturbances the LM tests are properly sized and powerful even in relatively small samples. Interestingly, with small samples and a relatively small signal-to-noise ratio as considered in the Monte Carlo study, there is no gain from using robust LM test statistics rather than non-robust ones. The LM tests are easy to calculate and their power is reasonably high for all three tests considered. Under normal disturbances 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., Magnus, J.R. (2005). *Matrix Algebra*. Cambridge: Cambridge University Press.

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

Anselin, L. (2001). Rao's score tests in spatial econometrics. *Journal of Statistical Planning and Inference* 97(1):113-139.

Anselin, L., Bera, A.K. (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. In: Ullah, A., Giles, H., eds. *Handbook of Applied Economic Statistics*. New York: Marcel Dekker, pp. 237-290.

Anselin, L., Le Gallo, J., Jayet, H. (2008). Spatial panel econometrics. In: Mátyás, L., Sevestre, P., eds. *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*. Springer: Berlin, pp. 625-660.

Baltagi, B.H. (2008). *Econometric Analysis of Panel Data*. Chichester: Wiley.

Baltagi, B.H., Song, S.H., Koh, W. (2003). Testing panel data models with spatial error correlation. *Journal of Econometrics* 117(1):123-150.

Chudik, A., Pesaran M. H., Tosetti, E. (2011). Weak and strong cross section dependence and estimation of large panels. *The Econometrics Journal* 14(1):C45-C90.

Hartley, H.O., Rao, J.N.K. (1967). Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika* 54(1-2): 93-108.

Harville, D.A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* 72(358):320-338.

Hemmerle, W.J., Hartley, H.O. (1973). Computing maximum likelihood estimates for the mixed A.O.V. model using the W-transformation. *Technometrics* 15(4):819-831.

Horn, R., Johnson, C. (1985). *Matrix Analysis*. Cambridge: Cambridge University Press.

Kapoor, M., Kelejian, H.H., Prucha, I.R. (2007). Panel data models with spatially correlated error components. *Journal of Econometrics* 140(1):97-130.

Kelejian, H.H., Prucha, I.R. (2001). On the asymptotic distribution of the Moran I test with applications. *Journal of Econometrics* 104(2):219-257.

Kelejian, H.H. and I.R. Prucha (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics* 157(1):53-67.

Lee, L.-F. (2004a). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica* 72(6):1899-1926.

Lee, L.-F. (2004b). Supplement to: Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. Ohio State University.

Lee, L.-F., Yu, J. (2010a). Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics* 154(2):165-185.

Lee, L.-F., Yu, J. (2010b). Spatial panels: Random components vs. fixed effects. unpublished manuscript.

Pesaran, M. H., Tosetti, E. (2011). Large panels with common factors and spatial correlation. *Journal of Econometrics* 161(2):182-202.

Rao, R.C. (1973). *Linear Statistical Inference and its Applications*. New Jersey: John Wiley & Sons.

Sarafidis, V., Wansbeek, T. (2011). Cross-sectional dependence in panel data analysis. *Econometric Reviews*, forthcoming.

Teuscher, F., Herrendörfer, G., Guiard, V. (1994). The estimation of skewness and kurtosis of random effects in the linear model. *Biometrical Journal* 36(6):661-672.

Yang, Z. (2010). A robust LM test for spatial error components. *Regional Science and Urban Economics* 40(5):299-310.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* 50(1):1-25.

Appendix A: Score and Information Matrix

Below we make use of the following derivatives to obtain the score and the relevant part of the information matrix:[12]

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

Appendix B: Identification and Consistency

In the sequel, we use subscript 0 to indicate true parameter values where necessary. First, we state the full set of Assumptions.

Assumptions[13]

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 *i.i.d.* $(0, \sigma_\mu^2)$ with $0 < \sigma_\mu^2 < b_\mu < \infty$. $\boldsymbol{\nu}$ is the vector of remainder errors and its elements are *i.i.d.* $(0, \sigma_\nu^2)$ with $0 < \sigma_\nu^2 < b_\nu < \infty$. The elements of $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ are independent of each other. Furthermore $E[|\mu_i|^{4+\eta_\mu}] < \infty$ and for some $\eta_\mu > 0$, and $E[|\nu_{it}|^{4+\eta_\nu}] < \infty$ and for some $\eta_\nu > 0$.

A2 (**spatial correlation**):

(i) Both $\mathbf{u}_1$ and $\mathbf{u}_{2t}$ are spatially correlated with the same $(N \times N)$ nonstochastic spatial weighting matrix $\mathbf{W}$ whose elements may depend on N.

[12] Hartley and Rao (1971) and Hemmerle and Hartley (1973) give a general useful formula that helps in obtaining the score of $\boldsymbol{\theta} = (\sigma_\nu^2, \sigma_\mu^2, \rho_1, \rho_2)'$: $\frac{\partial L}{\partial \theta_r} = -\frac{1}{2} tr \left(\boldsymbol{\Omega}_u^{-1} \frac{\partial \boldsymbol{\Omega}_u}{\partial \theta_r} \right) + \frac{1}{2} \mathbf{u}' \left(\boldsymbol{\Omega}_u^{-1} \frac{\partial \boldsymbol{\Omega}_u}{\partial \theta_r} \boldsymbol{\Omega}_u^{-1} \right) \mathbf{u}, \quad r = 1, ..., 4$. To derive the relevant part of the information matrix, we use the general differentiation result given in Harville (1977): $J_{rs} = E\left[-\frac{\partial^2 L}{\partial \theta_r \theta_s} \right] = \frac{1}{2} tr \left[\boldsymbol{\Omega}_u^{-1} \frac{\partial \boldsymbol{\Omega}_u}{\partial \theta_r} \boldsymbol{\Omega}_u^{-1} \frac{\partial \boldsymbol{\Omega}_u}{\partial \theta_s} \right]$.

[13] To 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. For a similar set of assumptions and a discussion of them see Lee (2004a) and Lee and Yu (2010a and 2010b).

The elements of $\mathbf{W}$ are non-negative and $w_{ii} = 0$.

(ii) The row and column sums of $\mathbf{W}$ are uniformly bounded in absolute value.

(iii) The parameter space for ρ_r is a closed interval contained in $-1/\lambda_{\max} < \rho_r < 1/\lambda_{\max}$ for all N and $r = 1, 2$, where $\lambda_{\max}$ is the largest absolute eigenvalue of $\mathbf{W}$. $\lambda_{\max}$ is assumed to be bounded away from zero by some fixed positive constant.

(iv) Let $\mathbf{A} = \mathbf{I}_N - \rho_1 \mathbf{W}$ and $\mathbf{B} = \mathbf{I}_N - \rho_2 \mathbf{W}$. The non-stochastic matrices $\mathbf{A}$, $\mathbf{B}$ are non-singular for all ρ_r in the parameter space and have bounded row and column sums, uniformly in N. Also, its inverses have bounded row and column sums, uniformly in N and uniformly in the parameter space of ρ_1 and ρ_2 .

(v) The inverse $\boldsymbol{\Sigma}_u^{-1}\left(\sigma_\mu^2, \sigma_\nu^2, \rho_1, \rho_2\right) = (\overline{\mathbf{J}}_T \otimes (\frac{T\sigma_\mu^2}{\sigma_\nu^2}(\mathbf{A}'\mathbf{A})^{-1} + (\mathbf{B}'\mathbf{B})^{-1})^{-1}) + (\mathbf{E}_T \otimes \mathbf{B}'\mathbf{B})$ has bounded row and column sum uniformly in N and uniformly in the parameter space of $\left(\sigma_\mu^2, \sigma_\nu^2, \rho_1, \rho_2\right)$.[14]

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

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

A4 (**identification of** $\boldsymbol{\vartheta}$): For every $\boldsymbol{\vartheta} \in \boldsymbol{\Xi}$, $\boldsymbol{\vartheta} \neq \boldsymbol{\vartheta}_0$, and any $\varepsilon > 0$:
$\limsup_{N\to\infty} \max_{\boldsymbol{\vartheta} \in \overline{\mathbf{N}}_\varepsilon(\boldsymbol{\vartheta}_0)} (-\frac{1}{2} \ln(\frac{1}{NT} tr[\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta}_0) \boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}]) - \frac{1}{2} \frac{1}{NT} \ln[\det \boldsymbol{\Sigma}_u(\boldsymbol{\vartheta}) / \det \boldsymbol{\Sigma}_u(\boldsymbol{\vartheta}_0)]) < 0$, where $\overline{\mathbf{N}}_\varepsilon(\boldsymbol{\vartheta}_0)$ is the complement of an open neighborhood of $\boldsymbol{\vartheta}_0$ of diameter ε.

[14] Under H_0^C we have $\boldsymbol{\Sigma}_u^{-1}\left(\sigma_\mu^2, \sigma_\nu^2, \rho_1, \rho_2\right) = (\overline{\mathbf{J}}_T \otimes (\frac{T\sigma_\mu^2}{\sigma_\nu^2}(\mathbf{A}'\mathbf{A})^{-1} + (\mathbf{A}'\mathbf{A})^{-1})^{-1}) + (\mathbf{E}_T \otimes \mathbf{A}'\mathbf{A}) = (\overline{\mathbf{J}}_T \otimes (\frac{T\sigma_\mu^2 + \sigma_\nu}{\sigma_\nu^2}\mathbf{I}_N) + (\mathbf{E}_T \otimes \mathbf{I}_N)\, (\mathbf{I}_T \otimes \mathbf{A}'\mathbf{A})$. Hence, in this case a sufficient condition for Assumption A2 (v) is A2 (iv). Note Lemma 1 shows that this inverse exists for all $\left(\sigma_\mu^2, \sigma_\nu^2, \rho_1, \rho_2\right)$ in the parameter space.

A5 (identification of $\boldsymbol{\beta}$ under H_0^C): The non-random matrix $\mathbf{X}$ has full column rank $K < N$ and its elements are uniformly bounded by some finite constant. Further, let $\mathbf{Q}_0 = \mathbf{E}_T \otimes \mathbf{I}_N$ and $\mathbf{Q}_1 = \overline{\mathbf{J}}_T \otimes \mathbf{I}_N$ and define $\mathbf{X}^*(\rho) = \mathbf{I}_T \otimes \mathbf{A}$. The non-random matrices $\lim_{N\to\infty}(\frac{1}{NT}\mathbf{X}^*(\rho)\mathbf{Q}_i\mathbf{X}^*(\rho))$, $i = 0, 1$ are finite. The nonrandom matrices $\lim_{N\to\infty}(\frac{1}{NT}\mathbf{X}'\mathbf{X})$, $\lim_{N\to\infty}(\frac{1}{NT}\mathbf{X}^*(\rho)'\mathbf{X}^*(\rho))$ and $\lim_{N\to\infty}(\frac{1}{NT}\mathbf{X}'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{X})$ are finite and non-singular.
A6 (positive variance of LM tests): $NT^{-1}2(\alpha^2\sigma_1^4 + (1-\alpha)^2(T-1)\sigma_\nu^4)\cdot tr[(\mathbf{H}(\mathbf{A}'\mathbf{A})^{-1})^2] - (NT)^{-1}\, 3\sum_{i=1}^N l_{ii}^2 \left(\alpha^2 T^2 + T\left((1-\alpha) + \frac{2\alpha-1}{T}\right)^2\right) > b_Q$ for some $b_Q > 0$, $\sigma_1^2 = T\sigma_\mu^2 + \sigma_\nu^2$ and $0 \le \alpha \le 1$. $\mathbf{H}$ and l_{ii} are defined in Lemma 4 below.

Consistency of the (quasi-)ML estimates under the general model.

In proving the consistency of (quasi-)MLE, we make use of the following Lemmata.[15]

Lemma 1 *Under the maintained assumptions A1-A3, (i) the row and column sums of* $(\mathbf{A}'\mathbf{A})^{-1}$ *and* $(\mathbf{B}'\mathbf{B})^{-1}$ *are bounded in absolute value, uniformly in* N *and in* $\boldsymbol{\vartheta} \in \boldsymbol{\Xi}$. *(ii) the row and column sums of* $\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})$ *are bounded in absolute value, uniformly in* N *and in* $\boldsymbol{\vartheta} \in \boldsymbol{\Xi}$. *(iii)* $\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}$ *exists.*

Lemma 2 *Under assumptions A1-A3, the matrices* $\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})$ *and* $\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}$ *are positive definite.*

The proof of consistency of the maximum likelihood estimates is based on the concentrated log-likelihood which is

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

As non-stochastic counterpart of $L^c(\boldsymbol{\vartheta})$ we use

$$Q(\boldsymbol{\vartheta}) = \underset{\sigma_\nu^2,\boldsymbol{\beta}}{max} E[L(\boldsymbol{\theta})] = -\frac{NT}{2}(\ln 2\pi + 1) - \frac{NT}{2}\ln\sigma_\nu^{*2}(\boldsymbol{\vartheta}) - \frac{1}{2}\ln\det\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta}).$$

[15] The proofs of these Lemmas are skipped to save space. However, they are included in the long version of the Appendix which is available form the authors.

Theorem 3 *Let Assumptions A1-A5 hold: Then (i) the maximum likelihood estimates of* $\boldsymbol{\vartheta}$ *are unique and consistent. (ii) Assume in addition that* H_0^C *holds:* $\left(\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\vartheta}})-\boldsymbol{\beta}_0\right)\xrightarrow{p}\mathbf{0}$, *where* $\widehat{\boldsymbol{\vartheta}}$ *is a consistent estimator of* $\boldsymbol{\vartheta}$.

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

$$\begin{aligned}
&\frac{1}{NT}tr[\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{M}(\boldsymbol{\vartheta})\mathbf{u}(\boldsymbol{\beta}_0)\mathbf{u}(\boldsymbol{\beta}_0)']\\
=\;&\frac{\sigma_{\nu,0}^2}{NT}tr\left[\left(\mathbf{X}'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{u}(\boldsymbol{\beta}_0)\mathbf{u}(\boldsymbol{\beta}_0)'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{X}\right]\\
\leq\;&\frac{\sigma_{\nu,0}^2}{NT}tr\left[\left(\frac{1}{NT}\mathbf{X}'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{X}\right)^{-1}\right]tr\left[\frac{1}{NT}\left(\mathbf{X}'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{u}(\boldsymbol{\beta}_0)\mathbf{u}(\boldsymbol{\beta}_0)'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{X}\right)\right]\\
\leq\;&\frac{\sigma_{\nu,0}^2}{NT}Kc_1tr\left[\frac{1}{NT}\left(\mathbf{X}'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{u}(\boldsymbol{\beta}_0)\mathbf{u}(\boldsymbol{\beta}_0)'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{X}\right)\right].
\end{aligned}$$

The third line follows since $\left(\mathbf{X}'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{X}\right)^{-1}$ and $\mathbf{X}'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{u}(\boldsymbol{\beta}_0)\mathbf{u}(\boldsymbol{\beta}_0)'*$ $\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{X}$ are positive definite matrices (see Abadir and Magnus, 2005, p. 216 and 329) for all $\boldsymbol{\vartheta}\in\boldsymbol{\Xi}$ and the elements of $\left(\frac{1}{NT}\mathbf{X}'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{X}\right)^{-1}$ are uniformly bounded by some positive constant, say c_1, uniformly in the parameter space of $\boldsymbol{\vartheta}$ by Assumptions A2 (v) and A5 (see also Kapoor, Kelejian and Prucha (2007, p. 118f.). This implies

$$\sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Xi}}\left(\sigma_\nu^{*2}(\boldsymbol{\vartheta})-\widehat{\sigma}_\nu^2(\boldsymbol{\vartheta})\right)\leq\frac{\sigma_{\nu,0}^2}{NT}Kc_1tr\left[\frac{1}{NT}\left(\mathbf{X}'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{u}(\boldsymbol{\beta}_0)\mathbf{u}(\boldsymbol{\beta}_0)'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{X}\right)\right].$$

Now.

$$\begin{aligned}
&\lim_{N\to\infty}\frac{1}{NT}E\left[tr\left[\frac{1}{NT}\left(\mathbf{X}'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{u}(\boldsymbol{\beta}_0)\mathbf{u}(\boldsymbol{\beta}_0)'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{X}\right)\right]\right]\\
=\;&\lim_{N\to\infty}\frac{1}{NT}tr\left[\frac{1}{NT}\left(\mathbf{X}'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta}_0)\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{X}\right)\right]\leq\lim_{N\to\infty}\frac{\sigma_{\nu,0}^2}{NT}Kc_2=0.
\end{aligned}$$

This follows from Assumptions A2 and A5 and the observations made in Kapoor, Kelejian and Prucha (2007, p. 118f.). In particular, $\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta}_0)\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}$ possesses bounded row and column sums, uniformly in N and uniformly in the parameter space of $\boldsymbol{\vartheta}$ using Assumption A2 (v), and the elements of $\mathbf{X}$ are uniformly bounded by Assumption A5. Then the elements of $\frac{1}{NT}\mathbf{X}'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta}_0)*$ $\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{X}$ are bounded, uniformly in N and uniformly in the parameter space of $\boldsymbol{\vartheta}$, say by some constant c_2. Next observe that

$$\begin{aligned} & Var[\tfrac{1}{NT}\left(\mathbf{X}'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{u}(\boldsymbol{\beta}_0)\mathbf{u}(\boldsymbol{\beta}_0)'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{X}\right)] \\ = & \tfrac{2\sigma^4_{\nu,0}}{(NT)^2}tr\left[\left(\tfrac{1}{NT}\left(\mathbf{X}'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta}_0)\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{X}\right)\right)^2\right] \leq \tfrac{2\sigma^4_{\nu,0}}{(NT)^2}K^2c_2^2. \end{aligned}$$

By Chebyshev's inequality, we conclude that $plim_{N\to\infty}\frac{1}{NT}(\mathbf{X}'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{u}(\boldsymbol{\beta}_0)*$ $\mathbf{u}(\boldsymbol{\beta}_0)'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{X}) = 0$ and, hence,

$$\underset{\boldsymbol{\vartheta}\in\boldsymbol{\Xi}}{sup}\left(\sigma^{*2}_\nu(\boldsymbol{\vartheta}) - \widehat{\sigma}^2_\nu(\boldsymbol{\vartheta})\right) = o_p(1).$$

Using the mean value theorem it follows that $\ln\widehat{\sigma}^2_\nu(\boldsymbol{\vartheta}) = \ln\sigma^{*2}_\nu(\boldsymbol{\vartheta}) + \frac{\widehat{\sigma}^2_\nu(\boldsymbol{\vartheta})-\sigma^{*2}_\nu(\boldsymbol{\vartheta})}{\overline{\sigma}^2_\nu(\boldsymbol{\vartheta})}$ with the $\overline{\sigma}^2_\nu(\boldsymbol{\vartheta})$ lying in between $\sigma^{*2}_\nu(\boldsymbol{\vartheta})$. Since $\widehat{\sigma}^2_\nu(\boldsymbol{\vartheta}) - \sigma^{*2}_\nu(\boldsymbol{\vartheta}) = o_p(1)$ uniformly in $\boldsymbol{\Xi}$, $\widehat{\sigma}^2_\nu(\boldsymbol{\vartheta})$ will be bounded away from zero uniformly in probability if $\sigma^{*2}_\nu(\boldsymbol{\vartheta})$ is bounded away from zero. Below we show that $\limsup_{N\to\infty}\max_{\boldsymbol{\vartheta}\in\overline{\mathbf{N}}_\varepsilon(\boldsymbol{\vartheta}_0)}\frac{1}{NT}(Q(\boldsymbol{\vartheta})$ $-\ Q(\boldsymbol{\vartheta}_0)) < 0$ under the present assumptions so that

$$\begin{aligned} & \tfrac{1}{NT}\left(Q(\boldsymbol{\vartheta}) - Q(\boldsymbol{\vartheta}_0)\right) \\ = & -\tfrac{1}{2}\ln\sigma^{*2}_\nu(\boldsymbol{\vartheta}) + \tfrac{1}{2}\ln\sigma^{*2}_\nu(\boldsymbol{\vartheta}_0) - \tfrac{1}{2NT}\ln\left(\det\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})/\det\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta}_0)\right) \\ = & -\tfrac{1}{2}\ln\sigma^{*2}_\nu(\boldsymbol{\vartheta}) + \tfrac{1}{2}\ln\sigma^{*2}_\nu(\boldsymbol{\vartheta}_0) + \tfrac{1}{2NT}\ln\left(\det\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}/\det\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta}_0)^{-1}\right) < 0 \end{aligned}$$

or

$$\ln\sigma^{*2}_\nu(\boldsymbol{\vartheta}) > \ln\sigma^{*2}_\nu(\boldsymbol{\vartheta}_0) + \tfrac{1}{NT}\ln\left(\det\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}/\det\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta}_0)^{-1}\right)$$

uniformly in $\boldsymbol{\vartheta}\in\overline{\mathbf{N}}_\varepsilon(\boldsymbol{\vartheta}_0)$, where $\overline{\mathbf{N}}_\varepsilon(\boldsymbol{\vartheta}_0)$ is the complement of an open neighborhood of $\boldsymbol{\vartheta}_0$ of diameter ε. $\sigma^{*2}_\nu(\boldsymbol{\vartheta}_0) > 0$ by Assumption A1. By Lemmata 1 and 2 $\det\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}/\det\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta}_0)^{-1} > 0$, uniformly in N and uniformly in the parameter space of $\boldsymbol{\vartheta}$ and we conclude that $\sigma^{*2}_\nu(\boldsymbol{\vartheta})$ is bounded away from zero and $\overline{\sigma}^2_\nu(\boldsymbol{\vartheta}) = O_P(1)$ uniformly in $\boldsymbol{\vartheta}$. Therefore, we obtain $\underset{\boldsymbol{\vartheta}\in\boldsymbol{\Xi}}{sup}\frac{2}{NT}\left|L^c(\boldsymbol{\vartheta}) - Q(\boldsymbol{\vartheta})\right| =$ $\underset{\boldsymbol{\vartheta}\in\boldsymbol{\Xi}}{sup}\left|\ln\widehat{\sigma}^2_\nu(\boldsymbol{\vartheta}) - \ln\sigma^{*2}_\nu(\boldsymbol{\vartheta})\right| = \underset{\boldsymbol{\vartheta}\in\boldsymbol{\Xi}}{sup}\frac{1}{\overline{\sigma}^2_\nu(\boldsymbol{\vartheta})}\left|\widehat{\sigma}^2_\nu(\boldsymbol{\vartheta}) - \sigma^{*2}_\nu(\boldsymbol{\vartheta})\right| = o_p(1)$ uniformly in $\boldsymbol{\Xi}$.

Secondly, we have to prove the following uniqueness identification condition (see Lee, 2004a). For any $\varepsilon > 0$, $\limsup_{N\to\infty}\max_{\boldsymbol{\vartheta}\in\overline{\mathbf{N}}_\varepsilon(\boldsymbol{\vartheta}_0)}\frac{1}{NT}(Q(\boldsymbol{\vartheta}) -$

$Q(\boldsymbol{\vartheta}_0)) < 0$, where $\overline{\mathbf{N}}_\varepsilon(\boldsymbol{\vartheta}_0)$ is the complement of an open neighborhood of $\boldsymbol{\vartheta}_0$ of diameter ε. Note, $Q(\boldsymbol{\vartheta}) - Q(\boldsymbol{\vartheta}_0) = -\frac{NT}{2}[\ln \sigma_\nu^{*2}(\boldsymbol{\vartheta}) - \ln \sigma_\nu^{*2}(\boldsymbol{\vartheta}_0)] - \frac{1}{2}\ln[\det \boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})/\det \boldsymbol{\Sigma}_u(\boldsymbol{\vartheta}_0)]$. Now, $\ln \sigma_\nu^{*2}(\boldsymbol{\vartheta}) - \ln \sigma_\nu^{*2}(\boldsymbol{\vartheta}_0) = \ln tr \frac{1}{NT}[\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta}_0)\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}] - \ln \frac{1}{NT} tr[\mathbf{I}_{NT}] = \ln tr \frac{1}{NT}[\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta}_0)\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}]$ and $\limsup_{N\to\infty} \max_{\boldsymbol{\vartheta}\in\overline{\mathbf{N}}_\varepsilon(\boldsymbol{\vartheta}_0)} \frac{1}{NT}(Q(\boldsymbol{\vartheta}) - Q(\boldsymbol{\vartheta}_0)) = \limsup_{N\to\infty} \max_{\boldsymbol{\vartheta}\in\overline{\mathbf{N}}_\varepsilon(\boldsymbol{\vartheta}_0)}(-\frac{1}{2}\ln \frac{1}{NT} tr[\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta}_0)\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}] - \frac{1}{2}\frac{1}{NT}\ln(\det \boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})/\det \boldsymbol{\Sigma}_u(\boldsymbol{\vartheta}_0))) < 0$ 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, the consistency of $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\vartheta}})$ under H_0^A or H_0^C is established by observing that our assumptions imply those made in Theorem 4, part b, given in Kapoor, Kelejian and Prucha (2007). Hence, we conclude that under H_0^A or H_0^C $(NT)^{1/2}\left(\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\vartheta}}) - \widehat{\boldsymbol{\beta}}(\boldsymbol{\vartheta})\right) \xrightarrow{p} \mathbf{0}$, since $\widehat{\boldsymbol{\vartheta}}$ is a consistent estimator of $\boldsymbol{\vartheta}$ as shown above. Note, $\widehat{\boldsymbol{\beta}}(\boldsymbol{\vartheta})$ is a $(NT)^{1/2}$-consistent estimator of $\boldsymbol{\beta}_0$ and the consistency of $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\vartheta}})$ follows. See Lee and Yu (2010b) for a similar proof. ■

Appendix C: LM Test for random effects

The following Lemma is useful in proving Theorems 6 and 7 that derive the asymptotic distribution of the LM tests for the random effects model and the KKP model.

Lemma 4 *Assume that Assumptions A1, A2 and A6 hold and that* $\rho_1 = \rho_2 = \rho$. *Consider the quadratic form* $Q = (\mathbf{Z}_\mu \mathbf{A}^{-1}\boldsymbol{\mu} + (\mathbf{I}_T \otimes \mathbf{A}^{-1})\boldsymbol{\nu})' \left((\alpha \overline{\mathbf{J}}_T + (1-\alpha)\mathbf{E}_T) \otimes \mathbf{H}\right) \cdot (\mathbf{Z}_\mu \mathbf{A}^{-1}\boldsymbol{\mu} + (\mathbf{I}_T \otimes \mathbf{A}^{-1})\boldsymbol{\nu})$, *where* $\mathbf{H}$ *is a conformable symmetric matrix and* $0 \le \alpha \le 1$ *is a real number. Then,*

$$E[Q] = (\alpha\sigma_1^2 + (1-\alpha)\sigma_\nu^2(T-1))tr[\mathbf{H}(\mathbf{A}'\mathbf{A})^{-1}]$$
$$Var[Q] = 2(\alpha^2\sigma_1^4 + (1-\alpha)^2(T-1)\sigma_\nu^4)tr[(\mathbf{H}(\mathbf{A}'\mathbf{A})^{-1})^2] + \alpha^2 T^2 \sum_{i=1}^{N} l_{ii}^2 \sigma_\mu^4 \left(\mu_\mu^{(4)} - 3\right) + ((1-\alpha) + \tfrac{2\alpha - 1}{T})^2 \sum_{i=N+1}^{NT+N} c_{ii}^2 \sigma_\nu^4 \left(\mu_\nu^{(4)} - 3\right)$$

with $\mathbf{L} = \mathbf{A}'^{-1}\mathbf{H}\mathbf{A}^{-1}$, $\mu_\mu^{(4)} = \frac{E[\mu^4]}{\sigma_\mu^4}$, *and* $\mu_\nu^{(4)} = \frac{E[\nu^4]}{\sigma_\nu^4}$. l_{ii} *and* c_{ii} *denote the iith* *elements of* $\mathbf{L}$ *and* $\mathbf{C}$*, respectively, where the latter is defined below and*

$$\frac{Q-E[Q]}{\sqrt{Var[Q]}} \xrightarrow{d} N(0,1).$$

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

$$Q \; : \; = \boldsymbol{\xi}'\mathbf{C}\boldsymbol{\xi} = \boldsymbol{\xi}' \begin{bmatrix} \alpha T\mathbf{L} & \alpha\mathbf{L} & .. & \alpha\mathbf{L} \\ \alpha\mathbf{L} & \mathbf{L}((1-\alpha)+\frac{2\alpha-1}{T}) & .. & \mathbf{L}(\frac{2\alpha-1}{T}) \\ .. & .. & .. & .. \\ \alpha\mathbf{L} & \mathbf{L}(\frac{2\alpha-1}{T}) & .. & \mathbf{L}((1-\alpha)+\frac{2\alpha-1}{T}) \end{bmatrix} \boldsymbol{\xi}$$

$$= \alpha T\boldsymbol{\mu}'\mathbf{L}\boldsymbol{\mu} + 2\alpha\sum_{t=1}^{T}\boldsymbol{\nu}_t'\mathbf{L}\boldsymbol{\mu} + (1-\alpha)\sum_{t=1}^{T}\boldsymbol{\nu}_t'\mathbf{L}\boldsymbol{\nu}_t + (2\alpha-1)\frac{1}{T}\left(\sum_{t=1}^{T}\boldsymbol{\nu}_t'\right)\mathbf{L}\left(\sum_{t=1}^{T}\boldsymbol{\nu}_t\right),$$

where $\mathbf{L} = \mathbf{A}'^{-1}\mathbf{H}\mathbf{A}^{-1}$, $tr(\mathbf{L}) = tr[\mathbf{H}(\mathbf{A}'\mathbf{A})^{-1}]$. $\boldsymbol{\xi} = (\boldsymbol{\mu}', \boldsymbol{\nu}_1', ..., \boldsymbol{\nu}_T')'$ with $E[\boldsymbol{\xi}] =$ 0 and

$$Var[\boldsymbol{\xi}] := \Omega_\xi = \begin{bmatrix} \sigma_\mu^2\mathbf{I}_N & 0 & .. & 0 \\ 0 & \sigma_\nu^2\mathbf{I}_N & .. & 0 \\ .. & .. & .. & .. \\ 0 & .. & .. & \sigma_\nu^2\mathbf{I}_N \end{bmatrix}.$$

Let $\Omega_\xi = \mathbf{S}\mathbf{S}'$ with

$$\mathbf{S} = \begin{bmatrix} \sigma_\mu\mathbf{I}_N & 0 & .. & 0 \\ 0 & \sigma_\nu\mathbf{I}_N & .. & 0 \\ .. & .. & .. & .. \\ 0 & .. & .. & \sigma_\nu\mathbf{I}_N \end{bmatrix}$$

and define $\boldsymbol{\eta} = \mathbf{S}^{-1}\boldsymbol{\xi}$ so that $\mu_\nu^{(4)} = E[\left(\frac{\mu_i}{\sigma_\mu}\right)^4]$ and $\mu_\nu^{(4)} = E[\left(\frac{\nu_i}{\sigma_\nu}\right)^4]$. $c_{ii,*}$ is the iith element of

$$\mathbf{S}'\mathbf{C}\mathbf{S} = \begin{bmatrix} \sigma_\mu^2\alpha T\mathbf{L} & \sigma_\mu\sigma_\nu\alpha\mathbf{L} & .. & \sigma_\mu\sigma_\nu\alpha\mathbf{L} \\ \sigma_\nu\sigma_\mu\alpha\mathbf{L} & \sigma_\nu^2\mathbf{L}((1-\alpha)+\frac{2\alpha-1}{T}) & .. & \sigma_\nu^2\mathbf{L}(\frac{2\alpha-1}{T}) \\ .. & .. & .. & .. \\ \sigma_\nu\sigma_\mu\alpha\mathbf{L} & \sigma_\nu^2\mathbf{L}(\frac{2\alpha-1}{T}) & .. & \sigma_\nu^2\mathbf{L}((1-\alpha)+\frac{2\alpha-1}{T}) \end{bmatrix}.$$

It can easily be verified that

$$\begin{aligned} E[Q] &= (\alpha\sigma_1^2 + (1-\alpha)\sigma_\nu^2(T-1))tr[\mathbf{H}(\mathbf{A}'\mathbf{A})^{-1}] \\ Var[Q] &= \mathbf{2}tr\left(\mathbf{C}\Omega_\xi\mathbf{C}\Omega_\xi\right) + \sum_{i=1}^{N} c_{ii,*}^2\left(\mu_\mu^{(4)} - 3\right) + \sum_{i=N+1}^{NT+N} c_{ii,*}^2\left(\mu_\nu^{(4)} - 3\right) \\ &= 2(\alpha^2\sigma_1^4 + (1-\alpha)^2(T-1)\sigma_\nu^4)tr[(\mathbf{H}(\mathbf{A}'\mathbf{A})^{-1})^2] \\ &\quad + \alpha^2T^2\sum_{i=1}^{N} l_{ii}^2\sigma_\mu^4\left(\mu_\mu^{(4)} - 3\right) + T\left((1-\alpha) + \tfrac{2\alpha-1}{T}\right)^2\sum_{i=1}^{N} l_{ii}^2\sigma_\nu^4\left(\mu_\nu^{(4)} - 3\right). \end{aligned}$$

For $\alpha = 1$ one obtains

$$\begin{aligned} E[Q] &= \sigma_1^2 tr[\mathbf{H}(\mathbf{A}'\mathbf{A})^{-1}] \\ Var[Q] &= 2\sigma_1^4 tr[(\mathbf{H}(\mathbf{A}'\mathbf{A})^{-1})^2] + \sum_{i=1}^{N} l_{ii}^2\left(\sigma_\mu^4 T^2\left(\mu_\mu^{(4)} - 3\right) + \tfrac{1}{T}\sigma_\nu^4\left(\mu_\nu^{(4)} - 3\right)\right) \end{aligned}$$

and for $\alpha = 0$

$$\begin{aligned} E[Q] &= \sigma_\nu^2(T-1)tr[\mathbf{H}(\mathbf{A}'\mathbf{A})^{-1}] \\ Var[Q] &= 2(T-1)\sigma_\nu^4 tr[(\mathbf{H}(\mathbf{A}'\mathbf{A})^{-1})^2] + \tfrac{(T-1)^2}{T}\sum_{i=1}^{N} l_{ii}^2\sigma_\nu^4\left(\mu_\nu^{(4)} - 3\right). \end{aligned}$$

The present assumptions imply that $Var[Q]$ is uniformly bounded away from zero by some positive constant under and that the row and column sums of $\mathbf{A}$, $(\mathbf{A}'\mathbf{A})^{-1}$ and $\mathbf{H}$ are uniformly bounded and so are those of $\mathbf{L}$. Since the elements of $\boldsymbol{\xi}$ are independently distributed by Assumption A1, 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 the claim of the Lemma follows. ■

Under $H_0^A : \rho_1 = \rho_2 = 0$, $\mathbf{B} = \mathbf{A} = \mathbf{I}_N$. Using the general formulas for the score and the information matrix given above one can show that the corresponding LM test statistic is given by

$$\widetilde{LM}_A = \tfrac{1}{2b_A\tilde{\sigma}_1^4}\widetilde{G}_A^2 + \tfrac{1}{2b_A(T-1)\tilde{\sigma}_\nu^4}\widetilde{M}_A^2,$$

where $\widetilde{G}_A = \tilde{\mathbf{u}}'\left[\bar{\mathbf{J}}_T \otimes (\mathbf{W}' + \mathbf{W})\right]\tilde{\mathbf{u}}$, $\widetilde{M}_A = \tilde{\mathbf{u}}'\left[\mathbf{E}_T \otimes (\mathbf{W}' + \mathbf{W})\right]\tilde{\mathbf{u}}$ and $b_A = tr\left[(\mathbf{W}' + \mathbf{W})^2\right]$.

Theorem 5 (LM_A) *Suppose Assumptions A1 - A5 hold and* H_0^A: $\rho_1 = \rho_2 = 0$ *is true. Then,* $\widetilde{LM}_A = \frac{1}{2b_A\widetilde{\sigma}_1^4}\widetilde{G}_A^2 + \frac{1}{2b_A(T-1)\widetilde{\sigma}_\nu^4}\widetilde{M}_A^2$ *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 $G_A = \mathbf{u}'\mathbf{G}_A\mathbf{u}$ and $M_A = \mathbf{u}'\mathbf{M}_A\mathbf{u}$, where $\mathbf{G}_A = \overline{\mathbf{J}}_T \otimes (\mathbf{W}' + \mathbf{W})$, and $\mathbf{M}_A = \mathbf{E}_T \otimes (\mathbf{W}' + \mathbf{W})$.

(i) We can apply Lemma 4 by setting $\alpha = 1$ and $\mathbf{A} = \mathbf{I}_N$ so that $\mathbf{H} =(\mathbf{W}'+\mathbf{W})$ with $tr[\mathbf{H}] = 0$, because $tr[\mathbf{W}] = 0$. Also observe that $l_{ii} = 0$ under H_0^A. 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. $\sigma_1^2\sqrt{2b_A}$ is bounded away from zero by some positive constant as shown in Lemma 4, so $\frac{G_A}{\sigma_1^2\sqrt{2b_A}} \xrightarrow{d} N(0, 1)$.

(ii) Setting $\alpha = 0$ in Lemma 4 implies that $\frac{M_A}{\sigma_\nu^2\sqrt{2(T-1)b_A}} \xrightarrow{d} N(0, 1)$.

(iii) Inspection of the proof in Lemma 4 establishes the independence of G_A and M_A. From Lemma 4 it follows that $\frac{\alpha_1'}{\sigma_1^2\sqrt{2b_A}}G_A + \frac{\alpha_2'}{\sigma_\nu^2\sqrt{2(T-1)b_A}}M_A$ with $\frac{\alpha_1'}{\sigma_1^2\sqrt{2b_A}} + \frac{\alpha_2'}{\sigma_\nu^2\sqrt{2(T-1)b_A}} = 1$ is also asymptotically normal and, hence, the vector of quadratic forms $\left[\frac{G_A}{\sigma_1^2\sqrt{2b_A}}, \frac{M_A}{\sigma_\nu^2\sqrt{2(T-1)b_A}}\right]'$ converges to a bivariate standard normal by the Cramér-Wold device. Consequently, $LM_A = \frac{1}{2b_A\sigma_1^4}G_A^2 + \frac{1}{2b_A(T-1)\sigma_\nu^4}M_A^2$ is asymptotically distributed as χ_2^2.

(iv) Notice that $\frac{1}{\sqrt{NT}}\widetilde{\mathbf{u}}'\mathbf{G}_A\widetilde{\mathbf{u}} - \frac{1}{\sqrt{NT}}\mathbf{u}'\mathbf{G}_A\mathbf{u} = \frac{2}{NT}\mathbf{u}'\mathbf{G}_A\mathbf{X}\sqrt{NT}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + (NT)^{-\frac{3}{2}} \cdot \sqrt{NT}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)'\mathbf{X}'\mathbf{G}_A\mathbf{X}\sqrt{NT}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$. Given a $\sqrt{N}$-consistent estimator of $\boldsymbol{\beta}_0$ under H_0^A, say $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\mathbf{u}} = \mathbf{y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}$, we have $\frac{1}{\sqrt{NT}}\widetilde{\mathbf{u}}'\mathbf{G}_A\widetilde{\mathbf{u}} - \frac{1}{\sqrt{NT}}\mathbf{u}'\mathbf{G}_A\mathbf{u} = o_p(1)$, since $\mathbf{X}$ and $\mathbf{G}_A$ are non-stochastic matrices (see Lemma 1 in Kelejian and Prucha, 2001, p. 229). Similarly, $\frac{1}{\sqrt{NT}}\widetilde{\mathbf{u}}'\mathbf{M}_A\widetilde{\mathbf{u}} - \frac{1}{\sqrt{NT}}\mathbf{u}'\mathbf{M}_A\mathbf{u} = o_p(1)$. Further, $(NT)^{-1}2\sigma_1^4 b_A > c_1 > 0$ for some constant c_1 and $(NT)^{-1}2\sigma_\nu^4 \cdot (T-1)b_A > c_\nu > 0$ for some constant c_ν, since $\sigma_\mu^2 > 0$ and $\sigma_\nu^2 > 0$ by Assumption A1 and $0 < c_{b_A} \leq b_A$ by Assumption A2. As shown in Appendix B, $\widetilde{\sigma}_1^2 = \sigma_1^2 + o_p(1)$ and $\widetilde{\sigma}_\nu^2 = \sigma_\nu^2 + o_p(1)$. Then, Theorem 2 of Kelejian and Prucha (2001, p. 230) implies that $\frac{\widetilde{G}_A}{\sqrt{2\widetilde{\sigma}_1^4 b_A^2}} - \frac{G_A}{\sqrt{2\sigma_1^4 b_A^2}} = o_p(1)$ and $\frac{\widetilde{M}_A}{\sqrt{2\widetilde{\sigma}_\nu^4(T-1)b_A}} - \frac{M_A}{\sqrt{2\sigma_\nu^4(T-1)b_A}} = o_p(1)$. Hence, $\widetilde{LM}_A - LM_A = o_p(1)$. ■

Appendix D: LM Test for the Anselin Model

Remember that under H_0^B

$$\mathbf{\Omega}_u = (\overline{\mathbf{J}}_T \otimes T\sigma_\mu^2 \mathbf{I}_N + \sigma_\nu^2(\mathbf{B}'\mathbf{B})^{-1}) + \sigma_\nu^2(\mathbf{E}_T \otimes (\mathbf{B}'\mathbf{B})^{-1}).$$

We diagonalize $\mathbf{\Omega}_u = \mathbf{SS}'$ so that $\mathbf{\Omega}_u^{-1} = \mathbf{S}'^{-1}\mathbf{S}^{-1}$. In the following the index r stands for restricted estimation so that H_0: $\rho_1 = 0$ is true. Following Kelejian Prucha (2010), let $\eta = \mathbf{S}^{-1}\mathbf{u}$ and $E[\eta_{it}^3] = \mu_\eta^{(3)}$ and $E[\eta_{it}^4] = \mu_\eta^{(4)}$and let θ_k refers to σ_μ^2, σ_ν^2, ρ_1 or ρ_2. In general, one obtains

$$\begin{aligned}
\frac{\partial L}{\partial \boldsymbol{\beta}} \;&: = \mathbf{s}_\beta(\boldsymbol{\beta}, \boldsymbol{\theta}_r) = \mathbf{X}'\mathbf{\Omega}_u^{-1}\mathbf{u} \\
\left.\frac{\partial L}{\partial \theta_k}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_r} \;&: = \mathbf{s}_{\theta_k}(\boldsymbol{\theta}_r) = -\tfrac{1}{2}tr\left[\mathbf{\Omega}_u^{-1}\frac{\partial \mathbf{\Omega}_u}{\partial \theta_k}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_r} + \tfrac{1}{2}\mathbf{u}'\left[\mathbf{\Omega}_u^{-1}\frac{\partial \mathbf{\Omega}_u}{\partial \theta_k}\mathbf{\Omega}_u^{-1}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_r}\mathbf{u} \\
E[\mathbf{s}_{\theta_k}(\boldsymbol{\theta}_r)] &= -\tfrac{1}{2}tr\left[\mathbf{\Omega}_u^{-1}\frac{\partial \mathbf{\Omega}_u}{\partial \theta_k}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_r} + \tfrac{1}{2}tr\left[\left(\mathbf{\Omega}_u^{-1}\frac{\partial \mathbf{\Omega}_u}{\partial \theta_k}\mathbf{\Omega}_u^{-1}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_r}\mathbf{\Omega}_u\right] \\
Cov[s_{\theta_k}(\boldsymbol{\theta}_r), \mathbf{s}_\beta(\boldsymbol{\beta}, \boldsymbol{\theta}_r)] &= \mathbf{0} \\
Cov[s_{\theta_k}(\boldsymbol{\theta}_r), s_{\theta_l}(\boldsymbol{\theta}_r)] &= tr\left[\left(\mathbf{\Omega}_u^{-1}\frac{\partial \mathbf{\Omega}_u}{\partial \theta_k}\mathbf{\Omega}_u^{-1}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_r}\mathbf{\Omega}_u\left(\mathbf{\Omega}_u^{-1}\frac{\partial \mathbf{\Omega}_u}{\partial \theta_k}\mathbf{\Omega}_u^{-1}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_r}\mathbf{\Omega}_u\right] \\
&+ \tfrac{1}{2}\sum_{i=1}^{NT} a_{k,ii}^* a_{l,ii}^* (\mu_\eta^{(4)} - 3),
\end{aligned}$$

where $a_{k,ii*}^2$ is an element of $\mathbf{A}_k^* = \mathbf{S}'\mathbf{\Omega}_{u,r}^{-1} \left.\frac{\partial \mathbf{\Omega}_u}{\partial \theta_k}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_r} \mathbf{\Omega}_{u,r}^{-1}\mathbf{S}$. Note, since $\mathbf{s}_\beta(\boldsymbol{\beta}, \boldsymbol{\theta}_r)$ is linear in $\mathbf{u}$ and $s_{\theta_k}(\boldsymbol{\theta}_r)$ is a quadratic form in $\mathbf{u}$, $Cov[s_{\theta_k}(\boldsymbol{\theta}_r), \mathbf{s}_\beta(\boldsymbol{\beta}, \boldsymbol{\theta}_r)] = \mathbf{0}$ and Ω_u is block diagonal. So we need a matrix of correction factors with elements $\frac{1}{2}\sum_{i=1}^{NT} a_{k,ii}^* a_{l,ii}^*(\mu_\eta^{(4)} - 3)$, which can be calculated numerically. In particular, $\mu_\eta^{(4)} = E[(\mathbf{S}^{-1}\mathbf{u})^4]$ can be estimated from $\widehat{\mathbf{S}}^{-1}\widehat{\mathbf{u}}$ using $\mathbf{\Omega}_u = \mathbf{SS}' = \mathbf{P\Lambda P}'$ or $\mathbf{S} = \mathbf{P\Lambda}^{\frac{1}{2}}$, since $\mathbf{\Omega}_u$ is a real symmetric matrix. It follows that $\mathbf{\Omega}_u^{-1} = \mathbf{S}'^{-1}\mathbf{S}^{-1}$, $Var(\mathbf{S}^{-1}\mathbf{u}) = \mathbf{S}^{-1}\mathbf{\Omega}_u\mathbf{S}'^{-1} = \mathbf{S}^{-1}\mathbf{SS}'\mathbf{S}'^{-1} = \mathbf{I}$. Observe that

$$\tfrac{1}{2}\mathbf{u}'\left[\mathbf{\Omega}_u^{-1}\frac{\partial \mathbf{\Omega}_u}{\partial \theta_k}\mathbf{\Omega}_u^{-1}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_r}\mathbf{u} = \tfrac{1}{2}\boldsymbol{\eta}'\left(\mathbf{S}^{-1} \left.\frac{\partial \mathbf{\Omega}_u}{\partial \theta_k}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_r}\mathbf{S}'^{-1}\right)\boldsymbol{\eta},$$

where the elements of $\boldsymbol{\eta}$ are $iid(0,1)$ so that $\mathbf{A}_k = \mathbf{S}^{-1} \left. \frac{\partial \boldsymbol{\Omega}_u}{\partial \theta_k} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_r} \mathbf{S}'^{-1}$

$$Cov[s_{\theta_k}(\boldsymbol{\theta}_r), s_{\theta_l}(\boldsymbol{\theta}_r)] = tr\left[\boldsymbol{\Omega}_u^{-1} \left.\frac{\partial \boldsymbol{\Omega}_u}{\partial \theta_k}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_r} \boldsymbol{\Omega}_u^{-1} \left.\frac{\partial \boldsymbol{\Omega}_u}{\partial \theta_k}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_r}\right] + \tfrac{1}{2}\sum_{i=1}^{NT} a_{k,ii} a_{l,ii} (\mu_\eta^{(4)} - 3)$$

Defining the 4×4 matrix $\boldsymbol{\Sigma}_\theta$ with klth element $[\frac{1}{2}\sum_{i=1}^{NT} a_{k,ii} a_{l,ii} (\mu_\eta^{(4)} - 3)]$, $\mathbf{R} = [0,0,1,0]$, the robust LM-test statistic following White (1982) is given by

$$LM_{B,robust} = \widehat{\mathbf{d}}_\theta' \widehat{\mathbf{J}}_\theta^{-1} \mathbf{R}' \left(\mathbf{R}\left(\widehat{\mathbf{J}}_\theta^{-1} + \widehat{\mathbf{J}}_\theta^{-1} \widehat{\Sigma}_\theta \widehat{\mathbf{J}}_\theta^{-1}\right)\mathbf{R}'\right)^{-1} \mathbf{R}\widehat{\mathbf{J}}_\theta^{-1} \widehat{\mathbf{d}}_\theta'$$

and asymptotically distributed as χ_1^2.

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. $H_1^B : \Delta \neq 0$, Under H_0^C, $\mathbf{B} = \mathbf{A}$, $\boldsymbol{\Omega}_u = (\sigma_1^2 \overline{\mathbf{J}}_T + \sigma_\nu^2 \mathbf{E}_T) \otimes (\mathbf{A}'\mathbf{A})^{-1}$ and $\boldsymbol{\Omega}_u^{-1} = (\frac{1}{\sigma_1^2}\overline{\mathbf{J}}_T + \frac{1}{\sigma_\nu^2}\mathbf{E}_T) \otimes (\mathbf{A}'\mathbf{A})$. Using the general formulas for the score and for the information matrix given above, the LM test statistic can be derived as

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

where $\bar{b}_C = \bar{e}_C - \bar{d}_C^2/N$ and $\overline{G}_C = \overline{\mathbf{u}}'(\overline{\mathbf{J}}_T \otimes \overline{\mathbf{F}})\overline{\mathbf{u}} - \bar{\sigma}_1^2 tr[\overline{\mathbf{D}}]$.

Theorem 6 **(LM_C)** *Suppose Assumptions A1 - A6 hold and H_0^c: $\rho_1 = \rho_2 = \rho$ is true. Let $\overline{\mathbf{H}} = (\mathbf{W}'\overline{\mathbf{A}} + \overline{\mathbf{A}}'\mathbf{W})$, $\overline{\mathbf{D}} = \overline{\mathbf{H}}(\overline{\mathbf{A}}'\overline{\mathbf{A}})^{-1}$, $\mathbf{L} = \mathbf{A}'^{-1}\mathbf{H}\mathbf{A}^{-1}$ with elements l_{ij}, $\bar{b}_C = \bar{e}_C - \bar{d}_C^2/N$, $\bar{d}_C = tr[\overline{\mathbf{D}}]$, $\bar{e}_C = tr[\overline{\mathbf{D}}^2]$, $\overline{G}_{Cb} = \overline{\mathbf{u}}'(\overline{\mathbf{J}}_T \otimes \overline{\mathbf{F}})\overline{\mathbf{u}}$, $d_b = \frac{\sum_{i=1}^N l_{ii}^2 \left(\sigma_\mu^4 T^2 \left(\mu_\mu^{(4)} - 3\right) + \frac{1}{T}\sigma_\nu^4 \left(\mu_\nu^{(4)} - 3\right)\right)}{2\sigma_1^4 tr[\mathbf{D}^2]}$ and $d_w' = \frac{\sum_{i=1}^N l_{ii}^2 \left(\frac{1}{T}\sigma_\nu^4 \left(\mu_\nu^{(4)} - 3\right)\right)}{(T-1)^2 2\sigma_\nu^4 tr[\mathbf{D}^2]}$. Then, $\overline{LM}_{C,robust} = \frac{T}{2\bar{b}_C(T-1)\bar{\sigma}_1^4}\left(\overline{G}_{Cb} - \bar{\sigma}_1^2 tr[\overline{\mathbf{D}}]\right)^2 \left(\frac{1}{1+(d_b+d_w')\frac{T-1}{T}}\right)$ is asymptotically distributed as χ_1^2. Under normality, $\overline{LM}_C = \frac{T}{2\bar{b}_C(T-1)\bar{\sigma}_1^4}\left(\overline{G}_{Cb} - \bar{\sigma}_1^2 tr[\overline{\mathbf{D}}]\right)^2$ and is asymptotically distributed as χ_1^2.*

Proof. We will make use of the following first order conditions evaluated under H_0^C :

$$\left.\frac{\partial L}{\partial \Delta}\right|_{H_0^C} = -\frac{T\overline{\sigma}_\mu^2}{2\overline{\sigma}_1^2} tr[\overline{\mathbf{D}}] + \frac{1}{2}\mathbf{u}'(\frac{T\overline{\sigma}_\mu^2}{\overline{\sigma}_1^4}\overline{\mathbf{J}}_T \otimes \overline{\mathbf{H}})\mathbf{u} = 0 \quad (10)$$

$$\left.\frac{\partial L}{\partial \rho_2}\right|_{H_0^C} = -\frac{T}{2} tr[\overline{\mathbf{D}}] + \frac{1}{2}\mathbf{u}'[(\frac{1}{\overline{\sigma}_1^2}\overline{\mathbf{J}}_T + \frac{1}{\overline{\sigma}_\nu^2}\mathbf{E}_T) \otimes \overline{\mathbf{H}}]\mathbf{u} = 0. \quad (11)$$

From the first order condition (11)

$$\left.\frac{\partial L}{\partial \rho_2}\right|_{H_0^C} = -\frac{T}{2} tr[\mathbf{D}] + \frac{1}{2}\mathbf{u}'[(\frac{1}{\sigma_1^2}\overline{\mathbf{J}}_T + \frac{1}{\sigma_\nu^2}\mathbf{E}_T) \otimes \overline{\mathbf{H}}]\mathbf{u},$$

we obtain

$$\overline{\sigma}_1^2 tr[\overline{\mathbf{D}}] = \frac{1}{T}\overline{\mathbf{u}}'(\overline{\mathbf{J}}_T \otimes \overline{\mathbf{H}})\overline{\mathbf{u}} + \frac{\overline{\sigma}_1^2}{T\overline{\sigma}_\nu^2}\overline{\mathbf{u}}'(\mathbf{E}_T \otimes \overline{\mathbf{H}})\overline{\mathbf{u}}.$$

Inserting the ML-estimates denoted by a bar in (10) gives the estimated score as

$$\overline{s}_\Delta(\boldsymbol{\theta})|_{H_0^C} = \frac{T-1}{T}\overline{\mathbf{u}}'(\overline{\mathbf{J}}_T \otimes \overline{\mathbf{H}})\overline{\mathbf{u}} - \frac{\overline{\sigma}_1^2}{T\overline{\sigma}_\nu^2}\overline{\mathbf{u}}'(\mathbf{E}_T \otimes \overline{\mathbf{H}})\overline{\mathbf{u}}.$$

Below we will show that $(NT)^{-\frac{1}{2}}\left(\overline{s}_\Delta(\boldsymbol{\theta})|_{H_0^C} - s_\Delta(\boldsymbol{\theta})|_{H_0^C}\right) + o_p(1)$, so we derive the asymptotic distribution of $s_\Delta(\boldsymbol{\theta})|_{H_0^C}$ to establish that of the LM test. Observe that

$$E[s_\Delta(\boldsymbol{\theta})|_{H_0^C}] = \frac{T-1}{T}\sigma_1^2 tr(\mathbf{D}) - \frac{\sigma_1^2}{T\sigma_\nu^2}(T-1)\sigma_\nu^2 tr(\mathbf{D}) = 0$$

$$Var[s_\Delta(\boldsymbol{\theta})|_{H_0^C}] = 2\left(\frac{T-1}{T}\right)^2\sigma_1^4 tr[\mathbf{D}^2] + \left(\frac{T-1}{T}\right)^2\sum_{i=1}^N l_{ii}^2\sigma_\mu^4\left(T^2\left(\mu_\mu^{(4)}-3\right) + \frac{1}{T}\sigma_\nu^4\left(\mu_\nu^{(4)}-3\right)\right)$$
$$+\frac{\sigma_1^4}{T^2\sigma_\nu^4}(T-1)\sigma_\nu^4 tr[\mathbf{D}^2] + \frac{\sigma_1^4}{T^2\sigma_\nu^4}\frac{(T-1)^2}{T}\sigma_\nu^4\sum_{i=1}^N l_{ii}^2\left(\mu_\nu^{(4)}-3\right).$$

using

$$\mathbf{\Omega}_u|_{H_0^C} = \sigma_1^2[\overline{\mathbf{J}}_T \otimes (\mathbf{A}'\mathbf{A})^{-1}] + \sigma_\nu^2[\mathbf{E}_T \otimes (\mathbf{A}'\mathbf{A})^{-1}]$$

and Lemma 4 under $\alpha = 0$ with Q defined as in Lemma 4.

$$E[Q] = \sigma_1^2 tr[\mathbf{H}(\mathbf{A}'\mathbf{A})^{-1}]$$
$$Var[Q] = 2\sigma_1^4 tr[(\mathbf{H}(\mathbf{A}'\mathbf{A})^{-1})^2] + \sum_{i=1}^N l_{ii}^2\sigma_\mu^4\left(T^2\left(\mu_\mu^{(4)}-3\right) + \frac{1}{T}\sigma_\nu^4\left(\mu_\nu^{(4)}-3\right)\right),$$

and, lastly, under $\alpha = 1$

$$\begin{aligned} E[Q] &= \sigma_\nu^2 (T-1) tr[\mathbf{H}(\mathbf{A}'\mathbf{A})^{-1}] \\ Var[Q] &= 2(T-1)\sigma_\nu^4 tr[(\mathbf{H}(\mathbf{A}'\mathbf{A})^{-1})^2] + \tfrac{(T-1)^2}{T}\sigma_\nu^4 \sum_{i=1}^N l_{ii}^2 \left(\mu_\nu^{(4)} - 3\right). \end{aligned}$$

Collecting terms yields

$$\begin{aligned} Var[s_\Delta(\boldsymbol{\theta})|_{H_0^C}] &= 2\sigma_1^4 tr[\mathbf{D}^2]\left(\tfrac{(T-1)^2}{T^2} + \tfrac{T-1}{T^2}\right) \\ &\quad + \left(\tfrac{T-1}{T}\right)^2 \sum_{i=1}^N l_{ii}^2 \sigma_\mu^4 \left(T^2\left(\mu_\mu^{(4)} - 3\right) + \tfrac{1}{T}\sigma_\nu^4\left(\mu_\nu^{(4)} - 3\right)\right) \\ &\quad + \tfrac{\sigma_1^4}{T^2\sigma_\nu^4}\tfrac{(T-1)^2}{T}\sigma_\nu^4 \sum_{i=1}^N l_{ii}^2 \left(\mu_\nu^{(4)} - 3\right). \\ &= 2\sigma_1^4 tr[\mathbf{D}^2]\tfrac{T-1}{T} + \left(\tfrac{T-1}{T}\right)^2 c_b + \tfrac{\sigma_1^4}{T^2\sigma_\nu^4}c_w, \end{aligned}$$

where we define $c_b = \sum_{i=1}^N l_{ii}^2 \left(\sigma_\mu^4 T^2\left(\mu_\mu^{(4)} - 3\right) + \frac{1}{T}\sigma_\nu^4\left(\mu_\nu^{(4)} - 3\right)\right)$ and $c_w = \sum_{i=1}^N l_{ii}^2 \frac{(T-1)^2}{T}\sigma_\nu^4\left(\mu_\nu^{(4)} - 3\right)$ and use $\frac{(T-1)^2}{T^2} + \frac{T-1}{T^2} = \frac{T-1}{T^2}(T-1+1) = \frac{T-1}{T}$. Next we derive the standardized score as

$$Q = \frac{s_\Delta(\boldsymbol{\theta})|_{H_0^C} - E[s_\Delta(\boldsymbol{\theta})|_{H_0^C}]}{\sqrt{Var[s_\Delta(\boldsymbol{\theta})|_{H_0^C}]}} = \frac{\frac{T-1}{T}\mathbf{u}'(\overline{\mathbf{J}}_T \otimes \mathbf{H})\mathbf{u} - \frac{\sigma_1^2}{T\sigma_\nu^2}\mathbf{u}'(\mathbf{E}_T \otimes \mathbf{H})\mathbf{u}}{\sqrt{2\sigma_1^4 tr[\mathbf{D}^2]\frac{T-1}{T} + \left(\frac{T-1}{T}\right)^2 c_b + \frac{\sigma_1^4}{T^2\sigma_\nu^4}c_w}}.$$

Below we show that $Q_b = \frac{\mathbf{u}'(\overline{\mathbf{J}}_T\otimes\mathbf{H})\mathbf{u} - \sigma_1^2 tr[\mathbf{D}]}{\sqrt{2\sigma_1^4 tr[\mathbf{D}^2] + c_b}} \xrightarrow{d} N(0,1)$ and $Q_w = \frac{\mathbf{u}'(\overline{\mathbf{E}}_T\otimes\mathbf{H})\mathbf{u} - \sigma_\nu^2(T-1)tr[\mathbf{D}]}{\sqrt{2(T-1)\sigma_\nu^4 tr[\mathbf{D}^2] + c_w}}$ $\xrightarrow{d} N(0,1)$. Since the two quadratic forms are independent it follows that $Q \xrightarrow{d} N(0,1)$, where

$$Q = \frac{Q_b\sqrt{(T-1)^2 + (T-1)^2 d_b} - Q_w\sqrt{T-1+d_w}}{\sqrt{(T-1)T + (T-1)^2 d_b + d_w}}.$$

and $d_b = \frac{c_b}{2\sigma_1^4 tr[\mathbf{D}^2]}$ and $d_w = \frac{c_w}{2\sigma_\nu^4 tr[\mathbf{D}^2]}$. Inserting the quadratic forms in the nominator of Q yields

$$\begin{aligned} &Q_b\sqrt{(T-1)^2 + (T-1)^2 d_b} - Q_w\sqrt{T-1+d_w} \\ &= G_{Cb}\frac{T-1}{\sqrt{2\sigma_1^4 tr[\mathbf{D}^2]}} - G_{Cw}\frac{1}{\sqrt{2\sigma_\nu^4 tr[\mathbf{D}^2]}}, \end{aligned}$$

where we define $G_{Cb} = \mathbf{u}'(\overline{\mathbf{J}}_T \otimes \mathbf{H})\mathbf{u}$ and $G_{Cw} = \mathbf{u}'(\overline{\mathbf{E}}_T \otimes \mathbf{H})\mathbf{u}$. Remember the denominator is given by $\sqrt{(T-1)T + (T-1)^2 d_b + d_w}$. The test can then be based on

$$\sqrt{LM_{C,robust}} = \frac{G_{Cb}(T-1) - G_{Cw}\frac{\sigma_1^2}{\sigma_\nu^2}}{\sqrt{2\sigma_1^4 tr[\mathbf{D}^2]}\sqrt{(T-1)T + (T-1)^2 d_b + d_w}}.$$

Under normality the test statistic is given by

$$\sqrt{LM_C} = \frac{G_{Cb}(T-1) - G_{Cw}\frac{\sigma_1^2}{\sigma_\nu^2}}{\sqrt{2\sigma_1^4 tr[\mathbf{D}^2]}\sqrt{(T-1)\,T}}.$$

Lastly it can be shown that with the higher moments being estimated consistently that $\overline{LM}_{C,robust} - LM_{C,robust} = o_p(1)$. [16] ■

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.

[16] The proofs of this last claim is skipped to save space. Details are given in the long version of the Appendix which is available form the authors.

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

		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	LR_B	LM_C	LR_C
-0.80	-0.80	1.000	1.000	0.938	0.964	**0.039**	**0.041**
-0.80	-0.50	1.000	1.000	0.985	0.992	0.590	0.565
-0.80	-0.20	0.997	0.998	0.989	0.991	0.919	0.922
-0.80	0.00	0.979	0.982	0.989	0.991	0.982	0.985
-0.80	0.20	0.997	0.997	0.989	0.993	0.999	0.999
-0.80	0.50	1.000	1.000	0.972	0.977	1.000	1.000
-0.80	0.80	1.000	1.000	0.925	0.938	1.000	1.000
-0.50	-0.80	1.000	1.000	0.562	0.595	0.172	0.307
-0.50	-0.50	1.000	1.000	0.692	0.711	**0.046**	**0.046**
-0.50	-0.20	0.913	0.925	0.727	0.742	0.318	0.324
-0.50	0.00	0.614	0.646	0.702	0.729	0.661	0.685
-0.50	0.20	0.888	0.886	0.690	0.724	0.868	0.894
-0.50	0.50	1.000	1.000	0.613	0.632	0.985	0.992
-0.50	0.80	1.000	1.000	0.430	0.450	0.999	1.000
-0.20	-0.80	1.000	1.000	0.144	0.153	0.643	0.755
-0.20	-0.50	1.000	1.000	0.175	0.183	0.209	0.231
-0.20	-0.20	0.663	0.669	0.164	0.167	**0.042**	**0.045**
-0.20	0.00	0.130	0.139	0.158	0.169	0.157	0.171
-0.20	0.20	0.696	0.660	0.186	0.203	0.453	0.499
-0.20	0.50	1.000	1.000	0.131	0.142	0.863	0.910
-0.20	0.80	1.000	1.000	0.095	0.097	0.976	0.996
0.00	-0.80	1.000	1.000	**0.043**	**0.058**	0.822	0.899
0.00	-0.50	1.000	1.000	**0.043**	**0.055**	0.501	0.509
0.00	-0.20	0.582	0.574	**0.045**	**0.059**	0.106	0.099
0.00	0.00	**0.043**	**0.053**	**0.049**	**0.058**	**0.054**	**0.059**
0.00	0.20	0.646	0.602	**0.042**	**0.047**	0.133	0.154
0.00	0.50	1.000	1.000	**0.049**	**0.051**	0.595	0.672
0.00	0.80	1.000	1.000	**0.050**	**0.053**	0.898	0.962
0.20	-0.80	1.000	1.000	0.117	0.092	0.962	0.983
0.20	-0.50	1.000	1.000	0.147	0.126	0.818	0.827
0.20	-0.20	0.605	0.593	0.174	0.142	0.402	0.382
0.20	0.00	0.130	0.110	0.148	0.125	0.131	0.111
0.20	0.20	0.686	0.649	0.171	0.140	**0.048**	**0.053**
0.20	0.50	1.000	1.000	0.134	0.116	0.283	0.348
0.20	0.80	1.000	1.000	0.093	0.082	0.798	0.909
0.50	-0.80	1.000	1.000	0.667	0.632	0.999	0.999
0.50	-0.50	1.000	1.000	0.761	0.728	0.989	0.988
0.50	-0.20	0.901	0.889	0.781	0.739	0.903	0.886
0.50	0.00	0.700	0.664	0.767	0.746	0.706	0.650
0.50	0.20	0.934	0.923	0.771	0.750	0.372	0.302
0.50	0.50	1.000	1.000	0.683	0.662	**0.044**	**0.054**
0.50	0.80	1.000	1.000	0.397	0.402	0.434	0.590
0.80	-0.80	1.000	1.000	0.994	0.995	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	0.998	0.999	0.999	0.997	0.996
0.80	0.20	1.000	1.000	1.000	1.000	0.988	0.977
0.80	0.50	1.000	1.000	0.990	0.997	0.781	0.699
0.80	0.80	1.000	1.000	0.847	0.947	**0.033**	**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, σ^2_μ=5, σ^2_v=15)

		Random effects model		Anselin model		Kelejian-Prucha model	
		H_0^A: ρ_1=0, ρ_2=0		H_0^B: ρ_1=0		H_0^C: ρ_1=ρ_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, σ^2_μ=15, σ^2_v=5)

		Random effects model		Anselin model		Kelejian-Prucha model	
		H_0^A: ρ_1=0, ρ_2=0		H_0^B: ρ_1=0		H_0^C: ρ_1=ρ_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-Kelejian-Prucha models; share of rejections in 2000 replications

(N=50, T=5, σ^2_μ=10, σ^2_ν=10)

n=50, T=5

			$\nu_{it} \sim t(5)$			ν_{it} ~lognormal(0,10)		
	ρ_1	ρ_2	LM	LM-robust	LR	LM	LM-robust	LR
Random effects model, H_0^A: ρ1=0, ρ2=0	0.00	0.00	0.046	0.046	0.053	0.043	0.046	0.049
Anselin model, H_0^B: ρ_1=0	0.00	-0.80	0.043	0.050	0.054	0.045	0.023	0.050
	0.00	-0.50	0.042	0.041	0.053	0.042	0.025	0.047
	0.00	-0.20	0.048	0.047	0.052	0.041	0.024	0.049
	0.00	0.00	0.045	0.043	0.054	0.044	0.035	0.056
	0.00	0.20	0.044	0.045	0.052	0.043	0.027	0.051
	0.00	0.50	0.056	0.053	0.060	0.040	0.021	0.044
	0.00	0.80	0.053	0.070	0.050	0.047	0.022	0.044
Kapoor-Kelejian-Prucha model, H_0^C: ρ_1=ρ_2	-0.80	-0.80	0.044	0.033	0.041	0.045	0.031	0.048
	-0.50	-0.50	0.056	0.051	0.056	0.046	0.039	0.046
	-0.20	-0.20	0.040	0.038	0.043	0.053	0.052	0.058
	0.00	0.00	0.048	0.048	0.052	0.049	0.049	0.056
	0.20	0.20	0.058	0.058	0.060	0.038	0.038	0.049
	0.50	0.50	0.054	0.044	0.063	0.037	0.029	0.053
	0.80	0.80	0.033	0.022	0.054	0.023	0.016	0.048

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

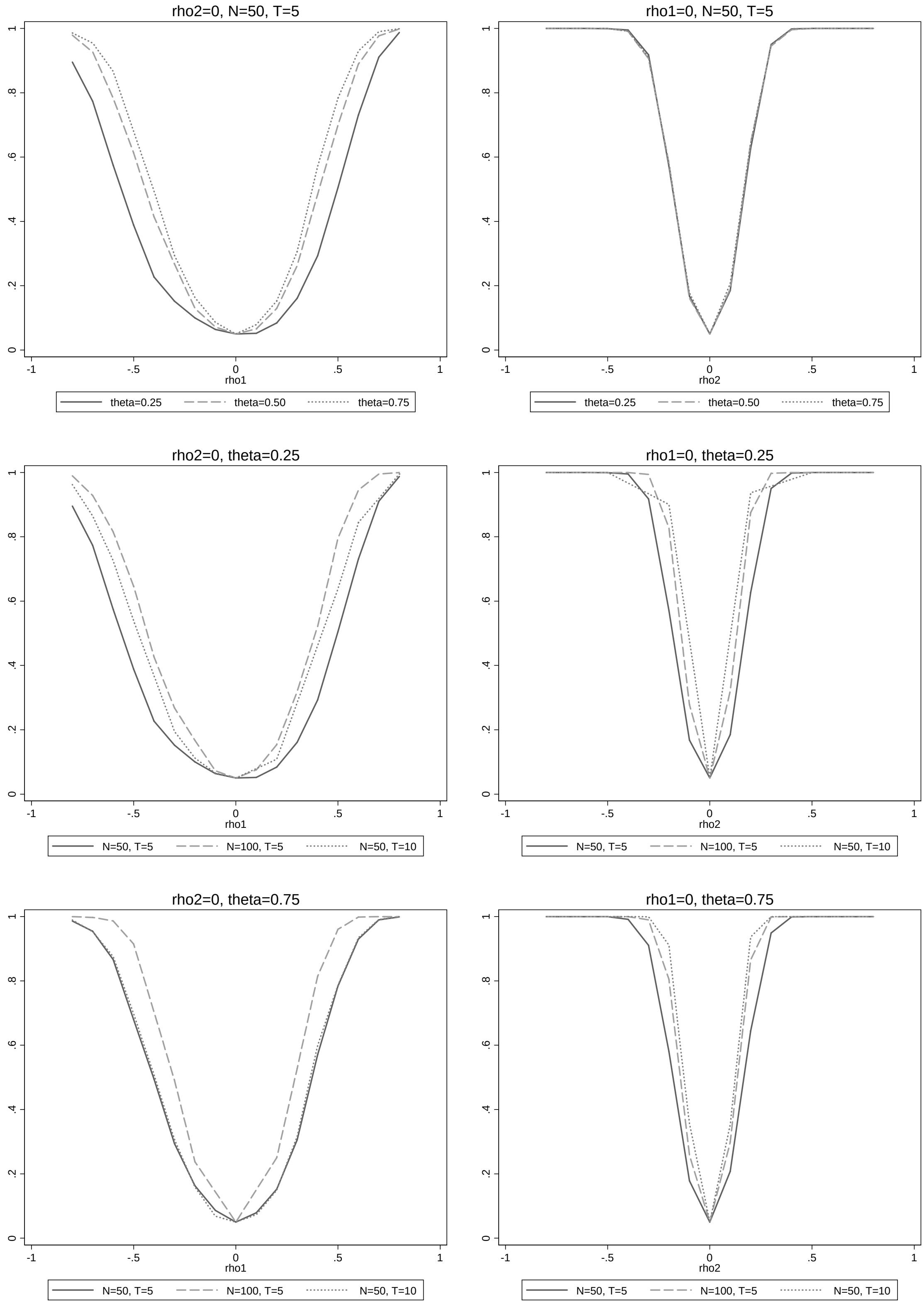


Figure 2: The power of the LM test, Anselin model

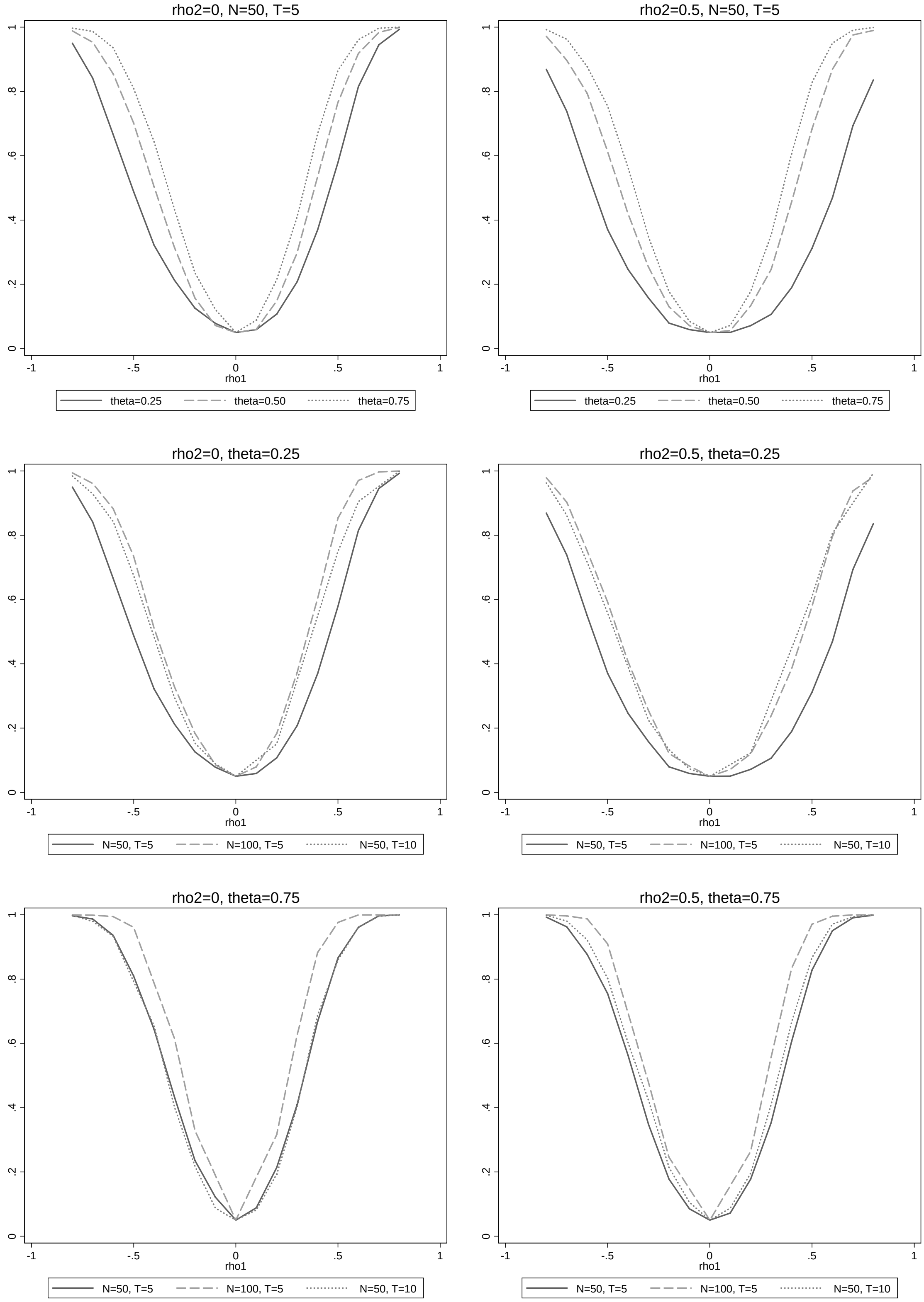


Figure 3: The power of the LM test, KKP model - part I

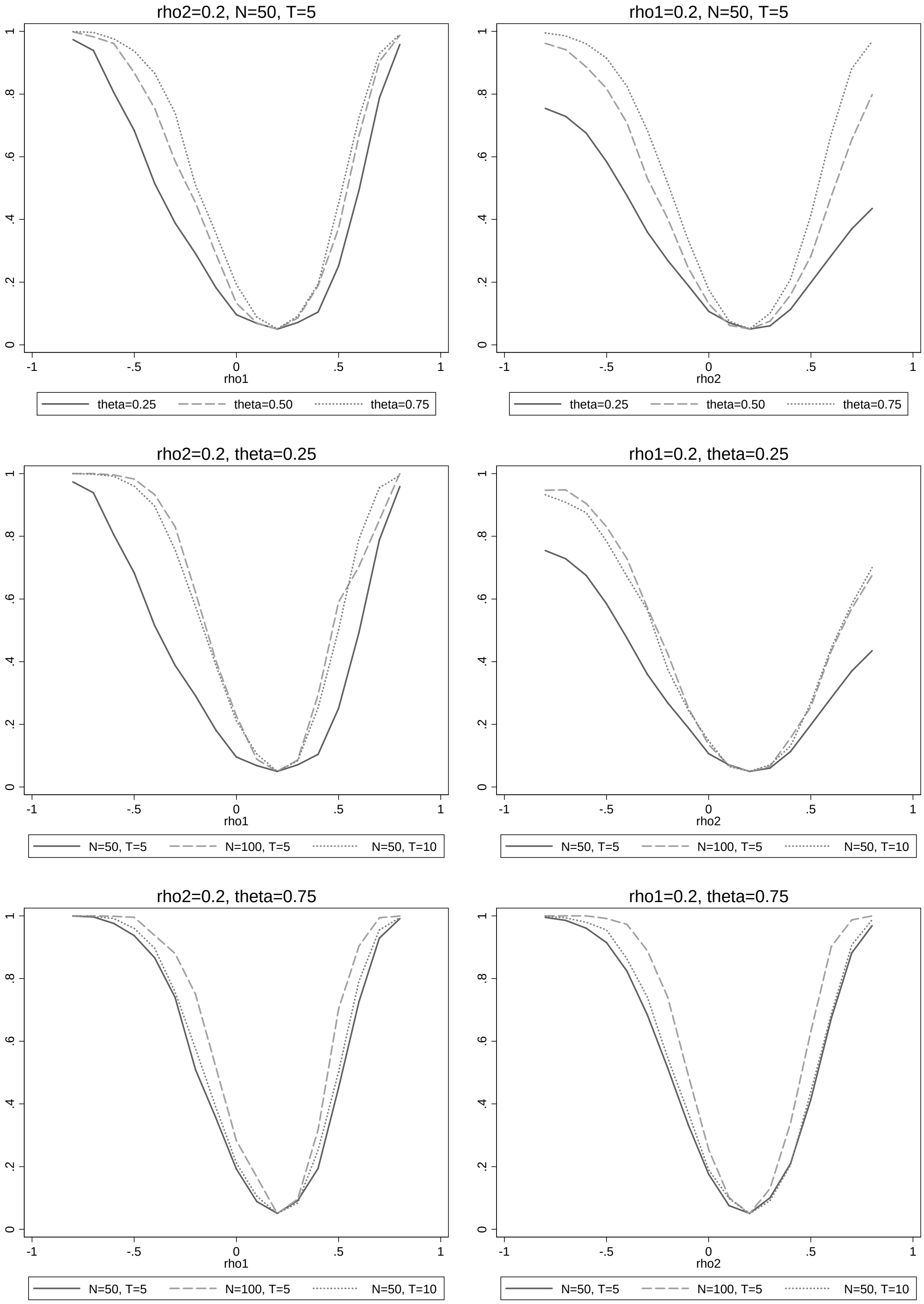
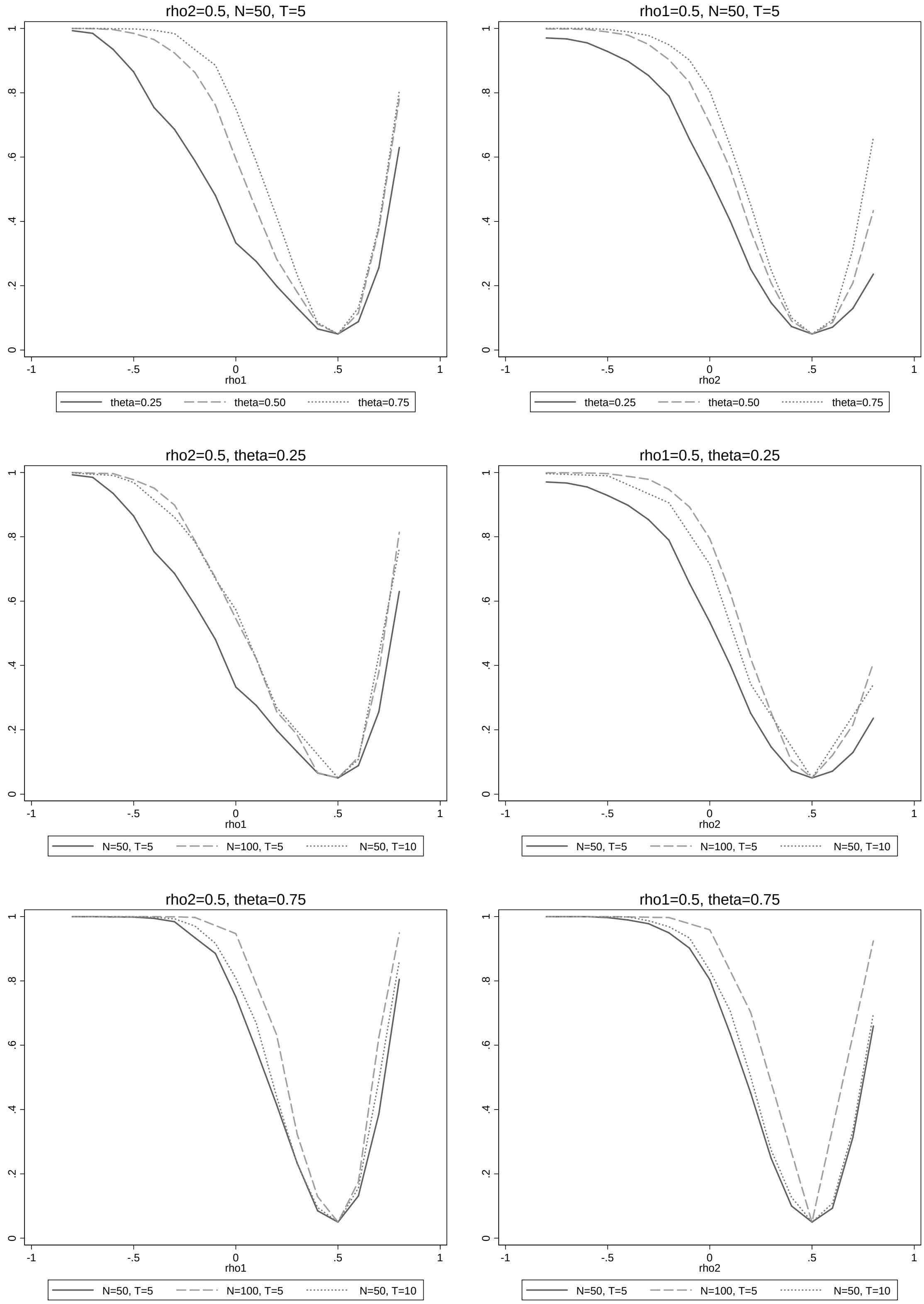


Figure 4: The power of the LM test, KKP model - part II



Detailed Appendices to "A Generalized Spatial Panel Data Model with Random Effects"

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

January 27, 2012

Appendix A: Score and Information Matrix

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

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

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

*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: ETH Zurich and CEPR, WEH E6, Weinbergstrasse 35, 8092 Zurich, Switzerland, E-mail: egger@kof.ethz.ch;

***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.

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 (quasi-)MLE is zero.

First, we drive 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:

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

Observe, that

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

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

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

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

Appendix B: Identification and Consistency

In the sequel, we use subscript 0 to indicate true parameter values where necessary. First, we state the full set of Assumptions.

Assumptions[1]

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 $i.i.d.$ $(0, \sigma_\mu^2)$ with $0 < \sigma_\mu^2 < b_\mu < \infty$. $\boldsymbol{\nu}$ is the vector of remainder errors and its elements are $i.i.d.$ $(0, \sigma_\nu^2)$ with $0 < \sigma_\nu^2 < b_\nu < \infty$. The elements of $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ are independent of each other. Furthermore $E[|\mu_i|^{4+\eta_\mu}] < \infty$ and for some $\eta_\mu > 0$, and $E[|\nu_{it}|^{4+\eta_\nu}] < \infty$ and for some $\eta_\nu > 0$.

A2 (**spatial correlation**):

(i) Both $\mathbf{u}_1$ and $\mathbf{u}_{2t}$ are spatially correlated with the same $(N \times N)$ non-stochastic spatial weighting matrix $\mathbf{W}$ whose elements may depend on N. The elements of $\mathbf{W}$ are non-negative and $w_{ii} = 0$.

(ii) The row and column sums of $\mathbf{W}$ are uniformly bounded in absolute value.

(iii) The parameter space for ρ_r is a closed interval contained in $-1/\lambda_{\max} < \rho_r < 1/\lambda_{\max}$ for all N and $r = 1, 2$, where $\lambda_{\max}$ is the largest absolute eigenvalue of $\mathbf{W}$. $\lambda_{\max}$ is assumed to be bounded away from zero by some fixed positive constant.

(iv) Let $\mathbf{A} = \mathbf{I}_N - \rho_1 \mathbf{W}$ and $\mathbf{B} = \mathbf{I}_N - \rho_2 \mathbf{W}$. The non-stochastic matrices $\mathbf{A}$, $\mathbf{B}$ are non-singular for all ρ_r in the parameter space and have bounded row and column sums, uniformly in N. Also, its inverses have bounded

[1] To 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. For a similar set of assumptions and a discussion of them see Lee (2004a) and Lee and Yu (2010).

row and column sums, uniformly in N and uniformly in the parameter space of ρ_1 and ρ_2 .

(v) The inverse $\mathbf{\Sigma}_u^{-1}\left(\sigma_\mu^2, \sigma_\nu^2, \rho_1, \rho_2\right) = (\overline{\mathbf{J}}_T \otimes (\frac{T\sigma_\mu^2}{\sigma_\nu^2}(\mathbf{A}'\mathbf{A})^{-1} + (\mathbf{B}'\mathbf{B})^{-1})^{-1}) + (\mathbf{E}_T \otimes \mathbf{B}'\mathbf{B})$ has bounded row and column sum uniformly in N and uniformly in the parameter space of $\left(\sigma_\mu^2, \sigma_\nu^2, \rho_1, \rho_2\right)$.[2]

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

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

A4 (**identification of** $\boldsymbol{\vartheta}$): For every $\boldsymbol{\vartheta}\in\mathbf{\Xi}$, $\boldsymbol{\vartheta} \neq \boldsymbol{\vartheta}_0$, and any $\varepsilon > 0$: $\limsup_{N\to\infty} \max_{\boldsymbol{\vartheta}\in\overline{\mathbf{N}}_\varepsilon(\boldsymbol{\vartheta}_0)}(-\frac{1}{2}\ln(\frac{1}{NT}tr[\mathbf{\Sigma}_u(\boldsymbol{\vartheta}_0)\mathbf{\Sigma}_u(\boldsymbol{\vartheta})^{-1}]) - \frac{1}{2}\frac{1}{NT}\ln[\det\mathbf{\Sigma}_u(\boldsymbol{\vartheta})/\det\mathbf{\Sigma}_u(\boldsymbol{\vartheta}_0)]) < 0$, where $\overline{\mathbf{N}}_\varepsilon(\boldsymbol{\vartheta}_0)$ is the complement of an open neighborhood of $\boldsymbol{\vartheta}_0$ of diameter ε.

A5 (**identification of** $\boldsymbol{\beta}$ under H_0^C): The non-random matrix $\mathbf{X}$ has full column rank $K < N$ and its elements are uniformly bounded by some finite constant. Further, let $\mathbf{Q}_0 = \mathbf{E}_T \otimes \mathbf{I}_N$ and $\mathbf{Q}_1 = \overline{\mathbf{J}}_T \otimes \mathbf{I}_N$ and define $\mathbf{X}^*(\rho) = \mathbf{I}_T \otimes \mathbf{A}$. The non-random matrices $\lim_{N\to\infty}(\frac{1}{NT}\mathbf{X}^*(\rho)\mathbf{Q}_i\mathbf{X}^*(\rho))$, $i = 0, 1$ are finite. The nonrandom matrices $\lim_{N\to\infty}(\frac{1}{NT}\mathbf{X}'\mathbf{X})$, $\lim_{N\to\infty}(\frac{1}{NT}\mathbf{X}^*(\rho)'\mathbf{X}^*(\rho))$ and $\lim_{N\to\infty}(\frac{1}{NT}\mathbf{X}'\mathbf{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{X})$ are finite and non-singular.

[2] Under H_0^C we have $\mathbf{\Sigma}_u^{-1}\left(\sigma_\mu^2, \sigma_\nu^2, \rho_1, \rho_2\right) = (\overline{\mathbf{J}}_T \otimes (\frac{T\sigma_\mu^2}{\sigma_\nu^2}(\mathbf{A}'\mathbf{A})^{-1} + (\mathbf{A}'\mathbf{A})^{-1})^{-1}) + (\mathbf{E}_T \otimes \mathbf{A}'\mathbf{A}) = (\overline{\mathbf{J}}_T \otimes (\frac{T\sigma_\mu^2 + \sigma_\nu}{\sigma_\nu^2}\mathbf{I}_N) + (\mathbf{E}_T \otimes \mathbf{I}_N)\left(\mathbf{I}_T \otimes \mathbf{A}'\mathbf{A}\right)$. Hence, in this case a sufficient condition for Assumption A2 (v) is A2 (iv). Note Lemma 1 shows that this inverse exists for all $\left(\sigma_\mu^2, \sigma_\nu^2, \rho_1, \rho_2\right)$ in the parameter space.

A6 (**positive variance of LM tests**): $NT^{-1}2(\alpha^2\sigma_1^4+(1-\alpha)^2(T-1)\sigma_\nu^4)tr[(\mathbf{H}(\mathbf{A}'\mathbf{A})^{-1})^2]-(NT)^{-1}3\sum_{i=1}^{N}l_{ii}^2\left(\alpha^2T^2+T\left((1-\alpha)+\frac{2\alpha-1}{T}\right)^2\right)>b_Q$ for some $b_Q>0$, $\sigma_1^2=T\sigma_\mu^2+\sigma_\nu^2$ and $0\leq\alpha\leq1$. $\mathbf{H}$ and l_{ii} are defined in Lemma 4 below.

Consistency of the (quasi-)ML estimates under the general model.

In proving the consistency of (quasi-)MLE, we make use of the following lemmas.

Lemma 1 *Under the maintained assumptions A1-A3, (i) the row and column sums of* $(\mathbf{A}'\mathbf{A})^{-1}$ *and* $(\mathbf{B}'\mathbf{B})^{-1}$ *are bounded in absolute value, uniformly in* N *and in* $\boldsymbol{\vartheta}\in\boldsymbol{\Xi}$*. (ii) the row and column sums of* $\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})$ *are bounded in absolute value, uniformly in* N *and in* $\boldsymbol{\vartheta}\in\boldsymbol{\Xi}$*. (iii)* $\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}$ *exists.*

Proof. 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, uniformly in N and in $\boldsymbol{\vartheta}\in\boldsymbol{\Xi}$. Since this property is preserved when multiplying matrices of proper dimension (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 uniformly in N and in $\boldsymbol{\vartheta}\in\boldsymbol{\Xi}$, say, by constants c_A and c_B, respectively.

(ii) The row and column sums of $\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})$ are uniformly bounded in absolute value by Assumptions A2 and A3. To see this, denote the typical element of $\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})$ by $\sigma_{ij}(\boldsymbol{\vartheta})$. Then, $max_i\sum_j\sigma_{ij}(\boldsymbol{\vartheta})\leq T\phi c_A+c_B<\infty$ and $max_j\sum_i\sigma_{ij}(\boldsymbol{\vartheta})\leq T\phi c_A+c_B<\infty$.

(iii) Since $\boldsymbol{\Sigma}_u=(\overline{\mathbf{J}}_T\otimes(T\phi(\mathbf{A}'\mathbf{A})^{-1}+(\mathbf{B}'\mathbf{B})^{-1}))+(\mathbf{E}_T\otimes(\mathbf{B}'\mathbf{B})^{-1})$ and $(\mathbf{B}'\mathbf{B})^{-1}$ exists by Assumption A2, it remains to be shown that $(T\phi(\mathbf{A}'\mathbf{A})^{-1}+(\mathbf{B}'\mathbf{B})^{-1})$

is invertible. Using the updating formula we have $(T\phi(\mathbf{A}'\mathbf{A})^{-1}+ (\mathbf{B}'\mathbf{B})^{-1})^{-1}$ $= \mathbf{B}'\mathbf{B} - \mathbf{B}'\mathbf{B}\left(\frac{1}{T\phi}\mathbf{A}'\mathbf{A} + \mathbf{B}'\mathbf{B}\right)^{-1}\mathbf{B}'\mathbf{B}$. The inverse will exist if $\det(\frac{1}{T\phi}\mathbf{A}'\mathbf{A} + \mathbf{B}'\mathbf{B}) \neq 0$. Observe that $\frac{1}{T\phi} > 0$, $\mathbf{A}$ and $\mathbf{B}$ have full rank by Assumption A2 (iv), and that $\mathbf{A}'\mathbf{A}$ and $\mathbf{B}'\mathbf{B}$ are positive definite. We have $\det(\frac{1}{T\phi}\mathbf{A}'\mathbf{A}+\mathbf{B}'\mathbf{B})$ $\geq \det(\frac{1}{T\phi}\mathbf{A}'\mathbf{A}) + \det(\mathbf{B}'\mathbf{B}) > 0$ for all $\boldsymbol{\vartheta} \in \boldsymbol{\Xi}$ (see Abadir and Magnus, 2005, p. 215 and p. 325) and the claim follows. ■

Lemma 2 *Under assumptions A1-A3, the matrices* $\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})$ *and* $\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}$ *are positive definite.*

Proof. Observe that $\det[\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})] = \det[T\phi(\mathbf{A}'\mathbf{A})^{-1}+(\mathbf{B}'\mathbf{B})^{-1}]\det[(\mathbf{B}'\mathbf{B})^{-1}]^{T-1}$ 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) as shown above. Therefore, $\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})$ and $\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}$ are positive definite. ■

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

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

In the following, we use a hat to indicate the maximum likelihood estimates of parameters. The first order conditions for $\boldsymbol{\beta}$ and σ_ν^2 are given by

$$\begin{aligned}
\frac{\partial L(\boldsymbol{\beta},\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} &= \frac{1}{\sigma_\nu^2}\mathbf{X}'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{y} - \mathbf{X}'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{X}\boldsymbol{\beta} \qquad (2)\\
&\Rightarrow \widehat{\boldsymbol{\beta}}(\boldsymbol{\vartheta}) = \left(\mathbf{X}'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{y}\\
\frac{\partial L(\boldsymbol{\beta},\boldsymbol{\theta})}{\partial \sigma_\nu^2} &= -\frac{NT}{2\sigma_\nu^2} + \frac{1}{2\sigma_\nu^4}\mathbf{u}(\widehat{\boldsymbol{\beta}}(\boldsymbol{\vartheta}))'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{u}(\widehat{\boldsymbol{\beta}}(\boldsymbol{\vartheta}))\\
&\Rightarrow \widehat{\sigma}_\nu^2(\boldsymbol{\vartheta}) = \frac{\mathbf{u}(\widehat{\boldsymbol{\beta}}(\boldsymbol{\vartheta}))'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{u}(\widehat{\boldsymbol{\beta}}(\boldsymbol{\vartheta}))}{NT}.
\end{aligned}$$

The concentrated log-likelihood function then reads

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

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

$$E[L(\boldsymbol{\beta}_0,\boldsymbol{\theta})] = -\frac{n}{2}\ln 2\pi - \frac{NT}{2}\ln\sigma_\nu^2 - \frac{1}{2}\ln\left[\det\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})\right] - \frac{\sigma_{v0}^2}{2\sigma_\nu^2}tr[\boldsymbol{\Sigma}(\boldsymbol{\vartheta})^{-1}\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta}_0)]$$

and

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

The non-stochastic counterpart to the concentrated likelihood is given by

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

Theorem 3 *Let Assumptions A1-A5 hold: Then (i) the maximum likelihood estimates of $\boldsymbol{\vartheta}$ are unique and consistent. (ii) Assume in addition that H_0^C holds: $\left(\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\vartheta}}) - \boldsymbol{\beta}_0\right) \xrightarrow{p} \mathbf{0}$, where $\widehat{\boldsymbol{\vartheta}}$ is a consistent estimator of $\boldsymbol{\vartheta}$.*

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 $\frac{1}{NT}\left(L^c(\boldsymbol{\vartheta}) - Q(\boldsymbol{\vartheta})\right) = -\frac{1}{2}(\ln\widehat{\sigma}^2_\nu(\boldsymbol{\vartheta}) - \ln\sigma^{*2}_\nu(\boldsymbol{\vartheta}))$ and that $\widehat{\sigma}^2_\nu(\boldsymbol{\vartheta}) = \frac{1}{NT}\mathbf{u}(\widehat{\boldsymbol{\beta}}(\boldsymbol{\vartheta}))'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{u}(\widehat{\boldsymbol{\beta}}(\boldsymbol{\vartheta})) = \frac{1}{NT}\mathbf{u}(\boldsymbol{\beta}_0)'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{u}(\boldsymbol{\beta}_0) - \mathbf{u}(\boldsymbol{\beta}_0)'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{M}(\boldsymbol{\vartheta})\mathbf{u}(\boldsymbol{\beta}_0) = \frac{1}{NT}tr[\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}(\mathbf{I}_{NT} - \mathbf{M}(\boldsymbol{\vartheta}))\mathbf{u}(\boldsymbol{\beta}_0)\mathbf{u}(\boldsymbol{\beta}_0)']$, where $\mathbf{M}(\boldsymbol{\vartheta}) \equiv \mathbf{X}\left(\mathbf{X}'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}$. Hence, $\ln\widehat{\sigma}^2_\nu(\boldsymbol{\vartheta}) - \ln\sigma^{*2}_\nu(\boldsymbol{\vartheta}) = \frac{1}{NT}tr[\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{M}(\boldsymbol{\vartheta})*$
$\mathbf{u}(\boldsymbol{\beta}_0)\mathbf{u}(\boldsymbol{\beta}_0)']$. Observe, that

$$
\begin{aligned}
& \tfrac{1}{NT}tr[\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{M}(\boldsymbol{\vartheta})\mathbf{u}(\boldsymbol{\beta}_0)\mathbf{u}(\boldsymbol{\beta}_0)'] \\
= & \tfrac{\sigma^2_{\nu,0}}{NT}tr\left[\left(\mathbf{X}'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{u}(\boldsymbol{\beta}_0)\mathbf{u}(\boldsymbol{\beta}_0)'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{X}\right] \\
\leq & \tfrac{\sigma^2_{\nu,0}}{NT}tr\left[\left(\tfrac{1}{NT}\mathbf{X}'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{X}\right)^{-1}\right]tr\left[\tfrac{1}{NT}\left(\mathbf{X}'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{u}(\boldsymbol{\beta}_0)\mathbf{u}(\boldsymbol{\beta}_0)'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{X}\right)\right] \\
\leq & \tfrac{\sigma^2_{\nu,0}}{NT}Kc_1 tr\left[\tfrac{1}{NT}\left(\mathbf{X}'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{u}(\boldsymbol{\beta}_0)\mathbf{u}(\boldsymbol{\beta}_0)'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{X}\right)\right].
\end{aligned}
$$

The third line follows since $\left(\mathbf{X}'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{X}\right)^{-1}$ and $\mathbf{X}'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{u}(\boldsymbol{\beta}_0)\mathbf{u}(\boldsymbol{\beta}_0)'*$ $\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{X}$ are positive definite matrices (see Abadir and Magnus, 2005, p. 216 and 329) for all $\boldsymbol{\vartheta} \in \boldsymbol{\Xi}$ and the elements of $\left(\frac{1}{NT}\mathbf{X}'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{X}\right)^{-1}$ are uniformly bounded by some positive constant, say c_1, uniformly in the parameter space of $\boldsymbol{\vartheta}$ by Assumptions A2 (v) and A5 (see also Kapoor, Kelejian and Prucha (2007, p. 118f.). This implies

$$
\sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Xi}}\left(\sigma^{*2}_\nu(\boldsymbol{\vartheta}) - \widehat{\sigma}^2_\nu(\boldsymbol{\vartheta})\right) \leq \tfrac{\sigma^2_{\nu,0}}{NT}Kc_1 tr\left[\tfrac{1}{NT}\left(\mathbf{X}'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{u}(\boldsymbol{\beta}_0)\mathbf{u}(\boldsymbol{\beta}_0)'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{X}\right)\right].
$$

Now.

$$
\begin{aligned}
& \lim_{N\to\infty}\tfrac{1}{NT}E\left[tr\left[\tfrac{1}{NT}\left(\mathbf{X}'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{u}(\boldsymbol{\beta}_0)\mathbf{u}(\boldsymbol{\beta}_0)'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{X}\right)\right]\right] \\
= & \lim_{N\to\infty}\tfrac{1}{NT}tr\left[\tfrac{1}{NT}\left(\mathbf{X}'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta}_0)\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{X}\right)\right] \leq \lim_{N\to\infty}\tfrac{\sigma^2_{\nu,0}}{NT}Kc_2 = 0.
\end{aligned}
$$

This follows from Assumptions A2 and A5 and the observations made in Kapoor, Kelejian and Prucha (2007, p. 118f.). In particular, we have that

$\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta}_0)\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}$ possesses bounded row and column sums, uniformly in N and uniformly in the parameter space of $\boldsymbol{\vartheta}$ using Assumption A2 (v), and the elements of $\mathbf{X}$ are uniformly bounded by Assumption A5. Then the elements of $\frac{1}{NT}\mathbf{X}'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta}_0)\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{X}$ are bounded, uniformly in N and uniformly in the parameter space of $\boldsymbol{\vartheta}$, say by some constant c_2. Next observe that

$$\begin{aligned}
& Var[\tfrac{1}{NT}\left(\mathbf{X}'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{u}(\boldsymbol{\beta}_0)\mathbf{u}(\boldsymbol{\beta}_0)'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{X}\right)] \\
= & \tfrac{2\sigma_{\nu,0}^4}{(NT)^2} tr\left[\left(\tfrac{1}{NT}\left(\mathbf{X}'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta}_0)\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{X}\right)\right)^2\right] \\
\leq & \tfrac{2\sigma_{\nu,0}^4}{(NT)^2} K^2 c_{2.}^2
\end{aligned}$$

By Chebyshev's inequality, we conclude that $plim_{N\to\infty}\frac{1}{NT}(\mathbf{X}'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{u}(\boldsymbol{\beta}_0)*$ $\mathbf{u}(\boldsymbol{\beta}_0)'\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}\mathbf{X}) = 0$ and, hence,

$$\sup_{\boldsymbol{\vartheta}\in\boldsymbol{\Xi}}\left(\sigma_\nu^{*2}(\boldsymbol{\vartheta}) - \widehat{\sigma}_\nu^2(\boldsymbol{\vartheta})\right) = o_p(1).$$

Using the mean value theorem it follows that $\ln\widehat{\sigma}_\nu^2(\boldsymbol{\vartheta}) = \ln\sigma_\nu^{*2}(\boldsymbol{\vartheta}) + \frac{\widehat{\sigma}_\nu^2(\boldsymbol{\vartheta}) - \sigma_\nu^{*2}(\boldsymbol{\vartheta})}{\overline{\sigma}_\nu^2(\boldsymbol{\vartheta})}$ with the $\overline{\sigma}_\nu^2(\boldsymbol{\vartheta})$ lying in between $\sigma_\nu^{*2}(\boldsymbol{\vartheta})$. Since $\widehat{\sigma}_\nu^2(\boldsymbol{\vartheta}) - \sigma_\nu^{*2}(\boldsymbol{\vartheta}) = o_p(1)$ uniformly in $\boldsymbol{\Xi}$, $\widehat{\sigma}_\nu^2(\boldsymbol{\vartheta})$ will be bounded away from zero uniformly in probability if $\sigma_\nu^{*2}(\boldsymbol{\vartheta})$ is bounded away from zero.

Below we show that $\limsup_{N\to\infty}\max_{\boldsymbol{\vartheta}\in\overline{\mathbf{N}}_\varepsilon(\boldsymbol{\vartheta}_0)}\frac{1}{NT}(Q(\boldsymbol{\vartheta}) - Q(\boldsymbol{\vartheta}_0)) < 0$ under the present assumptions so that

$$\begin{aligned}
& \tfrac{1}{NT}\left(Q(\boldsymbol{\vartheta}) - Q(\boldsymbol{\vartheta}_0)\right) \\
= & -\tfrac{1}{2}\ln\sigma_\nu^{*2}(\boldsymbol{\vartheta}) + \tfrac{1}{2}\ln\sigma_\nu^{*2}(\boldsymbol{\vartheta}_0) - \tfrac{1}{2NT}\ln\left(\det\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})/\det\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta}_0)\right) \\
= & -\tfrac{1}{2}\ln\sigma_\nu^{*2}(\boldsymbol{\vartheta}) + \tfrac{1}{2}\ln\sigma_\nu^{*2}(\boldsymbol{\vartheta}_0) + \tfrac{1}{2NT}\ln\left(\det\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}/\det\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta}_0)^{-1}\right) < 0
\end{aligned}$$

or

$$\ln\sigma_\nu^{*2}(\boldsymbol{\vartheta}) > \ln\sigma_\nu^{*2}(\boldsymbol{\vartheta}_0) + \tfrac{1}{NT}\ln\left(\det\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}/\det\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta}_0)^{-1}\right)$$

uniformly in $\boldsymbol{\vartheta} \in \overline{\mathbf{N}}_{\varepsilon}(\boldsymbol{\vartheta}_0)$,where $\overline{\mathbf{N}}_{\varepsilon}(\boldsymbol{\vartheta}_0)$ is the complement of an open neighborhood of $\boldsymbol{\vartheta}_0$ of diameter ε. $\sigma_{\nu}^{*2}(\boldsymbol{\vartheta}_0) > 0$ by Assumption A1. By Lemmata 1 and 2 $\det \boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1} / \det \boldsymbol{\Sigma}_u(\boldsymbol{\vartheta}_0)^{-1} > 0$, uniformly in N and uniformly in the parameter space of $\boldsymbol{\vartheta}$ and we conclude that $\sigma_{\nu}^{*2}(\boldsymbol{\vartheta})$ is bounded away from zero and $\overline{\sigma}_{\nu}^{2}(\boldsymbol{\vartheta}) = O_P(1)$ uniformly in $\boldsymbol{\vartheta}$. Therefore, we obtain $\underset{\boldsymbol{\vartheta}\in\boldsymbol{\Xi}}{sup} \frac{2}{NT} \left| L^c(\boldsymbol{\vartheta}) - Q(\boldsymbol{\vartheta}) \right| = \underset{\boldsymbol{\vartheta}\in\boldsymbol{\Xi}}{sup} \left| \ln \widehat{\sigma}_{\nu}^{2}(\boldsymbol{\vartheta}) - \ln \sigma_{\nu}^{*2}(\boldsymbol{\vartheta}) \right| = \underset{\boldsymbol{\vartheta}\in\boldsymbol{\Xi}}{sup} \frac{1}{\overline{\sigma}_{\nu}^{2}(\boldsymbol{\vartheta})} \left| \widehat{\sigma}_{\nu}^{2}(\boldsymbol{\vartheta}) - \sigma_{\nu}^{*2}(\boldsymbol{\vartheta}) \right|$ $= o_p(1)$ uniformly in $\boldsymbol{\Xi}$.

Secondly, we have to prove the following uniqueness identification condition (see Lee, 2004a). For any $\varepsilon > 0$, $\limsup_{N\to\infty} \max_{\boldsymbol{\vartheta}\in\overline{\mathbf{N}}_{\varepsilon}(\boldsymbol{\vartheta}_0)} \frac{1}{NT}(Q(\boldsymbol{\vartheta}) - Q(\boldsymbol{\vartheta}_0)) < 0$, where $\overline{\mathbf{N}}_{\varepsilon}(\boldsymbol{\vartheta}_0)$ is the complement of an open neighborhood of $\boldsymbol{\vartheta}_0$ of diameter ε. Note, $Q(\boldsymbol{\vartheta}) - Q(\boldsymbol{\vartheta}_0) = -\frac{NT}{2}[\ln \sigma_{\nu}^{*2}(\boldsymbol{\vartheta}) - \ln \sigma_{\nu}^{*2}(\boldsymbol{\vartheta}_0)] - \frac{1}{2}\ln[\det \boldsymbol{\Sigma}_u(\boldsymbol{\vartheta}) / \det \boldsymbol{\Sigma}_u(\boldsymbol{\vartheta}_0)]$. Now, $\ln \sigma_{\nu}^{*2}(\boldsymbol{\vartheta}) - \ln \sigma_{\nu}^{*2}(\boldsymbol{\vartheta}_0) = \ln tr \frac{1}{NT}[\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta}_0)\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}]$ $- \ln \frac{1}{NT} tr[\mathbf{I}_{NT}] = \ln tr \frac{1}{NT}[\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta}_0)\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}]$ and $\limsup_{N\to\infty} \max_{\boldsymbol{\vartheta}\in\overline{\mathbf{N}}_{\varepsilon}(\boldsymbol{\vartheta}_0)} \frac{1}{NT}(Q(\boldsymbol{\vartheta})$ $- Q(\boldsymbol{\vartheta}_0)) = \limsup_{N\to\infty} \max_{\boldsymbol{\vartheta}\in\overline{\mathbf{N}}_{\varepsilon}(\boldsymbol{\vartheta}_0)} (-\frac{1}{2}\ln \frac{1}{NT} tr[\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta}_0)\boldsymbol{\Sigma}_u(\boldsymbol{\vartheta})^{-1}] -$ $\frac{1}{2}\frac{1}{NT}\ln(\det \boldsymbol{\Sigma}_u(\boldsymbol{\vartheta}) / \det \boldsymbol{\Sigma}_u(\boldsymbol{\vartheta}_0))) < 0$ 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, the consistency of $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\vartheta}})$ under H_0^A or H_0^C is established by observing that our assumptions imply those made in Theorem 4, part b, given in Kapoor, Kelejian and Prucha (2006). Hence, we conclude that under H_0^A or H_0^C $(NT)^{1/2}\left(\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\vartheta}}) - \widehat{\boldsymbol{\beta}}(\boldsymbol{\vartheta})\right) \xrightarrow{p} \mathbf{0}$, since $\widehat{\boldsymbol{\vartheta}}$ is a consistent estimator of $\boldsymbol{\vartheta}$ as shown above. Note, $\widehat{\boldsymbol{\beta}}(\boldsymbol{\vartheta})$ is a $(NT)^{1/2}$-consistent estimator of $\boldsymbol{\beta}_0$ and the consistency of $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\vartheta}})$ follows. See Lee and Yu (2010b) for a similar proof. ■

Appendix C: LM Test for random effects

Below, Theorems 6 and 7 derive the asymptotic distribution of the LM tests for the random effects model and the KKP model. The following lemma is useful in proving these theorems.

Lemma 4 *Assume that Assumptions A1, A2 and A6 hold and that* $\rho_1 = \rho_2 = \rho$. *Consider the quadratic form* $Q = (\mathbf{Z}_\mu \mathbf{A}^{-1}\boldsymbol{\mu} + (\mathbf{I}_T \otimes \mathbf{A}^{-1})\boldsymbol{\nu})' \left((\alpha \overline{\mathbf{J}}_T + (1-\alpha)\mathbf{E}_T) \otimes \mathbf{H}\right) \cdot (\mathbf{Z}_\mu \mathbf{A}^{-1}\boldsymbol{\mu} + (\mathbf{I}_T \otimes \mathbf{A}^{-1})\boldsymbol{\nu})$, *where where* $\mathbf{H}$ *is a conformable symmetric matrix and* $0 \leq \alpha \leq 1$ *is a real number. Then,*

$$E[Q] = (\alpha\sigma_1^2 + (1-\alpha)\sigma_\nu^2(T-1))tr[\mathbf{H}(\mathbf{A}'\mathbf{A})^{-1}].$$

$$\begin{aligned} Var[Q] &= 2(\alpha^2\sigma_1^4 + (1-\alpha)^2 (T-1)\,\sigma_\nu^4)tr[(\mathbf{H}(\mathbf{A}'\mathbf{A})^{-1})^2] \\ &\quad + \alpha^2 T^2 \sum_{i=1}^{N} l_{ii}^2 \sigma_\mu^4 \left(\mu_\mu^{(4)} - 3\right) + ((1-\alpha) + \tfrac{2\alpha-1}{T})^2 \sum_{i=N+1}^{NT+N} c_{ii}^2 \sigma_\nu^4 \left(\mu_\nu^{(4)} - 3\right) \end{aligned}$$

with $\mathbf{L} = \mathbf{A}'^{-1}\mathbf{H}\mathbf{A}^{-1}$, $\mu_\mu^{(4)} = \frac{E[\mu^4]}{\sigma_\mu^4}$, *and* $\mu_\nu^{(4)} = \frac{E[\nu^4]}{\sigma_\nu^4}$. l_{ii} *and* c_{ii} *denote the iith elements of* $\mathbf{L}$ *and* $\mathbf{C}$, *respectively, where the latter is defined below.*
Then

$$\frac{Q - E[Q]}{\sqrt{Var[Q]}} \overset{d}{\rightarrow} N(0,1).$$

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

$$\begin{aligned} Q \;&:= \boldsymbol{\xi}'\mathbf{C}\boldsymbol{\xi} = \boldsymbol{\xi}' \begin{bmatrix} \alpha T\mathbf{L} & \alpha\mathbf{L} & .. & \alpha\mathbf{L} \\ \alpha\mathbf{L} & \mathbf{L}((1-\alpha) + \frac{2\alpha-1}{T}) & .. & \mathbf{L}(\frac{2\alpha-1}{T}) \\ .. & .. & .. & .. \\ \alpha\mathbf{L} & \mathbf{L}(\frac{2\alpha-1}{T}) & .. & \mathbf{L}((1-\alpha) + \frac{2\alpha-1}{T}) \end{bmatrix} \boldsymbol{\xi} \\ &= \alpha T\boldsymbol{\mu}'\mathbf{L}\boldsymbol{\mu} + 2\alpha \sum_{t=1}^{T} \boldsymbol{\nu}_t'\mathbf{L}\boldsymbol{\mu} + (1-\alpha)\sum_{t=1}^{T} \boldsymbol{\nu}_t'\mathbf{L}\boldsymbol{\nu}_t + (2\alpha-1)\tfrac{1}{T}\left(\sum_{t=1}^{T}\boldsymbol{\nu}_t'\right)\mathbf{L}\left(\sum_{t=1}^{T}\boldsymbol{\nu}_t\right), \end{aligned}$$

where $\mathbf{L} = \mathbf{A}'^{-1}\mathbf{H}\mathbf{A}^{-1}$, $tr(\mathbf{L}) = tr[(\mathbf{H}(\mathbf{A}'\mathbf{A})^{-1}]$. $\boldsymbol{\xi} = (\boldsymbol{\mu}', \boldsymbol{\nu}'_1, ..., \boldsymbol{\nu}'_T)'$ with $E[\boldsymbol{\xi}] = 0$ and

$$Var[\boldsymbol{\xi}] := \Omega_\xi = \begin{bmatrix} \sigma_\mu^2 \mathbf{I}_N & 0 & .. & 0 \\ 0 & \sigma_\nu^2 \mathbf{I}_N & .. & 0 \\ .. & .. & .. & .. \\ 0 & .. & .. & \sigma_\nu^2 \mathbf{I}_N \end{bmatrix}.$$

Let

$$\Omega_\xi = \mathbf{S}\mathbf{S}'$$

$$\mathbf{S} = \begin{bmatrix} \sigma_\mu \mathbf{I}_N & 0 & .. & 0 \\ 0 & \sigma_\nu \mathbf{I}_N & .. & 0 \\ .. & .. & .. & .. \\ 0 & .. & .. & \sigma_\nu \mathbf{I}_N \end{bmatrix}$$

and define $\boldsymbol{\eta} = \mathbf{S}^{-1}\boldsymbol{\xi}$ so that $\mu_\nu^{(4)} = E[\left(\frac{\mu_i}{\sigma_\mu}\right)]^4$ and $\mu_\nu^{(4)} = E[\left(\frac{\nu_i}{\sigma_\nu}\right)^4]$. $c_{ii,*}$ is the iith element of

$$\mathbf{S}'\mathbf{C}\mathbf{S} = \begin{bmatrix} \sigma_\mu^2 \alpha T \mathbf{L} & \sigma_\mu \sigma_\nu \alpha \mathbf{L} & .. & \sigma_\mu \sigma_\nu \alpha \mathbf{L} \\ \sigma_\nu \sigma_\mu \alpha \mathbf{L} & \sigma_\nu^2 \mathbf{L}((1-\alpha) + \frac{2\alpha - 1}{T}) & .. & \sigma_\nu^2 \mathbf{L}(\frac{2\alpha-1}{T}) \\ .. & .. & .. & .. \\ \sigma_\nu \sigma_\mu \alpha \mathbf{L} & \sigma_\nu^2 \mathbf{L}(\frac{2\alpha-1}{T}) & .. & \sigma_\nu^2 \mathbf{L}((1-\alpha) + \frac{2\alpha-1}{T}) \end{bmatrix}.$$

It can easily be verified that

$$E[Q] = (\alpha\sigma_1^2 + (1-\alpha)\sigma_\nu^2(T-1))tr[\mathbf{H}(\mathbf{A}'\mathbf{A})^{-1}]$$

and that

$$\begin{aligned} Var[Q] &= \mathbf{2}tr\left(\mathbf{C}\Omega_\xi\mathbf{C}\Omega_\xi\right) + \sum_{i=1}^{N} c_{ii,*}^2\left(\mu_\mu^{(4)} - 3\right) + \sum_{i=N+1}^{NT+N} c_{ii,*}^2\left(\mu_\nu^{(4)} - 3\right) \\ &= 2(\alpha^2\sigma_1^4 + (1-\alpha)^2(T-1)\sigma_\nu^4)tr[(\mathbf{H}(\mathbf{A}'\mathbf{A})^{-1})^2] \\ &\quad + \alpha^2T^2\sum_{i=1}^{N} l_{ii}^2\sigma_\mu^4\left(\mu_\mu^{(4)} - 3\right) + T\left((1-\alpha) + \tfrac{2\alpha-1}{T}\right)^2\sum_{i=1}^{N} l_{ii}^2\sigma_\nu^4\left(\mu_\nu^{(4)} - 3\right). \end{aligned}$$

For $\alpha = 1$ one obtains

$$\begin{aligned} E[Q] &= \sigma_1^2 tr[\mathbf{H}(\mathbf{A}'\mathbf{A})^{-1}] \\ Var[Q] &= 2\sigma_1^4 tr[(\mathbf{H}(\mathbf{A}'\mathbf{A})^{-1})^2] + \sum_{i=1}^{N} l_{ii}^2\left(\sigma_\mu^4T^2\left(\mu_\mu^{(4)} - 3\right) + \tfrac{1}{T}\sigma_\nu^4\left(\mu_\nu^{(4)} - 3\right)\right) \end{aligned}$$

and for $\alpha = 0$

$$\begin{aligned} E[Q] &= \sigma_\nu^2(T-1)tr[\mathbf{H}(\mathbf{A}'\mathbf{A})^{-1}] \\ Var[Q] &= 2(T-1)\sigma_\nu^4 tr[(\mathbf{H}(\mathbf{A}'\mathbf{A})^{-1})^2] + \tfrac{(T-1)^2}{T}\sum_{i=1}^{N} l_{ii}^2\sigma_\nu^4\left(\mu_\nu^{(4)} - 3\right). \end{aligned}$$

Observe that $Var[Q]$ is uniformly bounded away from zero by some positive constant under the present assumptions. Also, the assumptions imply that the row and column sums of $\mathbf{A}$, $(\mathbf{A}'\mathbf{A})^{-1}$ and $\mathbf{H}$ are uniformly bounded and so are those of $\mathbf{L}$. Since the elements of $\boldsymbol{\xi}$ are independently distributed by Assumption A1, 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 the claim of the lemma follows. ■

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

$$\widetilde{LM}_A = \widetilde{\mathbf{D}}_\theta'\widetilde{\mathbf{J}}_\theta^{-1}\widetilde{\mathbf{D}}_\theta, \tag{3}$$

where $\widetilde{\mathbf{D}}_\theta = (\partial L/\partial \boldsymbol{\theta})(\widetilde{\boldsymbol{\theta}})$ 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 (quasi-)MLE, $\widetilde{\boldsymbol{\theta}}$. $\widetilde{\mathbf{J}}_\theta = E[-\partial^2 L/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'](\widetilde{\boldsymbol{\theta}})$ is the part of the information matrix corresponding to $\boldsymbol{\theta}$, also evaluated at the restricted (quasi-)MLE, $\widetilde{\boldsymbol{\theta}}$.

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

$$\begin{aligned} \left.\frac{\partial L}{\partial \rho_1}\right|_{H_0^A} &= \tfrac{\sigma_\mu^2}{2\sigma_1^4}\mathbf{u}'\left[\mathbf{J}_T \otimes (\mathbf{W}' + \mathbf{W})\right]\mathbf{u} \\ \left.\frac{\partial L}{\partial \rho_2}\right|_{H_0^A} &= \tfrac{1}{2}\mathbf{u}'\left[(\tfrac{\sigma_\nu^2}{\sigma_1^4}\bar{\mathbf{J}}_T + \tfrac{1}{\sigma_\nu^2}\mathbf{E}_T) \otimes (\mathbf{W}' + \mathbf{W})\right]\mathbf{u} \end{aligned}$$

and

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

where $b_A = tr\left[(\mathbf{W}' + \mathbf{W})^2\right]$. Note the determinant of the submatrix $\widetilde{\mathbf{J}}_{\rho_1,\rho_2}$ is determined as

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

and its inverse is

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

Defining

$$\widetilde{G}_A = \widetilde{\mathbf{u}}'\left[\bar{\mathbf{J}}_T \otimes (\mathbf{W}' + \mathbf{W})\right]\widetilde{\mathbf{u}}, \qquad \widetilde{M}_A = \widetilde{\mathbf{u}}'\left[\mathbf{E}_T \otimes (\mathbf{W}' + \mathbf{W})\right]\widetilde{\mathbf{u}},$$

we have

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

Theorem 5 **(**LM_A**)** *Suppose Assumptions A1 - A5 hold and* $H_0^A : \rho_1 = \rho_2 = 0$ *is true. Then,* $\widetilde{LM}_A = \frac{1}{2b_A\widetilde{\sigma}_1^4}\widetilde{G}_A^2 + \frac{1}{2b_A(T-1)\widetilde{\sigma}_\nu^4}\widetilde{M}_A^2$ *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 $G_A = \mathbf{u}'\mathbf{G}_A\mathbf{u}$ and $M_A = \mathbf{u}'\mathbf{M}_A\mathbf{u}$, where $\mathbf{G}_A = \overline{\mathbf{J}}_T \otimes (\mathbf{W}' + \mathbf{W})$, and $\mathbf{M}_A = \mathbf{E}_T \otimes (\mathbf{W}' + \mathbf{W})$.

(i) We can apply Lemma 5 by setting $\alpha = 1$ and $\mathbf{A} = \mathbf{I}_N$ so that $\mathbf{H} =(\mathbf{W}'+\mathbf{W})$ with $tr[\mathbf{H}] = 0$, because $tr[\mathbf{W}] = 0$. Also observe that $l_{ii} = 0$ under H_0^A. 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. $\sigma_1^2\sqrt{2b_A}$ is bounded away from zero by some positive constant as shown in Lemma 5, so $\frac{G_A}{\sigma_1^2\sqrt{2b_A}}$ converges in distribution to the standard normal.

(ii) Setting $\alpha = 0$ in Lemma 5 implies that $\frac{M_A}{\sigma_\nu^2\sqrt{2(T-1)b_A}} \xrightarrow{d} N(0,1)$.

(iii) Inspection of the proof in Lemma 5 establishes the independence of G_A and M_A. From Lemma 5 it follows that $\frac{\alpha_1'}{\sigma_1^2\sqrt{2b_A}}G_A + \frac{\alpha_2'}{\sigma_\nu^2\sqrt{2(T-1)b_A}}M_A$ with $\frac{\alpha_1'}{\sigma_1^2\sqrt{2b_A}} + \frac{\alpha_2'}{\sigma_\nu^2\sqrt{2(T-1)b_A}} = 1$ is also asymptotically normal and, hence, the vector of quadratic forms $\left[\frac{G_A}{\sigma_1^2\sqrt{2b_A}}, \frac{M_A}{\sigma_\nu^2\sqrt{2(T-1)b_A}}\right]'$ converges to a bivariate standard normal by the Cramér-Wold device. Consequently, $LM_A = \frac{1}{2b_A\sigma_1^4}G_A^2 + \frac{1}{2b_A(T-1)\sigma_\nu^4}M_A^2$ is asymptotically distributed as χ_2^2.

(iv) Notice that $\frac{1}{\sqrt{NT}}\widetilde{\mathbf{u}}'\mathbf{G}_A\widetilde{\mathbf{u}} - \frac{1}{\sqrt{NT}}\mathbf{u}'\mathbf{G}_A\mathbf{u} = \frac{2}{NT}\mathbf{u}'\mathbf{G}_A\mathbf{X}\sqrt{NT}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + (NT)^{-\frac{3}{2}} \cdot \sqrt{NT}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)'\mathbf{X}'\mathbf{G}_A\mathbf{X}\sqrt{NT}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$. Given a $\sqrt{N}$-consistent estimator of $\boldsymbol{\beta}_0$, say $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\mathbf{u}} = \mathbf{y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}$, we have $\frac{1}{\sqrt{NT}}\widetilde{\mathbf{u}}'\mathbf{G}_A\widetilde{\mathbf{u}} - \frac{1}{\sqrt{NT}}\mathbf{u}'\mathbf{G}_A\mathbf{u} = o_p(1)$, since $\mathbf{X}$ and $\mathbf{G}_A$ are non-stochastic matrices (see Lemma 1 in Kelejian and Prucha, 2001, p. 229). Similarly, $\frac{1}{\sqrt{NT}}\widetilde{\mathbf{u}}'\mathbf{M}_A\widetilde{\mathbf{u}} - \frac{1}{\sqrt{NT}}\mathbf{u}'\mathbf{M}_A\mathbf{u} = o_p(1)$. Further, $(NT)^{-1}2\sigma_1^4 b_A > c_1 > 0$ for some constant c_1 and $(NT)^{-1}2\sigma_\nu^4{\cdot}(T-1)b_A > c_\nu > 0$ for some constant c_ν, since $\sigma_\mu^2 > 0$ and $\sigma_\nu^2 > 0$ by Assumption A1 and $0 < c_{b_A} \leq b_A$ by Assumption A2. As shown in Appendix B, $\widetilde{\sigma}_1^2 = \sigma_1^2 + o_p(1)$ and $\widetilde{\sigma}_\nu^2 = \sigma_\nu^2 + o_p(1)$. Then, Theorem 2 of Kelejian and Prucha (2001, p. 230) implies that $\frac{\widetilde{G}_A}{\sqrt{2\widetilde{\sigma}_1^4 b_A^2}} - \frac{G_A}{\sqrt{2\sigma_1^4 b_A^2}} = o_p(1)$ and $\frac{\widetilde{M}_A}{\sqrt{2\widetilde{\sigma}_\nu^4(T-1)b_A}} - \frac{M_A}{\sqrt{2\sigma_\nu^4(T-1)b_A}} = o_p(1)$. Hence, $\widetilde{LM}_A - LM_A = o_p(1)$. ■

Appendix D: LM Test for the Anselin Model

First, we drive the score and the relevant information submatrix of the general model. These results are than used for the special cases 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:

$$\begin{aligned} L_U(\boldsymbol{\beta}, \sigma_\nu^2, \sigma_\mu^2, \rho_1, \rho_2) &= -\tfrac{NT}{2}\ln 2\pi - \tfrac{1}{2}\ln\det[T\sigma_\mu^2(\mathbf{A}'\mathbf{A})^{-1} + \sigma_\nu^2(\mathbf{B}'\mathbf{B})^{-1}] \quad (4) \\ &\quad -\tfrac{T-1}{2}\ln\det(\sigma_\nu^2(\mathbf{B}'\mathbf{B})^{-\mathbf{1}}) - \tfrac{1}{2}\mathbf{u}'\boldsymbol{\Omega}_u^{-1}\mathbf{u}, \end{aligned}$$

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

We diagonalize $\boldsymbol{\Omega}_u = \mathbf{S}\mathbf{S}'$ so that $\boldsymbol{\Omega}_u^{-1} = \mathbf{S}'^{-1}\mathbf{S}^{-1}$. In the following the index r stands for restricted estimation so that H_0: $\rho_1 = 0$ is true. Following Kelejian Prucha (2010), let $\eta = \mathbf{S}^{-1}\mathbf{u}$ and $E[\eta_{it}^3] = \mu_\eta^{(3)}$ and $E[\eta_{it}^4] = \mu_\eta^{(4)}$and let θ_k refers to σ^2_μ, σ^2_ν, ρ_1 or ρ_2. In general, one obtains

$$
\begin{aligned}
\frac{\partial L}{\partial \boldsymbol{\beta}} \quad &: \; = \mathbf{s}_\beta(\boldsymbol{\beta}, \boldsymbol{\theta}_r) = \mathbf{X}'\boldsymbol{\Omega}_u^{-1}\mathbf{u} \\
\left.\frac{\partial L}{\partial \theta_k}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_r} \quad &: \; = \mathbf{s}_{\theta_k}(\boldsymbol{\theta}_r) = -\tfrac{1}{2} tr \left[\boldsymbol{\Omega}_u^{-1}\frac{\partial \boldsymbol{\Omega}_u}{\partial \theta_k}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_r} + \tfrac{1}{2}\mathbf{u}'\left[\boldsymbol{\Omega}_u^{-1}\frac{\partial \boldsymbol{\Omega}_u}{\partial \theta_k}\boldsymbol{\Omega}_u^{-1}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_r}\mathbf{u} \\
E[\mathbf{s}_{\theta_k}(\boldsymbol{\theta}_r)] &= -\tfrac{1}{2} tr \left[\boldsymbol{\Omega}_u^{-1}\frac{\partial \boldsymbol{\Omega}_u}{\partial \theta_k}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_r} + \tfrac{1}{2} tr \left[\left(\boldsymbol{\Omega}_u^{-1}\frac{\partial \boldsymbol{\Omega}_u}{\partial \theta_k}\boldsymbol{\Omega}_u^{-1}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_r}\boldsymbol{\Omega}_u\right] \\
Cov[s_{\theta_k}(\boldsymbol{\theta}_r), \mathbf{s}_\beta(\boldsymbol{\beta}, \boldsymbol{\theta}_r)] &= \mathbf{0} \\
Cov[s_{\theta_k}(\boldsymbol{\theta}_r), s_{\theta_l}(\boldsymbol{\theta}_r)] &= tr\left[\left(\boldsymbol{\Omega}_u^{-1}\frac{\partial \boldsymbol{\Omega}_u}{\partial \theta_k}\boldsymbol{\Omega}_u^{-1}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_r}\boldsymbol{\Omega}_u\left(\boldsymbol{\Omega}_u^{-1}\frac{\partial \boldsymbol{\Omega}_u}{\partial \theta_k}\boldsymbol{\Omega}_u^{-1}\right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_r}\boldsymbol{\Omega}_u\right] \\
&+ \tfrac{1}{2}\sum_{i=1}^{NT} a^*_{k,ii}a^*_{l,ii}(\mu^{(4)}_\eta - 3),
\end{aligned}
$$

where $a^2_{k,ii*}$ is an element of $\mathbf{A}^*_k = \mathbf{S}'\boldsymbol{\Omega}^{-1}_{u,r} \left.\frac{\partial \boldsymbol{\Omega}_u}{\partial \theta_k}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_r} \boldsymbol{\Omega}^{-1}_{u,r}\mathbf{S}$. Note, since $\mathbf{s}_\beta(\boldsymbol{\beta}, \boldsymbol{\theta}_r)$ is linear in $\mathbf{u}$ and $s_{\theta_k}(\boldsymbol{\theta}_r)$ is a quadratic form in $\mathbf{u}$, $Cov[s_{\theta_k}(\boldsymbol{\theta}_r), \mathbf{s}_\beta(\boldsymbol{\beta}, \boldsymbol{\theta}_r)] = \mathbf{0}$ and Ω_u is block diagonal. So we need a matrix of correction factors with elements $\frac{1}{2}\sum_{i=1}^{NT} a^*_{k,ii}a^*_{l,ii}(\mu^{(4)}_\eta - 3)$, which can be calculated numerically. In particular, $\mu^{(4)}_\eta = E[(\mathbf{S}^{-1}\mathbf{u})^4]$ can be estimated from $\widehat{\mathbf{S}}^{-1}\widehat{\mathbf{u}}$ using $\boldsymbol{\Omega}_u = \mathbf{S}\mathbf{S}' = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}'$ or $\mathbf{S} = \mathbf{P}\boldsymbol{\Lambda}^{\frac{1}{2}}$, since $\boldsymbol{\Omega}_u$ is a real symmetric matrix. It follows that $\boldsymbol{\Omega}_u^{-1} = \mathbf{S}'^{-1}\mathbf{S}^{-1}$, $Var(\mathbf{S}^{-1}\mathbf{u}) = \mathbf{S}^{-1}\boldsymbol{\Omega}_u\mathbf{S}'^{-1} = \mathbf{S}^{-1}\mathbf{S}\mathbf{S}'\mathbf{S}'^{-1} = \mathbf{I}$. Observe that

$$
\begin{aligned}
& \tfrac{1}{2}\mathbf{u}'\left[\boldsymbol{\Omega}_u^{-1}\frac{\partial \boldsymbol{\Omega}_u}{\partial \theta_k}\boldsymbol{\Omega}_u^{-1}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_r}\mathbf{u} \\
= \; & \tfrac{1}{2}\mathbf{u}'\left[\mathbf{S}'^{-1}\mathbf{S}^{-1}\left.\frac{\partial \boldsymbol{\Omega}_u}{\partial \theta_k}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_r}\mathbf{S}'^{-1}\mathbf{S}^{-1}\right]\mathbf{u} \\
= \; & \tfrac{1}{2}\mathbf{u}'\mathbf{S}'^{-1}\left(\mathbf{S}^{-1}\left.\frac{\partial \boldsymbol{\Omega}_u}{\partial \theta_k}\right|_{r}\mathbf{S}'^{-1}\right)\mathbf{S}^{-1}\mathbf{u} \\
= \; & \tfrac{1}{2}\boldsymbol{\eta}'\left(\mathbf{S}^{-1}\left.\frac{\partial \boldsymbol{\Omega}_u}{\partial \theta_k}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_r}\mathbf{S}'^{-1}\right)\boldsymbol{\eta}
\end{aligned}
$$

where the elements of $\boldsymbol{\eta}$ are $iid(0,1)$ so that $\mathbf{A}_k = \mathbf{S}^{-1} \left.\frac{\partial \boldsymbol{\Omega}_u}{\partial \theta_k}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_r} \mathbf{S}'^{-1}$

$$\begin{aligned} Cov[s_{\theta_k}(\boldsymbol{\theta}_r), s_{\theta_l}(\boldsymbol{\theta}_r)] &= tr\left[\mathbf{A}_k\mathbf{A}_l\right] + \tfrac{1}{2}\sum_{i=1}^{NT} a_{k,ii}a_{l,ii}(\mu_\eta^{(4)} - 3) \\ &= tr\left[\boldsymbol{\Omega}_u^{-1} \left.\frac{\partial \boldsymbol{\Omega}_u}{\partial \theta_k}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_r} \boldsymbol{\Omega}_u^{-1} \left.\frac{\partial \boldsymbol{\Omega}_u}{\partial \theta_k}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_r}\right] + \tfrac{1}{2}\sum_{i=1}^{NT} a_{k,ii}a_{l,ii}(\mu_\eta^{(4)} - 3) \end{aligned}$$

Defining the 4×4 matrix $\boldsymbol{\Sigma}_\theta$ with klth element $[\frac{1}{2}\sum_{i=1}^{NT} a_{k,ii}a_{l,ii}(\mu_\eta^{(4)} - 3)]$, $\mathbf{R} = [0,0,1,0]$, the robust LM-test statistic following White (1982) is given by

$$LM_{B,robust} = \widehat{\mathbf{d}}_\theta' \widehat{\mathbf{J}}_\theta^{-1}\mathbf{R}' \left(\mathbf{R}\left(\widehat{\mathbf{J}}_\theta^{-1} + \widehat{\mathbf{J}}_\theta^{-1}\widehat{\Sigma}_\theta\widehat{\mathbf{J}}_\theta^{-1}\right)\mathbf{R}'\right)^{-1} \mathbf{R}\widehat{\mathbf{J}}_\theta^{-1}\widehat{\mathbf{d}}_\theta'$$

and asymptotically distributed as $\chi^2(1)$.

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

$$\left.\frac{\partial L}{\partial \sigma_\nu^2}\right|_{H_0^C} = -\tfrac{N}{2\sigma_1^2} - \tfrac{N(T-1)}{2\sigma_\nu^2} + \tfrac{1}{2}\mathbf{u}'[(\tfrac{1}{\sigma_1^4}\overline{\mathbf{J}}_T + \tfrac{1}{\sigma_\nu^4}\mathbf{E}_T) \otimes \mathbf{A}'\mathbf{A}]\mathbf{u}$$

$$\left.\frac{\partial L}{\partial \sigma_\mu^2}\right|_{H_0^C} = -\tfrac{NT}{2\sigma_1^2} + \tfrac{1}{2}\mathbf{u}'[\tfrac{T}{\sigma_1^4}(\overline{\mathbf{J}}_T \otimes \mathbf{A}'\mathbf{A})]\mathbf{u}$$

$$\left.\frac{\partial L}{\partial \Delta}\right|_{H_0^C} = -\tfrac{T\sigma_\mu^2}{2\sigma_1^2}tr[\mathbf{D}] + \tfrac{1}{2}\mathbf{u}'(\tfrac{T\sigma_\mu^2}{\sigma_1^4}\overline{\mathbf{J}}_T \otimes \mathbf{F})\mathbf{u}$$

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

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

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

The restricted (quasi-)MLE estimates under H_0^C are labeled by a bar. In fact, this gives the (quasi-)MLE version of the KKP model and $\overline{\mathbf{u}} = \mathbf{y} - \mathbf{X}\overline{\boldsymbol{\beta}}$. The score with respect to each element of $\boldsymbol{\theta}$ evaluated at the restricted (quasi-)MLE $\overline{\boldsymbol{\theta}}$ is given by

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

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

$$\overline{\mathbf{J}}_\theta = \frac{1}{2\overline{\sigma}_1^4} \begin{bmatrix} NT \begin{bmatrix} \frac{(T-1)\overline{\sigma}_1^4 + \overline{\sigma}_\nu^4}{T\overline{\sigma}_\nu^4} & 1 \\ 1 & T \end{bmatrix} & T\overline{d}_C \begin{bmatrix} \overline{\sigma}_\mu^2 & \frac{(T-1)\overline{\sigma}_1^4 + \overline{\sigma}_\nu^2\overline{\sigma}_1^2}{T\overline{\sigma}_\nu^2} \\ T\overline{\sigma}_\mu^2 & \overline{\sigma}_1^2 \end{bmatrix} \\ T\overline{d}_C \begin{bmatrix} \overline{\sigma}_\mu^2 & T\overline{\sigma}_\mu^2 \\ \frac{(T-1)\overline{\sigma}_1^4 + \overline{\sigma}_\nu^2\overline{\sigma}_1^2}{T\overline{\sigma}_\nu^2} & \overline{\sigma}_1^2 \end{bmatrix} & T\overline{e}_C \begin{bmatrix} T\overline{\sigma}_\mu^4 & \overline{\sigma}_1^2\overline{\sigma}_\mu^2 \\ \overline{\sigma}_1^2\overline{\sigma}_\mu^2 & \overline{\sigma}_1^4 \end{bmatrix} \end{bmatrix}.$$

To derive the lower right block of the inverse $\overline{\mathbf{J}}_\theta^{-1}$, we employ the formula for the partitioned inverse to obtain

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

where $\overline{b}_C = \overline{e}_C - \overline{d}_C^2/N$. Defining $\overline{G}_{Cb} = \overline{\mathbf{u}}'(\overline{\mathbf{J}}_T \otimes \overline{\mathbf{F}})\overline{\mathbf{u}}$ the resulting LM statistic for H_0^C is given by

$$\overline{LM}_C = \overline{\mathbf{D}}_\theta' \overline{\mathbf{J}}_\theta^{-1} \overline{\mathbf{D}}_\theta = \frac{T\left(\overline{G}_{Cb} - \overline{\sigma}_1^2 tr[\overline{\mathbf{D}}]\right)^2}{2\overline{b}_C(T-1)\overline{\sigma}_1^4}.$$

Theorem 6 **(LM_C)** *Suppose Assumptions A1 - A6 hold and H_0^c: $\rho_1 = \rho_2 = \rho$ is true. Let $\overline{\mathbf{H}} = (\mathbf{W}'\overline{\mathbf{A}} + \overline{\mathbf{A}}'\mathbf{W})$, $\overline{\mathbf{D}} = \overline{\mathbf{H}}(\overline{\mathbf{A}}'\overline{\mathbf{A}})^{-1}$, $\mathbf{L} = \mathbf{A}'^{-1}\mathbf{H}\mathbf{A}^{-1}$ with elements l_{ij}, $\overline{b}_C = \overline{e}_C - \overline{d}_C^2/N$, $\overline{d}_C = tr[\overline{\mathbf{D}}]$, $\overline{e}_C = tr[\overline{\mathbf{D}}^2]$, $\overline{G}_{Cb} = \overline{\mathbf{u}}'(\overline{\mathbf{J}}_T \otimes \overline{\mathbf{F}})\overline{\mathbf{u}}$, $d_b = \frac{\sum_{i=1}^N l_{ii}^2\left(\sigma_\mu^4 T^2\left(\mu_\mu^{(4)}-3\right)+\frac{1}{T}\sigma_\nu^4\left(\mu_\nu^{(4)}-3\right)\right)}{2\sigma_1^4 tr[\mathbf{D}^2]}$ and $d_w' = \frac{\sum_{i=1}^N l_{ii}^2\left(\frac{1}{T}\sigma_\nu^4\left(\mu_\nu^{(4)}-3\right)\right)}{(T-1)^2 2\sigma_\nu^4 tr[\mathbf{D}^2]}$. Then, $\overline{LM}_{C,robust} = \frac{T}{2\overline{b}_C(T-1)\overline{\sigma}_1^4}\left(\overline{G}_{Cb} - \overline{\sigma}_1^2 tr[\overline{\mathbf{D}}]\right)^2\left(\frac{1}{1+(d_b+d_w')\frac{T-1}{T}}\right)$ is asymptotically distributed as χ_1^2. Under normality, $\overline{LM}_C = \frac{T}{2\overline{b}_C(T-1)\overline{\sigma}_1^4}\left(\overline{G}_{Cb} - \overline{\sigma}_1^2 tr[\overline{\mathbf{D}}]\right)^2$ and is asymptotically distributed as χ_1^2.*

Proof. We will make use of the following first order conditions evaluated under H_0^C :

$$\left.\frac{\partial L}{\partial \Delta}\right|_{H_0^C} = -\frac{T\overline{\sigma}_\mu^2}{2\overline{\sigma}_1^2} tr[\overline{\mathbf{D}}] + \frac{1}{2}\mathbf{u}'(\frac{T\overline{\sigma}_\mu^2}{\overline{\sigma}_1^4}\overline{\mathbf{J}}_T \otimes \overline{\mathbf{H}})\mathbf{u} = 0 \quad (5)$$

$$\left.\frac{\partial L}{\partial \rho_2}\right|_{H_0^C} = -\frac{T}{2} tr[\overline{\mathbf{D}}] + \frac{1}{2}\mathbf{u}'[(\frac{1}{\overline{\sigma}_1^2}\overline{\mathbf{J}}_T + \frac{1}{\overline{\sigma}_\nu^2}\mathbf{E}_T) \otimes \overline{\mathbf{H}}]\mathbf{u} = 0. \quad (6)$$

From the first order condition (6)

$$\left.\frac{\partial L}{\partial \rho_2}\right|_{H_0^C} = -\frac{T}{2} tr[\mathbf{D}] + \frac{1}{2}\mathbf{u}'[(\frac{1}{\sigma_1^2}\overline{\mathbf{J}}_T + \frac{1}{\sigma_\nu^2}\mathbf{E}_T) \otimes \overline{\mathbf{H}}]\mathbf{u}$$

we obtain

$$\overline{\sigma}_1^2 tr[\overline{\mathbf{D}}] = \frac{1}{T}\overline{\mathbf{u}}'(\overline{\mathbf{J}}_T \otimes \overline{\mathbf{H}})\overline{\mathbf{u}} + \frac{\overline{\sigma}_1^2}{T\overline{\sigma}_\nu^2}\overline{\mathbf{u}}'(\mathbf{E}_T \otimes \overline{\mathbf{H}})\overline{\mathbf{u}}$$

Inserting in (5) gives the estimated score as

$$\begin{aligned} \overline{s}_\Delta(\boldsymbol{\theta})|_{H_0^C} &= \overline{\mathbf{u}}'(\overline{\mathbf{J}}_T \otimes \overline{\mathbf{H}})\overline{\mathbf{u}} - \frac{1}{T}\overline{\mathbf{u}}'(\overline{\mathbf{J}}_T \otimes \overline{\mathbf{H}})\overline{\mathbf{u}} - \frac{\overline{\sigma}_1^2}{T\overline{\sigma}_\nu^2}\overline{\mathbf{u}}'(\mathbf{E}_T \otimes \overline{\mathbf{H}})\overline{\mathbf{u}} \\ &= \frac{T-1}{T}\overline{\mathbf{u}}'(\overline{\mathbf{J}}_T \otimes \overline{\mathbf{H}})\overline{\mathbf{u}} - \frac{\overline{\sigma}_1^2}{T\overline{\sigma}_\nu^2}\overline{\mathbf{u}}'(\mathbf{E}_T \otimes \overline{\mathbf{H}})\overline{\mathbf{u}}. \end{aligned}$$

Below we will show that $(NT)^{-\frac{1}{2}}\left(\overline{s}_\Delta(\boldsymbol{\theta})|_{H_0^C} - s_\Delta(\boldsymbol{\theta})|_{H_0^C}\right) + o_p(1)$, so we derive the asymptotic distribution of $s_\Delta(\boldsymbol{\theta})|_{H_0^C}$ to establish that of the LM test. Observe that

$$\begin{aligned} E[s_\Delta(\boldsymbol{\theta})|_{H_0^C}] &= E\left[\frac{T-1}{T}\mathbf{u}'(\overline{\mathbf{J}}_T \otimes \mathbf{H})\mathbf{u} - \frac{\sigma_1^2}{T\sigma_\nu^2}\mathbf{u}'(\mathbf{E}_T \otimes \mathbf{H})\mathbf{u}\right] \\ &= \frac{T-1}{T}\sigma_1^2 tr(\mathbf{D}) - \frac{\sigma_1^2}{T\sigma_\nu^2}(T-1)\sigma_\nu^2 tr(\mathbf{D}) = 0 \\ Var[s_\Delta(\boldsymbol{\theta})|_{H_0^C}] &= 2\left(\frac{T-1}{T}\right)^2 \sigma_1^4 tr[\mathbf{D}^2] + \left(\frac{T-1}{T}\right)^2 \sum_{i=1}^N l_{ii}^2 \sigma_\mu^4 \left(T^2\left(\mu_\mu^{(4)} - 3\right) + \frac{1}{T}\sigma_\nu^4\left(\mu_\nu^{(4)} - 3\right)\right) \\ &\quad + \frac{\sigma_1^4}{T^2\sigma_\nu^4}(T-1)\sigma_\nu^4 tr[\mathbf{D}^2] + \frac{\sigma_1^4}{T^2\sigma_\nu^4}\frac{(T-1)^2}{T}\sigma_\nu^4 \sum_{i=1}^N l_{ii}^2\left(\mu_\nu^{(4)} - 3\right). \end{aligned}$$

using

$$\boldsymbol{\Omega}_u|_{H_0^C} = \sigma_1^2[\overline{\mathbf{J}}_T \otimes (\mathbf{A}'\mathbf{A})^{-1}] + \sigma_\nu^2[\mathbf{E}_T \otimes (\mathbf{A}'\mathbf{A})^{-1}]$$

and Lemma 4 under $\alpha = 0$ with Q as in Lemma 4

$$\begin{aligned} E[Q] &= \sigma_1^2 tr[\mathbf{H}(\mathbf{A}'\mathbf{A})^{-1}] \\ Var[Q] &= 2\sigma_1^4 tr[(\mathbf{H}(\mathbf{A}'\mathbf{A})^{-1})^2] + \sum_{i=1}^{N} l_{ii}^2 \sigma_\mu^4 \left(T^2 \left(\mu_\mu^{(4)} - 3 \right) + \tfrac{1}{T}\sigma_\nu^4 \left(\mu_\nu^{(4)} - 3 \right) \right), \end{aligned}$$

and under $\alpha = 1$

$$\begin{aligned} E[Q] &= \sigma_\nu^2 (T-1) tr[\mathbf{H}(\mathbf{A}'\mathbf{A})^{-1}] \\ Var[Q] &= 2(T-1)\sigma_\nu^4 tr[(\mathbf{H}(\mathbf{A}'\mathbf{A})^{-1})^2] + \tfrac{(T-1)^2}{T}\sigma_\nu^4 \sum_{i=1}^{N} l_{ii}^2 \left(\mu_\nu^{(4)} - 3 \right). \end{aligned}$$

Collecting terms yields

$$\begin{aligned} Var[s_\Delta(\boldsymbol{\theta})|_{H_0^C}] &= 2\left(\tfrac{T-1}{T}\right)^2 \sigma_1^4 tr[\mathbf{D}^2] + \left(\tfrac{T-1}{T}\right)^2 \sum_{i=1}^{N} l_{ii}^2 \sigma_\mu^4 \left(T^2 \left(\mu_\mu^{(4)} - 3 \right) + \tfrac{1}{T}\sigma_\nu^4 \left(\mu_\nu^{(4)} - 3 \right) \right) \\ &\quad + 2\tfrac{\sigma_1^4}{T^2\sigma_\nu^4}(T-1)\sigma_\nu^4 tr[\mathbf{D}^2] + \tfrac{\sigma_1^4}{T^2\sigma_\nu^4}\tfrac{(T-1)^2}{T}\sigma_\nu^4 \sum_{i=1}^{N} l_{ii}^2 \left(\mu_\nu^{(4)} - 3 \right) \\ &= 2\sigma_1^4 tr[\mathbf{D}^2] \left(\tfrac{(T-1)^2}{T^2} + \tfrac{T-1}{T^2} \right) \\ &\quad + \left(\tfrac{T-1}{T}\right)^2 \sum_{i=1}^{N} l_{ii}^2 \sigma_\mu^4 \left(T^2 \left(\mu_\mu^{(4)} - 3 \right) + \tfrac{1}{T}\sigma_\nu^4 \left(\mu_\nu^{(4)} - 3 \right) \right) \\ &\quad + \tfrac{\sigma_1^4}{T^2\sigma_\nu^4}\tfrac{(T-1)^2}{T}\sigma_\nu^4 \sum_{i=1}^{N} l_{ii}^2 \left(\mu_\nu^{(4)} - 3 \right). \\ &= 2\sigma_1^4 tr[\mathbf{D}^2]\tfrac{T-1}{T} + \left(\tfrac{T-1}{T}\right)^2 c_b + \tfrac{\sigma_1^4}{T^2\sigma_\nu^4} c_w, \end{aligned}$$

where we define $c_b = \sum_{i=1}^N l_{ii}^2 \left(\sigma_\mu^4 T^2 \left(\mu_\mu^{(4)} - 3 \right) + \frac{1}{T}\sigma_\nu^4 \left(\mu_\nu^{(4)} - 3 \right) \right)$ and $c_w = \sum_{i=1}^N l_{ii}^2 \frac{(T-1)^2}{T}\sigma_\nu^4 \left(\mu_\nu^{(4)} - 3 \right)$ and use $\frac{(T-1)^2}{T^2} + \frac{T-1}{T^2} = \frac{T-1}{T^2}(T-1+1) = \frac{T-1}{T}$. Next we derive the standardized score as

$$Q = \frac{s_\Delta(\boldsymbol{\theta})|_{H_0^C} - E[s_\Delta(\boldsymbol{\theta})|_{H_0^C}]}{\sqrt{Var[s_\Delta(\boldsymbol{\theta})|_{H_0^C}]}} = \frac{\frac{T-1}{T}\mathbf{u}'(\overline{\mathbf{J}}_T \otimes \mathbf{H})\mathbf{u} - \frac{\sigma_1^2}{T\sigma_\nu^2}\mathbf{u}'(\mathbf{E}_T \otimes \mathbf{H})\mathbf{u}}{\sqrt{2\sigma_1^4 tr[\mathbf{D}^2]\frac{T-1}{T} + \left(\frac{T-1}{T}\right)^2 c_b + \frac{\sigma_1^4}{T^2\sigma_\nu^4} c_w}},$$

with

$$\begin{aligned} E[\mathbf{u}'(\overline{\mathbf{J}}_T \otimes \mathbf{H})\mathbf{u}] &= \sigma_1^2 tr[\mathbf{D}] \\ Var[\mathbf{u}'(\overline{\mathbf{J}}_T \otimes \mathbf{H})\mathbf{u}] &= 2\sigma_1^4 tr[\mathbf{D}^2] + c_b \end{aligned}$$

$$\begin{aligned} E[\mathbf{u}'(\overline{\mathbf{E}}_T \otimes \mathbf{H})\mathbf{u}] &= \sigma_1^2 (T-1) tr[\mathbf{D}] \\ Var[\mathbf{u}'(\overline{\mathbf{E}}_T \otimes \mathbf{H})\mathbf{u}] &= 2(T-1)\sigma_v^4 tr[\mathbf{D}^2] + c_w. \end{aligned}$$

Below we show that $Q_b = \frac{\mathbf{u}'(\overline{\mathbf{J}}_T \otimes \mathbf{H})\mathbf{u} - \sigma_1^2 tr[\mathbf{D}]}{\sqrt{2\sigma_1^4 tr[\mathbf{D}^2]+c_b}} \xrightarrow{d} N(0,1)$ and $Q_w = \frac{\mathbf{u}'(\overline{\mathbf{E}}_T \otimes \mathbf{H})\mathbf{u} - \sigma_\nu^2 (T-1) tr[\mathbf{D}]}{\sqrt{2(T-1)\sigma_\nu^4 tr[\mathbf{D}^2]+c_w}} \xrightarrow{d} N(0,1)$. Since the two quadratic forms are independent it follows that $Q \xrightarrow{d} N(0,1)$, where

$$Q = \frac{\frac{T-1}{T} Q_b \sqrt{2\sigma_1^4 tr[\mathbf{D}^2] + c_b} - \frac{\sigma_1^2}{T\sigma_\nu^2} Q_w \sqrt{2(T-1)\sigma_\nu^4 tr[\mathbf{D}^2] + c_w}}{\sqrt{2\sigma_1^4 tr[\mathbf{D}^2]\frac{T-1}{T} + \left(\frac{T-1}{T}\right)^2 c_b + \frac{\sigma_1^4}{T^2\sigma_\nu^4} c_w}}.$$

For convenience we define $d_b = \frac{c_b}{2\sigma_1^4 tr[\mathbf{D}^2]}$ and $d_w = \frac{c_w}{2\sigma_\nu^4 tr[\mathbf{D}^2]}$ and rewrite Q as

$$\begin{aligned} Q &= \frac{Q_b \sqrt{2\left(\frac{T-1}{T}\right)^2 \sigma_1^4 tr[\mathbf{D}^2] + c_b} - Q_w \frac{\sigma_1^2}{T\sigma_\nu^2} \sqrt{2(T-1)\sigma_\nu^4 tr[\mathbf{D}^2] + c_w}}{\sqrt{2\sigma_1^4 tr[\mathbf{D}^2]\frac{T-1}{T} + \left(\frac{T-1}{T}\right)^2 c_b + \frac{\sigma_1^4}{T^2\sigma_\nu^4} c_w}} \\ &= \frac{Q_b \sqrt{2\left(\frac{T-1}{T}\right)^2 \sigma_1^4 tr[\mathbf{D}^2] + c_b} - Q_w \sqrt{2\frac{T-1}{T^2}\sigma_1^4 tr[\mathbf{D}^2] + \frac{\sigma_1^4}{T^2\sigma_\nu^4} c_w}}{\sqrt{2\sigma_1^4 tr[\mathbf{D}^2]\frac{T-1}{T} + \left(\frac{T-1}{T}\right)^2 c_b + \frac{\sigma_1^4}{T^2\sigma_\nu^4} c_w}} \\ &= \frac{Q_b \frac{T-1}{T}\sigma_1^2 \sqrt{2tr[\mathbf{D}^2]}\sqrt{1+d_b} - Q_w \sigma_1^2 \sqrt{2tr[\mathbf{D}^2]} \sqrt{\frac{T-1}{T^2} + \frac{\sigma_1^4 c_w}{T^2\sigma_\nu^4\sigma_1^4\sqrt{2tr[\mathbf{D}^2]}}}}{\sigma_1^2\sqrt{2tr[\mathbf{D}^2]}\sqrt{\frac{T-1}{T} + \left(\frac{T-1}{T}\right)^2 \frac{c_b}{2\sigma_1^4 tr[\mathbf{D}^2]} + \frac{\sigma_1^4}{T^2\sigma_\nu^4}\frac{c_w}{2\sigma_1^4 tr[\mathbf{D}^2]}}} \\ &= \frac{Q_b\sqrt{(T-1)^2 + (T-1)^2 d_b} - Q_w\sqrt{T-1+d_w}}{\sqrt{(T-1)T + (T-1)^2 d_b + d_w}}. \end{aligned}$$

Inserting the quadratic forms in the nominator of Q yields

$$
\begin{aligned}
& Q_b\sqrt{(T-1)^2+(T-1)^2 d_b} - Q_w\sqrt{T-1+d_w} \\
= \; & \frac{\mathbf{u}'(\overline{\mathbf{J}}_T\otimes\mathbf{H})\mathbf{u}-\sigma_1^2 tr[\mathbf{D}]}{\sqrt{2\sigma_1^4 tr[\mathbf{D}^2]+c_b}}\sqrt{(T-1)^2+(T-1)^2 d_b} \\
& -\frac{\mathbf{u}'(\overline{\mathbf{E}}_T\otimes\mathbf{H})\mathbf{u}-\sigma_\nu^2(T-1) tr[\mathbf{D}]}{\sqrt{2(T-1)\sigma_\nu^4 tr[\mathbf{D}^2]+c_w}}\sqrt{T-1+d_w} \\
= \; & \frac{G_{Cb}}{\sqrt{2\sigma_1^4 tr[\mathbf{D}^2]+c_b}}\sqrt{(T-1)^2+(T-1)^2 d_b} - \frac{G_{Cw}}{\sqrt{2(T-1)\sigma_\nu^4 tr[\mathbf{D}^2]+c_w}}\sqrt{T-1+d_w} \\
& -\frac{\sigma_1^2 tr[\mathbf{D}]}{\sqrt{2\sigma_1^4 tr[\mathbf{D}^2]+c_b}}\sqrt{(T-1)^2+(T-1)^2 d_b} - \frac{\sigma_\nu^2(T-1) tr[\mathbf{D}]}{\sqrt{2(T-1)\sigma_\nu^4 tr[\mathbf{D}^2]+c_w}}\sqrt{T-1+d_w} \\
= \; & G_{Cb}\frac{\sqrt{(T-1)^2+(T-1)^2 d_b}}{\sqrt{2\sigma_1^4 tr[\mathbf{D}^2]}\sqrt{1+d_b}} - G_{Cw}\frac{\sqrt{T-1+d_w}}{\sqrt{2\sigma_\nu^4 tr[\mathbf{D}^2]}\sqrt{(T-1)+d_w}} \\
& -\frac{\sigma_1^2 tr[\mathbf{D}](T-1)}{\sqrt{2\sigma_1^4 tr[\mathbf{D}^2]}\sqrt{1+d_b}}\sqrt{1+d_b} + \frac{\sigma_\nu^2(T-1) tr[\mathbf{D}]}{\sqrt{2\sigma_\nu^4 tr[\mathbf{D}^2]}\sqrt{(T-1)+d_w}}\sqrt{T-1+d_w} \\
= \; & G_{Cb}\frac{T-1}{\sqrt{2\sigma_1^4 tr[\mathbf{D}^2]}} - G_{Cb}\frac{1}{\sqrt{2\sigma_\nu^4 tr[\mathbf{D}^2]}},
\end{aligned}
$$

where we define $G_{Cb}=\mathbf{u}'(\overline{\mathbf{J}}_T\otimes\mathbf{H})\mathbf{u}$ and $G_{Cw}=\mathbf{u}'(\overline{\mathbf{E}}_T\otimes\mathbf{H})\mathbf{u}$.

Remember the denominator is given by

$$\sqrt{(T-1)T+(T-1)^2 d_b+d_w}.$$

The test can then be based on

$$
\begin{aligned}
\sqrt{LM_{C,robust}} &= \frac{G_{Cb}\frac{T-1}{\sqrt{2\sigma_1^4 tr[\mathbf{D}^2]}} - G_{Cw}\frac{1}{\sqrt{2\sigma_\nu^4 tr[\mathbf{D}^2]}}}{\sqrt{(T-1)T+(T-1)^2 d_b+d_w}} \\
&= \frac{G_{Cb}(T-1) - G_{Cw}\frac{\sigma_1^2}{\sigma_\nu^2}}{\sqrt{2\sigma_1^4 tr[\mathbf{D}^2]}\sqrt{(T-1)T+(T-1)^2 d_b+d_w}}.
\end{aligned}
$$

Under normality the test statistic is given by

$$\sqrt{LM_C} = \frac{G_{Cb}(T-1) - G_{Cw}\frac{\sigma_1^2}{\sigma_\nu^2}}{\sqrt{2\sigma_1^4 tr[\mathbf{D}^2]}\sqrt{(T-1)\,T}}.$$

Observe that $\overline{\mathbf{u}} - \mathbf{u} = -\mathbf{X}\left(\bar{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)$, where $\mathbf{u} = (\boldsymbol{\iota}_T \otimes \mathbf{A}^{-1})\boldsymbol{\mu} + (\mathbf{I}_T \otimes \mathbf{A}^{-1})\boldsymbol{\nu}$ and

$$(NT)^{-1/2}\overline{\mathbf{u}}'(\overline{\mathbf{J}}_T \otimes \overline{\mathbf{H}})\overline{\mathbf{u}} = (NT)^{-1/2}\overline{\mathbf{u}}'(\overline{\mathbf{J}}_T \otimes (\mathbf{W} + \mathbf{W}' - 2\bar{\rho}\mathbf{W}'\mathbf{W}))\overline{\mathbf{u}} := \overline{Q}_{bC1} - 2\rho\overline{Q}_{bC2}.$$

Following Kelejian and Prucha (2001, Lemma 1), one obtains

$$\begin{aligned} \overline{Q}_{bC1} &= (NT)^{-1/2}\mathbf{u}'(\overline{\mathbf{J}}_T \otimes (\mathbf{W} + \mathbf{W}'))\mathbf{u} + o_p(1) \\ \overline{Q}_{bC2} &= (NT)^{-1/2}\mathbf{u}'(\overline{\mathbf{J}}_T \otimes \mathbf{W}'\mathbf{W})\mathbf{u} + o_p(1). \end{aligned}$$

Notice that

$$2\bar{\rho}\overline{Q}_{bC2} - 2\rho Q_{bC2} = 2\left(\bar{\rho} - \rho\right)\overline{Q}_{bC2} - 2\rho(\overline{Q}_{bC2} - Q_{bC2}) = o_p(1).$$

The last equality follows since $\bar{\rho}$ is a consistent estimator and $\overline{Q}_{bC2} = O_p(1)$ by Lemma 4, after setting $\mathbf{H} = \mathbf{W}'\mathbf{W}$ and $\alpha = 1$. Therefore,

$$\overline{Q}_{bC1} - Q_{bC1} + 2\bar{\rho}\overline{Q}_{bC2} - 2\rho Q_{bC2} = o_p(1).$$

Defining $Q_{bC} = Q_{bC1} - 2\rho Q_{bC1}$, we obtain $\overline{Q}_{bC} - Q_{bC} = o_p(1)$.and $(NT)^{-1/2}\overline{\mathbf{u}}'(\overline{\mathbf{J}}_T \otimes \overline{\mathbf{H}})\overline{\mathbf{u}} - (NT)^{-1/2}\mathbf{u}'(\overline{\mathbf{J}}_T \otimes \mathbf{H})\mathbf{u} = o_p(1)$. Now,

$$E[\mathbf{u}'(\overline{\mathbf{J}}_T \otimes \mathbf{H})\mathbf{u}] = \sigma_1^2 tr\left(\mathbf{H}(\mathbf{A}'\mathbf{A})^{-1}\right) = \sigma_1^2 tr\left(\mathbf{H}(\mathbf{A}'\mathbf{A})^{-1}\right)$$

using $tr\left(\mathbf{H}(\mathbf{A}'\mathbf{A})^{-1}\right) = tr\left(\mathbf{A}'^{-1}(\mathbf{W}'\mathbf{A} + \mathbf{A}'\mathbf{W})\mathbf{A}^{-1}\right) = tr(\mathbf{A}'^{-1}\mathbf{W}' + \mathbf{W}\mathbf{A}^{-1})$.Similarly, $(NT)^{-1/2}\overline{\mathbf{u}}'(\mathbf{E}_T \otimes \overline{\mathbf{H}})\overline{\mathbf{u}} - (NT)^{-1/2}\mathbf{u}'(\mathbf{E}_T \otimes \mathbf{H})\mathbf{u} = o_p(1)$. Defining $Q_{wC} = \frac{\sqrt{T-1}}{T}\mathbf{u}'(\mathbf{E}_T \otimes \mathbf{H})\mathbf{u}$, we also have $\overline{Q}_{wC} - Q_{wC} = o_p(1)$. Also, the two quadratic forms Q_{bC} and Q_{wC} are independent by Lemma 4. As a result we obtain

$$\sqrt{\overline{LM}_{C,robust}} = \frac{(T-1)\overline{G}_{Cb} - \frac{\overline{\sigma}_1^2}{\overline{\sigma}_\nu^2}\overline{G}_{Cw}}{\sqrt{2\overline{\sigma}_1^4 tr[\overline{\mathbf{D}}^2]}\sqrt{(T-1)\,T}}\sqrt{\frac{1}{1+(d_b+d_w')\,(\frac{T-1}{T})}}$$

$$= \frac{T\left(\overline{G}_{Cb} - \overline{\sigma}_1^2 tr[\overline{\mathbf{D}}]\right)}{\sqrt{2\overline{\sigma}_1^4 tr[\overline{\mathbf{D}}^2]}\sqrt{(T-1)\,T}}\sqrt{\frac{1}{1+\left(\overline{d}_b+\overline{d}_w'\right)(\frac{T-1}{T})}}$$

$$= \sqrt{\overline{LM}_C}\sqrt{\frac{1}{1+(d_b+d_w')\,(\frac{T-1}{T})}}$$

using $\overline{\sigma}_1^2 tr[\overline{\mathbf{D}}]{=}\frac{1}{T}\overline{\mathbf{u}}'(\overline{\mathbf{J}}_T \otimes \overline{\mathbf{H}})\overline{\mathbf{u}}{+}\frac{\overline{\sigma}_1^2}{T\overline{\sigma}_\nu^2}\overline{\mathbf{u}}'(\mathbf{E}_T \otimes \overline{\mathbf{H}})\overline{\mathbf{u}}$.

$$\sqrt{\overline{LM}_{C,robust}} = \frac{\sqrt{(T-1)^2+(T-1)^2 d_b + T - 1 + d_w}}{\sqrt{(T-1)^2+(T-1)^2\overline{d}_b + T - 1 + \overline{d}_w}} * \Bigg\{ \frac{(NT)^{-1/2}\sigma_1^2\sqrt{2tr[\mathbf{D}^2]}}{(NT)^{-1/2}\overline{\sigma}_1^2\sqrt{2tr[\overline{\mathbf{D}}^2]}} \cdot \left(\frac{(NT)^{-1/2}\,G_{Cb}+o_p(1)}{(NT)^{-1/2}\sigma_1^2\sqrt{2tr[\mathbf{D}^2]}\sqrt{(T-1)^2+(T-1)^2 d_b+T-1+d_w}}\right)(T-1)$$
$$- \begin{array}{c} \frac{(NT)^{-1/2}\sigma_\nu^2\sqrt{2tr[\mathbf{D}^2]}}{(NT)^{-1/2}\overline{\sigma}_\nu^2\sqrt{2tr[\overline{\mathbf{D}}^2]}} \cdot \\ \left(\frac{(NT)^{-1/2}G_{Cw}+o_p(1)}{(NT)^{-1/2}\sigma_\nu^2\sqrt{2tr[\mathbf{D}^2]}\sqrt{(T-1)^2+(T-1)^2 d_b+T-1+d_w}}\right)\end{array} \Bigg\}.$$

Notice that $\overline{\sigma}_1^2 = \sigma_1^2 + o_p(1)$, $\overline{\sigma}_\nu^2 = \sigma_\nu^2 + o_p(1)$ and $\sigma_1^2 > 0$ and $\sigma_\nu^2 > 0$ by Assumption A1. Using $\mathbf{H} = \mathbf{F} = (\mathbf{W}'\mathbf{A} + \mathbf{A}'\mathbf{W}) = \mathbf{W}' + \mathbf{W} - 2\rho\mathbf{W}'\mathbf{W})$ in Lemma 4, we conclude that $(NT)^{-1}\sigma_1^2\,(2tr[\mathbf{D}^2])$ and $(NT)^{-1}\sigma_\nu^4 2(tr[\mathbf{D}^2])$ are bounded away from zero by some positive constants. Furthermore, since $\overline{\rho} - \rho = o_p(1)$ we have $plim_{N\to\infty}\frac{(NT)^{-1/2}\sigma_1^2\sqrt{2tr[\mathbf{D}^2]}}{(NT)^{-1/2}\overline{\sigma}_1^2\sqrt{2tr[\overline{\mathbf{D}}^2]}} = 1$ and $plim_{N\to\infty}\frac{(NT)^{-1/2}\sigma_\nu^2\sqrt{2tr[\mathbf{D}^2]}}{(NT)^{-1/2}\overline{\sigma}_\nu^2\sqrt{2tr[\overline{\mathbf{D}}^2]}} = 1$. Assumptions A1 and A6 imply that $\sqrt{(T-1)^2+(T-1)^2 d_b + T - 1 + d_w}$ is bounded way from zero by some positive constant. With the higher moments being estimated consistently, $plim_{N\to\infty}\frac{\sqrt{(T-1)^2+(T-1)^2 d_b+T-1+d_w}}{\sqrt{(T-1)^2+(T-1)^2\overline{d}_b+T-1+\overline{d}_w}} = 1$, it follows that $\overline{LM}_{C,robust} - LM_{C,robust} = o_p(1)$. ■

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.