



Munich Personal RePEc Archive

## **GMM Gradient Tests for Spatial Dynamic Panel Data Models**

Taspinar, Suleyman and Dogan, Osman and Bera, Anil K.

2017

Online at <https://mpra.ub.uni-muenchen.de/83570/>  
MPRA Paper No. 83570, posted 02 Jan 2018 23:00 UTC

# GMM Gradient Tests for Spatial Dynamic Panel Data Models \*

Süleyman Taşpınar<sup>†</sup>      Osman Doğan<sup>‡</sup>      Anil K. Bera<sup>§</sup>

September 29, 2016

## Abstract

In this study, we formulate the adjusted gradient tests when the alternative model used to construct tests deviates from the true data generating process for a spatial dynamic panel data model (SDPD). Following Bera et al. (2010), we introduce these adjusted gradient tests along with the standard ones within a GMM framework. These tests can be used to detect the presence of (i) the contemporaneous spatial lag terms, (ii) the time lag term, and (iii) the spatial time lag terms in an higher order SDPD model. These adjusted tests have two advantages: (i) their null asymptotic distribution is a central chi-squared distribution irrespective of the misspecified alternative model, and (ii) their test statistics are computationally simple and require only the ordinary least-squares (OLS) estimates from a non-spatial two-way panel data model. We investigate the finite sample size and power properties of these tests through Monte Carlo studies. Our results indicates that the adjusted gradient tests have good finite sample properties.

JEL-Classification: C13, C21, C31.

Keywords: Spatial Dynamic Panel Data Model, SDPD, GMM, Robust LM Tests, GMM Gradient Tests, Inference.

---

\*This research was supported, in part, by a grant of computer time from the City University of New York High Performance Computing Center under NSF Grants CNS-0855217 and CNS-0958379. Please address all correspondence to Süleyman Taşpınar at STaspınar@qc.cuny.edu.

<sup>†</sup>Economics Program, Queens College, The City University of New York, United States, email: staspınar@qc.cuny.edu.

<sup>‡</sup>Istanbul Ulasim A.S., Project Department, Istanbul, Turkey., email: odogan10@gmail.com.

<sup>§</sup>Economics Program, University of Illinois, Illinois, United States, email: a-bera@uiuc.edu.

# 1 Introduction

In this study, we consider a spatial dynamic panel data model (SDPD) that includes a time lag term, spatial time lag terms and contemporaneous spatial lag terms. The model is in the form of a high order spatial autoregressive model by including high orders of contemporaneous spatial lag term and spatial time lag term. We formulate the GMM gradient tests, the adjusted GMM gradient tests and the  $C(\alpha)$  test to test hypothesis about the parameters of the time lag term, the spatial time lag terms and the contemporaneous spatial lag terms.

In the literature, the model specifications and estimation strategies, including the ML, GMM and Bayesian methods, receive considerably more attention than the specification testing and other forms of hypothesis tests for the SDPD models. For two recent surveys, see Anselin et al. (2008) and Lee and Yu (2010b). Lee and Yu (2010a, 2011, 2012a), Yu and Lee (2010), and Yu et al. (2008, 2012) consider the ML approach for dynamic spatial panel data models when both the number of individuals and the number of time periods are large under various scenarios. The MLE suggested in these studies has asymptotic bias and the limiting distributions of bias corrected versions are properly centered when the number of time periods grows faster than the number of individuals. Elhorst (2005), Lee and Yu (2015), and Su and Yang (2015) consider the ML approach for the dynamic panel data models that have spatial autoregressive processes in the disturbance terms. Parent and LeSage (2011) introduce the Bayesian MCMC method for a panel data model that accommodates dependence across space and time in the error components. Kapoor et al. (2007) extend the GMM approach of Kelejian and Prucha (2010) to a static spatial panel data model with error components. Lee and Yu (2014) consider the GMM approach for an SDPD model that has high orders of contemporaneous spatial lag term and spatial time lag term.

To date, the focus has been on the specification testing for the cross-sectional and the static spatial panel data models (Anselin et al. 1996; Baltagi and Yang 2013; Baltagi et al. 2003, 2007; Debarsy and Ertur 2010). In this study, we introduce GMM-based tests for an SDPD model that has high orders of contemporaneous spatial lag term and spatial time lag term. In particular, we first consider the GMM-gradient test (or the LM test) of Newey and West (1987), which can be used to test the non-linear restrictions on the parameter vector. We also consider the  $C(\alpha)$  test within the GMM framework for the same model. While the computation of GMM-gradient test requires an estimate of the optimal restricted GMME, the computation of  $C(\alpha)$  test statistic requires only a consistent estimate of the parameter vector. For both tests, we provide analytical justification for their asymptotic distributions within the context of our SDPD.

Within the ML framework, Davidson and MacKinnon (1987), Saikkonen (1989) and Bera and Yoon (1993) show that the usual LM tests are not robust to local mis-specifications in the alternative models. That is, the usual LM tests have non-central chi-squared distribution when the alternative model (locally) deviates from the true data generating process. Bera et al. (2010) extend this result to the GMM framework and show that the asymptotic distribution of the usual GMM-gradient test is a non-central chi-squared distribution when the alternative model deviates from the true data generating process. In such a context, the usual LM and GMM-gradient tests will over reject the true null hypothesis. Therefore, Bera and Yoon (1993) and Bera et al. (2010) suggest robust (or adjusted) versions that have, asymptotically, central chi-squared distributions irrespective of the local deviations of the alternative models from the true data generating process.

By following Bera et al. (2010), we construct various adjusted GMM-gradient tests for an SDPD model. These tests can be used to detect the presence of (i) the spatial lag terms, (ii) the time lag term, and (iii) the spatial time lag terms in an SDPD model. Besides being robust to local mis-specifications, these tests are computationally simple and require only estimates from a non-spatial two-way panel data model. Within the context of our SDPD, we analytically show the asymptotic

66 distribution of robust tests under both the null and local alternative hypotheses. We investigate  
 67 the size and power properties of our suggested robust tests through a Monte Carlo simulation. The  
 68 simulation results are in line with our theoretical findings and indicate that the robust tests have  
 good size and power properties.

70 The rest of this paper is organized in the following way. Section 2 presents the SDPD model  
 under consideration and discusses its assumptions. Section 3 lays out the details of the GMM  
 72 estimation approach for the model specification. Section 4 presents the GMM gradient tests, the  
 adjusted GMM gradient tests and the  $C(\alpha)$  test. Section 5 lays out the details of the Monte Carlo  
 74 design and presents the results. Section 6 closes with concluding remarks. Some of the technical  
 derivations are relegated to an appendix.

## 76 2 The Model Specification and Assumptions

Using the standard notation, an SDPD model with both individual and time fixed effects is stated  
 as

$$Y_{nt} = \sum_{j=1}^p \lambda_{j0} W_{nj} Y_{nt} + \gamma_0 Y_{n,t-1} + \sum_{j=1}^p \rho_{j0} W_{nj} Y_{n,t-1} + X_{nt} \beta_0 + \mathbf{c}_{n0} + \alpha_{t0} l_n + V_{nt} \quad (2.1)$$

for  $t = 1, 2, \dots, T$ , where  $Y_{nt} = (y_{1t}, y_{2t}, \dots, y_{nt})'$  is the  $n \times 1$  vector of a dependent variable,  $X_n$   
 78 is the  $n \times k_x$  matrix of non-stochastic exogenous variables with a matching parameter vector  $\beta_0$ ,  
 and  $V_{nt} = (v_{1t}, \dots, v_{nt})'$  is the  $n \times 1$  vector of disturbances (or innovations). The spatial lags of  
 80 the dependent variable at time  $t$  and  $t - 1$  are, respectively, denoted by  $W_{nj} Y_{nt}$  and  $W_{nj} Y_{n,t-1}$   
 for  $j = 1, \dots, p$ . Here,  $W_{nj}$ s are the  $n \times n$  spatial weight matrices of known constants with  
 82 zero diagonal elements,  $\lambda_0 = (\lambda_{10}, \dots, \lambda_{p0})'$  and  $\rho_0 = (\rho_{10}, \dots, \rho_{p0})'$  are the spatial autoregressive  
 parameters. The individual fixed effects are denoted by  $\mathbf{c}_{n0} = (c_{1,0}, \dots, c_{n,0})'$  and the time fixed  
 84 effect is denoted by  $\alpha_{t0} l_n$ , where  $l_n$  is the  $n \times 1$  vectors of ones. For the identification of fixed  
 effects, Lee and Yu (2014) impose the normalization  $l_n' \mathbf{c}_{n0} = 0$ . For the estimation of the model,  
 86 we assume that  $Y_{n0}$  is observable. Let  $\Theta$  be the parameter space of the model. In order to  
 distinguish the true parameter vector from other possible values in  $\Theta$ , we state the model with  
 88 the true parameter vector  $\theta_0 = (\lambda_0', \delta_0')$ , where  $\delta_0 = (\gamma_0, \rho_0', \beta_0')$ . Furthermore, for notational  
 simplicity we let  $S_n(\lambda) = (I_n - \sum_{j=1}^p \lambda_j W_{nj})$ ,  $S_n = S_n(\lambda_0)$ ,  $A_n = S_n^{-1}(\gamma_0 I_n + \sum_{j=1}^p \rho_j W_{nj})$ ,  
 90  $G_{nj}(\lambda) = W_{nj} S_n^{-1}(\lambda)$ ,  $G_{nj} = G_{nj}(\lambda_0)$  and  $N = n(T - 1)$ .

To avoid the incidental parameter problem, the model is transformed to wipe out the  
 fixed effects. The individual effects can be eliminated from the model by employing the or-  
 thonormal eigenvector matrix  $[F_{T,T-1}, \frac{1}{\sqrt{T}} l_T]$  of  $J_T = (I_T - \frac{1}{T} l_T l_T')$ , where  $F_{T,T-1}$  is the  
 $T \times (T - 1)$  eigenvectors matrix corresponding to the eigenvalue one and  $l_T$  is the  $T \times 1$  vec-  
 tor of ones corresponding to the eigenvalue zero.<sup>1</sup> This orthonormal transformation can be  
 applied by writing the model in an  $n \times T$  system. Hence, the dependent variable is trans-  
 formed as  $[Y_{n1}, Y_{n2}, \dots, Y_{nT}] \times F_{T,T-1} = [Y_{n1}^*, Y_{n2}^*, \dots, Y_{n,T-1}^*]$ , and also  $[Y_{n0}, Y_{n1}, \dots, Y_{n,T-1}] \times$   
 $F_{T,T-1} = [Y_{n0}^{(*,-1)}, Y_{n1}^{(*,-1)}, \dots, Y_{n,T-2}^{(*,-1)}]$ . Similarly,  $[X_{nj,1}, X_{nj,2}, \dots, X_{nj,T}] \times F_{T,T-1} =$   
 $[X_{nj,1}^*, X_{nj,2}^*, \dots, X_{nj,T-1}^*]$  for  $j = 1, \dots, k_x$ ,  $[V_{n1}, V_{n2}, \dots, V_{nT}] \times F_{T,T-1} = [V_{n1}^*, V_{n2}^*, \dots, V_{n,T-1}^*]$ ,  
 and  $[\alpha_{10}, \alpha_{20}, \dots, \alpha_{T0}] \times F_{T,T-1} = [\alpha_{10}^*, \alpha_{20}^*, \dots, \alpha_{T-1,0}^*]$ . Since the column of  $[F_{T,T-1}, \frac{1}{\sqrt{T}} l_T]$  are  
 orthonormal, we have  $[\mathbf{c}_{n0}, \mathbf{c}_{n0}, \dots, \mathbf{c}_{n0}] \times F_{T,T-1} = \mathbf{0}_{n \times (T-1)}$ . Thus, the transformed model does

<sup>1</sup>This orthonormal matrix has the following properties (i)  $J_T F_{T,T-1} = F_{T,T-1}$  and  $J_T l_T = \mathbf{0}_{T \times 1}$ , (ii)  
 $F_{T,T-1}' F_{T,T-1} = I_{T-1}$  and  $F_{T,T-1}' l_T = \mathbf{0}_{(T-1) \times 1}$ , (iii)  $F_{T,T-1} F_{T,T-1}' + \frac{1}{T} l_T l_T' = I_T$  and (iv)  $F_{T,T-1} F_{T,T-1}' = J_T$ .

not include the individual fixed effects and can be written as

$$Y_{nt}^* = \sum_{j=1}^p \lambda_{j0} W_{nj} Y_{nt}^* + \gamma_0 Y_{n,t-1}^{(*,-1)} + \sum_{j=1}^p \rho_{j0} W_{nj} Y_{n,t-1}^{(*,-1)} + X_{nt}^* \beta_0 + \alpha_{t0}^* l_n + V_{nt}^* \quad (2.2)$$

for  $t = 1, \dots, T-1$ . We consider the forward orthogonal difference (FOD) transformation for the orthonormal transformation. Hence, the terms in (2.2) can be explicitly stated as  $V_{nt}^* = \left(\frac{T-t}{T-t+1}\right)^{1/2} [V_{nt} - \frac{1}{T-t} \sum_{h=t+1}^T V_{nh}]$ ,  $Y_{n,t-1}^{(*,-1)} = \left(\frac{T-t}{T-t+1}\right)^{1/2} [Y_{n,t-1} - \frac{1}{T-t} \sum_{h=t}^{T-1} Y_{nh}]$ , and the others terms are defined similarly. Let  $\mathbf{V}_{n,T-1}^* = (V_{n1}^*, \dots, V_{n,T-1}^*)'$ . Then,  $\text{Var}(\mathbf{V}_{n,T-1}^*) = (F'_{T,T-1} \otimes I_n) \text{E}(\mathbf{V}_{nT} \mathbf{V}'_{nT}) (F_{T,T-1} \otimes I_n) = \sigma_0^2 I_N$  by Assumption 1. The transformed model in (2.2) still includes the time fixed effect  $\alpha_{t0}^* l_n$ , which can be eliminated by pre-multiplying the model with  $J_n = I_n - \frac{1}{n} l_n l_n'$ . The resulting model is free of the fixed effects, for  $t = 1, \dots, T-1$ ,

$$J_n Y_{nt}^* = \sum_{j=1}^p \lambda_{j0} J_n W_{nj} Y_{nt}^* + \gamma_0 J_n Y_{n,t-1}^{(*,-1)} + \sum_{j=1}^p \rho_{j0} J_n W_{nj} Y_{n,t-1}^{(*,-1)} + J_n X_{nt}^* \beta_0 + J_n V_{nt}^*. \quad (2.3)$$

The consistency and asymptotic normality of the GMME of  $\theta_0$  are established under Assumptions 1 through 5.<sup>2</sup>

**Assumption 1.** — The innovations  $v_{it}$ s are independently and identically distributed across  $i$  and  $t$ , and satisfy  $\text{E}(v_{it}) = 0$ ,  $\text{E}(v_{it}^2) = \sigma_0^2$ , and  $\text{E}|v_{it}|^{4+\eta} < \infty$  for some  $\eta > 0$  for all  $i$  and  $t$ .

**Assumption 2.** — The spatial weight matrix  $W_{nj}$ s is uniformly bounded in row and column sums in absolute value for  $j = 1, \dots, p$ , and  $\|\sum_{j=1}^p \lambda_{j0} W_{nj}\|_\infty < 1$ . Moreover,  $S_n^{-1}(\lambda)$  exists and is uniformly bounded in row and column sums in absolute value for all values of  $\lambda$  in a compact parameter space.

**Assumption 3.** — Let  $\eta > 0$  be a real number. Assume that  $X_{nt}$ ,  $\mathbf{c}_{n0}$ , and  $\alpha_{t0}$  are non-stochastic terms satisfying (i)  $\sup_{n,T} \frac{1}{nT} \sum_{t=1}^T \sum_{i=1}^n |x_{it,l}|^{2+\eta} < \infty$  for  $l = 1, \dots, k_x$ , where  $x_{it,l}$  is the  $(i, t)$ th element of the  $l$ th column, (ii)  $\lim_{n \rightarrow \infty} \frac{1}{n(T-1)} \sum_{t=1}^{T-1} X_{nt}^* J_n X_{nt}^*$  exists and is non-singular, and (iii)  $\sup_T \frac{1}{T} \sum_{t=1}^T |\alpha_{t0}|^{2+\eta} < \infty$  and  $\sup_n \frac{1}{n} \sum_{i=1}^n |c_{i0}|^{2+\eta} < \infty$ .

**Assumption 4.** — The DGP for the initial observations is  $Y_{n0} = \sum_{h=0}^{h^*} A_n^h S_n^{-1}(\mathbf{c}_{n0} + X_{n,-h} \beta_0 + \alpha_{-h,0} l_n + V_{n,-h})$ , where  $h^*$  could be finite or infinite.

**Assumption 5.** — The elements of  $\sum_{h=0}^\infty \text{abs}(A_n^h)$  are uniformly bounded in row and column sums in absolute value, where  $[\text{abs}(A_n)]_{ij} = |A_{n,ij}|$

### 3 The GMM Estimation Approach

In this section, we summarize the GMM estimation approach for (2.3) under both large  $T$  and finite  $T$  scenarios. The model in (2.3) indicates that IVs are needed for  $W_{nj} Y_{nt}^*$ ,  $Y_{n,t-1}^{(*,-1)}$ , and  $W_{nj} Y_{n,t-1}^{(*,-1)}$  for each  $t$ . Before, we introduce the set of moment functions, it will be convenient to introduce some further notations. Let  $Z_{nt}^* = [Y_{n,t-1}^{(*,-1)}, W_{n1} Y_{n,t-1}^{(*,-1)}, \dots, W_{np} Y_{n,t-1}^{(*,-1)}, X_{nt}^*]$ ,  $\mathbf{J}_{n,T-1} = I_{T-1} \otimes J_n$ , and  $\mathbf{V}_{n,T-1}^*(\theta) = (V_{n1}^*(\theta), \dots, V_{n,T-1}^*(\theta))'$  where  $V_{nt}^*(\theta) = S_{nt}(\lambda) Y_{nt}^* - Z_{nt}^* \delta - \alpha_t^* l_n$ . We consider the

<sup>2</sup>For interpretations and implications of these assumptions, see Lee and Yu (2014) and Kelejian and Prucha (2010).

following  $(m + q) \times 1$  vector of moment functions

$$g_{nT}(\theta) = \begin{pmatrix} \mathbf{V}_{n,T-1}^{*\prime}(\theta) \mathbf{J}_{n,T-1} \mathbf{P}_{n1,T-1} \mathbf{J}_{n,T-1} \mathbf{V}_{n,T-1}^*(\theta) \\ \mathbf{V}_{n,T-1}^{*\prime}(\theta) \mathbf{J}_{n,T-1} \mathbf{P}_{n2,T-1} \mathbf{J}_{n,T-1} \mathbf{V}_{n,T-1}^*(\theta) \\ \vdots \\ \mathbf{V}_{n,T-1}^{*\prime}(\theta) \mathbf{J}_{n,T-1} \mathbf{P}_{nm,T-1} \mathbf{J}_{n,T-1} \mathbf{V}_{n,T-1}^*(\theta) \\ \mathbf{Q}'_{n,T-1} \mathbf{J}_{n,T-1} \mathbf{V}_{n,T-1}^*(\theta) \end{pmatrix}. \quad (3.1)$$

108 In (3.1),  $\mathbf{P}_{nj,T-1} = I_{T-1} \otimes P_{nj}$ , where  $P_{nj}$  is the  $n \times n$  quadratic moment matrix satisfying  $\text{tr}(P_{nj} \mathbf{J}_n) = 0$  for  $j = 1, \dots, m$ , and  $\mathbf{Q}_{n,T-1} = (Q'_{n1}, \dots, Q'_{n,T-1})'$  is the  $N \times q$  linear IV matrix such  
110 that  $q \geq k_x + 2p + 1$ . Under Assumptions 1-4, it can be shown that  $\frac{1}{N} \frac{\partial g_{nT}(\theta_0)}{\partial \theta'} = D_{nT} + R_{nT} + O\left(\frac{1}{\sqrt{nT}}\right)$ , where  $D_{nT}$  is  $O(1)$  and  $R_{nT}$  is  $O\left(\frac{1}{T}\right)$ .<sup>3</sup>

Let  $\text{vec}_D(\cdot)$  be the operator that creates a column vector from the diagonal elements of an input square matrix. For the optimal GMM estimation, we need to calculate the covariance matrix of moment functions  $E(g'_{nT}(\theta_0) g_{nT}(\theta_0))$ , which can be approximated by

$$\begin{aligned} \Sigma_{nT} = \sigma_0^4 & \begin{pmatrix} \frac{1}{N} \Delta_{nm,T} & 0_{m \times q} \\ 0_{q \times m} & \frac{1}{\sigma_0^2} \frac{1}{N} \mathbf{Q}'_{n,T-1} \mathbf{J}_{n,T-1} \mathbf{Q}_{n,T-1} \end{pmatrix} \\ & + \frac{1}{N} \begin{pmatrix} (\mu_4 - 3\sigma_0^4) \omega'_{nm,T} \omega_{nm,T} & 0_{m \times q} \\ 0_{q \times m} & 0_{q \times m} \end{pmatrix}, \end{aligned} \quad (3.2)$$

112 where  $\omega_{nm,T} = [\text{vec}_D(\mathbf{J}_{n,T-1} \mathbf{P}_{n1,T-1} \mathbf{J}_{n,T-1}), \dots, \text{vec}_D(\mathbf{J}_{n,T-1} \mathbf{P}_{nm,T-1} \mathbf{J}_{n,T-1})]$ ,  
 $\Delta_{nm,T} = [\text{vec}(\mathbf{J}_{n,T-1} \mathbf{P}'_{n1,T-1} \mathbf{J}_{n,T-1}), \dots, \text{vec}(\mathbf{J}_{n,T-1} \mathbf{P}'_{nm,T-1} \mathbf{J}_{n,T-1})]' \times$   
114  $[\text{vec}(\mathbf{J}_{n,T-1} \mathbf{P}^s_{n1,T-1} \mathbf{J}_{n,T-1}), \dots, \text{vec}(\mathbf{J}_{n,T-1} \mathbf{P}^s_{nm,T-1} \mathbf{J}_{n,T-1})]$ , where  $A_n^s = A_n + A'_n$  for any square matrix  $A_n$ .

Let  $\widehat{\Sigma}_{nT}$  be a consistent estimate of  $\Sigma_{nT}$ . Then, the optimal GMME is defined by

$$\widehat{\theta}_{nT} = \underset{\theta \in \Theta}{\text{argmin}} g'_{nT}(\theta) \widehat{\Sigma}_{nT}^{-1} g_{nT}(\theta) \quad (3.3)$$

Under Assumptions 1 - 5, Lee and Yu (2014) show that when both  $T$  and  $n$  tend to infinity<sup>4</sup>:

$$\sqrt{N}(\widehat{\theta}_{nT} - \theta_0) \xrightarrow{d} N\left(0, \left[\text{plim}_{n,T \rightarrow \infty} D'_{nT} \Sigma_{nT}^{-1} D_{nT}\right]^{-1}\right). \quad (3.4)$$

116 When  $T$  is finite, the GMME in (3.4) is still consistent and unbiased but its limiting covariance matrix is different, since the additional term  $R_{nT} = O\left(\frac{1}{T}\right)$  does not vanish. Hence, when  $T$  is finite,  
118 the asymptotic covariance matrix of  $\sqrt{N}(\widehat{\theta}_{nT} - \theta_0)$  is given by  $[\text{plim}_{n \rightarrow \infty} (D_{nT} + R_{nT})' \Sigma_{nT}^{-1} (D_{nT} + R_{nT})]^{-1}$ .

<sup>3</sup>The explicit forms for  $D_{nT}$  and  $R_{nT}$  are not required for our testing results, hence they are not given here. For these terms, see Lee and Yu (2014).

<sup>4</sup> Lee and Yu (2014) state the identification conditions. Here, we simply assume that the parameter vector is identified.

120 **4 The GMM Gradient Tests**

In this section, we consider various version of the gradient test (LM test). Let  $r : \mathbb{R}^{2p+k_x+1} \rightarrow \mathbb{R}^{k_r}$  be a twice continuously differentiable function, and assume that  $R(\theta) = \frac{\partial r(\theta)}{\partial \theta'}$  has rank  $k_r$ . Consider the implicit restrictions denoted by the null hypothesis  $H_0 : r(\theta_0) = 0$ . Define  $\widehat{\theta}_{nT,r} = \operatorname{argmax}_{\{\theta:r(\theta)=0\}} \mathcal{Q}_n$ , where  $\mathcal{Q}_n = g'_{nT}(\theta) \widehat{\Sigma}_{nT}^{-1} g_{nT}(\theta)$ , as a restricted (or constrained) optimal GMME.

In order to give a general argument, consider the following partition of  $\theta = (\beta', \psi', \phi')'$ , where  $\psi$  and  $\phi$  are, respectively,  $k_\psi \times 1$  and  $k_\phi \times 1$  vectors such that  $k_\psi + k_\phi = 2p + 1$ . In the context of our model,  $\psi$  and  $\phi$  can be any combinations of the remaining parameters, namely,  $(\lambda', \gamma, \rho')'$ . Let  $G_a = \frac{1}{N} \frac{\partial g_{nT}(\theta)}{\partial a'}$ ,  $C_a = G'_a(\theta) \widehat{\Sigma}_{nT}^{-1} \bar{g}_{nT}(\theta)$ , where  $a \in \{\beta, \psi, \phi\}$  and  $\bar{g}_{nT} = \frac{1}{N} g_{nT}$ . Define  $G(\theta) = (G_\beta(\theta), G_\psi(\theta), G_\phi(\theta))$ , and  $C(\theta) = (C'_\beta(\theta), C'_\psi(\theta), C'_\phi(\theta))'$ , and  $B(\theta) = G'(\theta) \widehat{\Sigma}_{nT}^{-1} G(\theta)$ . Finally, let  $\mathcal{G}_a = \operatorname{plim}_{n,T \rightarrow \infty} \frac{1}{N} \frac{\partial g_{nT}(\theta_0)}{\partial a'}$  for  $a \in \{\beta, \psi, \phi\}$ . Define  $\mathcal{G} = (\mathcal{G}_\beta, \mathcal{G}_\psi, \mathcal{G}_\phi)$  and  $\mathcal{H} = \operatorname{plim}_{n,T \rightarrow \infty} (D_{nT} + R_{nT})' \widehat{\Sigma}_{nT}^{-1} (D_{nT} + R_{nT})$ . We consider the following partition of  $B(\theta)$  and  $\mathcal{H}$ :

$$B(\theta) = \begin{pmatrix} B_\beta(\theta) & B_{\beta\psi}(\theta) & B_{\beta\phi}(\theta) \\ B_{\psi\beta}(\theta) & B_\psi(\theta) & B_{\psi\phi}(\theta) \\ B_{\phi\beta}(\theta) & B_{\phi\psi}(\theta) & B_\phi(\theta) \end{pmatrix}, \quad \mathcal{H} = \begin{pmatrix} \mathcal{H}_\beta & \mathcal{H}_{\beta\psi} & \mathcal{H}_{\beta\phi} \\ \mathcal{H}_{\psi\beta} & \mathcal{H}_\psi & \mathcal{H}_{\psi\phi} \\ \mathcal{H}_{\phi\beta} & \mathcal{H}_{\phi\psi} & \mathcal{H}_\phi \end{pmatrix}. \quad (4.1)$$

With the notation introduced, the standard LM test statistic for  $H_0 : r(\theta_0) = 0$  is defined in the following way (Newey and West 1987):

$$LM = N C'(\widehat{\theta}_{nT,r}) B^{-1}(\widehat{\theta}_{nT,r}) C(\widehat{\theta}_{nT,r}). \quad (4.2)$$

A similar test is the  $C(\alpha)$  test.<sup>5</sup> This test is designed to deal with the nuisance parameters when testing the parameter of main interest (Bera and Biliias 2001). Lee and Yu (2012b) investigate the finite sample properties of this test for a cross-sectional autoregressive model. Their simulation results indicate that this test can be useful to test the possible presence of spatial correlation through a spatial lag in the spatial autoregressive (SAR) model. Here, we provide a general description of this test within the context of our SDPD model. By the implicit function theorem, the set of  $k_r$  restrictions on  $\theta_0$  can also be stated as  $h(\xi_0) = \theta_0$ , where  $h : \mathbb{R}^{\bar{q}} \rightarrow \mathbb{R}^{2p+k_x+1}$  is continuously differentiable,  $\xi_0$  contains the free parameters, and  $\bar{q} = 2p + k_x + 1 - k_r$ . Define  $\widehat{\xi}_{nT} = \operatorname{argmin}_\phi g'_{nT}(h(\xi)) \widehat{\Sigma}_{nT}^{-1} g_{nT}(h(\xi))$ . Then, we have  $\widehat{\theta}_{nT,r} = h(\widehat{\xi}_{nT})$ . Let  $\tilde{\xi}_{nT}$  be a consistent estimate of  $\xi_0$ . Denote  $G_\xi(\theta) = \frac{1}{N} \frac{\partial g_{nT}(\theta)}{\partial \xi'}$ ,  $C_\xi(\theta) = G'_\xi(\theta) \widehat{\Sigma}_{nT}^{-1} \bar{g}_{nT}(\theta)$ , and  $B_\xi(\theta) = G'_\xi(\theta) \widehat{\Sigma}_{nT}^{-1} G_\xi(\theta)$ . Following the formulation suggested by Breusch and Pagan (1980), we state the  $C(\alpha)$  test statistic in the following way

$$C(\alpha) = N [C'(h(\tilde{\xi}_{nT})) B^{-1}(h(\tilde{\xi}_{nT})) C(h(\tilde{\xi}_{nT})) - C'_\xi(h(\tilde{\xi}_{nT})) B_\xi^{-1}(h(\tilde{\xi}_{nT})) C_\xi(h(\tilde{\xi}_{nT}))]. \quad (4.3)$$

126 In (4.3), it is important to note that  $\tilde{\xi}_{nT}$  can be any consistent estimator. In the case where  $\tilde{\xi}_{nT}$  is an optimal GMME, the  $C(\alpha)$  statistic reduces to  $LM$  statistic, since  $C_\xi(h(\tilde{\xi}_{nT})) = 0$  by definition.<sup>6</sup>  
 128 The asymptotic distributions of  $C(\alpha)$  and  $LM$  are given in the following proposition.

<sup>5</sup>Breusch and Pagan (1980) call this test the pseudo-LM test, since its test statistic is very similar to the form of the LM statistic.

<sup>6</sup>In the context of ML estimation, the  $C(\alpha)$  statistic reduces to the LM statistic when the restricted MLE is used. For details, see Bera and Biliias (2001).

**Proposition 1.** — Given our stated assumptions, we have the following results under  $H_0 : r(\theta_0) = 0$ :

$$LM \xrightarrow{d} \chi_{k_r}^2, \quad \text{and} \quad C(\alpha) \xrightarrow{d} \chi_{k_r}^2. \quad (4.4)$$

*Proof.* See Section C.1. □

Next, we consider the following joint null hypothesis:

$$H_0 : \lambda_0 = 0, \rho_0 = 0, \gamma_0 = 0, \quad H_A : \text{At least one parameter is not equal to zero.} \quad (4.5)$$

Under the joint null hypothesis, the model reduces to a two-way non-spatial panel data model which can be estimated by an OLSE (for the estimation of two-way models, see Baltagi (2008) and Hsiao (2014)). The joint null hypothesis can be tested either by  $LM$  or  $C(\alpha)$ . Let  $\tilde{\theta}_{nT}$  be a constrained optimal GMME under the joint null hypothesis, and let  $\hat{\theta}_{nT}$  be any other consistent estimator of  $\theta_0$  under the null hypothesis. As stated in Newey and West (1987), the LM test statistic should be formulated with the optimal constrained GMME. Let  $\vartheta = (\lambda', \rho', \gamma)'$ . Then, the LM test statistic for the joint null hypothesis can be expressed as

$$LM_J(\tilde{\theta}_{nT}) = N C'_J(\tilde{\theta}_{nT}) [B_{\vartheta \cdot \beta}(\tilde{\theta}_{nT})]^{-1} C_J(\tilde{\theta}_{nT}), \quad (4.6)$$

where  $C'_J(\tilde{\theta}_{nT}) = (C'_\lambda(\tilde{\theta}_{nT}), C'_\rho(\tilde{\theta}_{nT}), C'_\gamma(\tilde{\theta}_{nT}))'$ ,  $B_{\vartheta \cdot \beta}(\tilde{\theta}_{nT}) = B_\vartheta(\tilde{\theta}_{nT}) - B_{\vartheta\beta}(\tilde{\theta}_{nT})B_\beta^{-1}(\tilde{\theta}_{nT})B_{\beta\vartheta}(\tilde{\theta}_{nT})$ ,  $B_{\vartheta\beta}(\tilde{\theta}_{nT}) = B'_{\beta\vartheta}(\tilde{\theta}_{nT}) = (B'_{\lambda\beta}(\tilde{\theta}_{nT}), B'_{\rho\beta}(\tilde{\theta}_{nT}), B'_{\gamma\beta}(\tilde{\theta}_{nT}))'$ , and

$$B_\vartheta(\tilde{\theta}_{nT}) = \begin{pmatrix} B_\lambda(\tilde{\theta}_{nT}) & B_{\lambda\rho}(\tilde{\theta}_{nT}) & B_{\lambda\gamma}(\tilde{\theta}_{nT}) \\ B_{\rho\lambda}(\tilde{\theta}_{nT}) & B_\rho(\tilde{\theta}_{nT}) & B_{\rho\gamma}(\tilde{\theta}_{nT}) \\ B_{\gamma\lambda}(\tilde{\theta}_{nT}) & B_{\gamma\rho}(\tilde{\theta}_{nT}) & B_\gamma(\tilde{\theta}_{nT}) \end{pmatrix}. \quad (4.7)$$

Similarly, the consistent estimator  $\hat{\theta}_{nT}$  can be used to formulate the following  $C(\alpha)$  test for the joint null hypothesis:

$$C_J(\alpha) = N [C'(\hat{\theta}_{nT})B^{-1}(\hat{\theta}_{nT})C(\hat{\theta}_{nT}) - C'_\beta(\hat{\theta}_{nT})B_\beta^{-1}(\hat{\theta}_{nT})C_\beta(\hat{\theta}_{nT})]. \quad (4.8)$$

The properties of the LM test can be investigated under a sequence of local alternatives (Bera and Biliias 2001; Bera and Yoon 1993; Bera et al. 2010; Davidson and MacKinnon 1987; Saikkonen 1989). Bera and Yoon (1993) and Bera et al. (2010) suggest robust LM tests when the alternative model is misspecified. We consider similar robust LM tests within the context of our model. In order to give a general result, we consider the LM test for  $H_0^\psi : \psi_0 = 0$  when  $H_0^\phi : \phi_0 = 0$ , which can be stated as

$$LM_\psi = N C'_\psi(\tilde{\theta}_{nT}) [B_{\psi \cdot \beta}(\tilde{\theta}_{nT})]^{-1} C_\psi(\tilde{\theta}_{nT}), \quad (4.9)$$

where  $B_{\psi \cdot \beta}(\tilde{\theta}_{nT}) = B_\psi(\tilde{\theta}_{nT}) - B_{\psi\beta}(\tilde{\theta}_{nT})B_\beta^{-1}(\tilde{\theta}_{nT})B_{\beta\psi}(\tilde{\theta}_{nT})$ . We investigate the asymptotic distribution of  $LM_\psi$  under the sequences of local alternatives  $H_A^\psi : \psi = \psi_0 + \delta_\psi/\sqrt{N}$ , and  $H_A^\phi : \phi = \phi_0 + \delta_\phi/\sqrt{N}$ , where  $(\psi'_0, \phi'_0)'$  is the vector of hypothesized values under the null, and  $\delta_\psi$  and  $\delta_\phi$  are bounded vectors. The distribution of (4.9), under  $H_A^\psi$  and  $H_A^\phi$ , can be investigated from the first order Taylor expansions of pseudo-scores  $C_\psi(\tilde{\theta}_{nT})$  and  $C_\beta(\tilde{\theta}_{nT})$  around



$\theta^* = (\beta'_0, \psi'_0 + \delta'_\psi/\sqrt{N}, \phi'_0 + \delta'_\phi/\sqrt{N})'$ . These expansions can be written as

$$\begin{aligned} \sqrt{N} C_\psi(\tilde{\theta}_{nT}) &= \sqrt{N} C_\psi(\theta^*) - G'_\psi(\theta^*) \widehat{\Sigma}_{nT}^{-1} G_\psi(\bar{\theta}) \delta_\psi - G'_\psi(\theta^*) \widehat{\Sigma}_{nT}^{-1} G_\phi(\bar{\theta}) \delta_\phi \\ &\quad + \sqrt{N} G'_\psi(\theta^*) \widehat{\Sigma}_{nT}^{-1} G_\beta(\bar{\theta}) (\tilde{\beta}_{nT} - \beta_0) + o_p(1), \end{aligned} \quad (4.10)$$

$$\begin{aligned} \sqrt{N} C_\beta(\tilde{\theta}_{nT}) &= \sqrt{N} C_\beta(\theta^*) - G'_\beta(\theta^*) \widehat{\Sigma}_{nT}^{-1} G_\psi(\bar{\theta}) \delta_\psi - G'_\beta(\theta^*) \widehat{\Sigma}_{nT}^{-1} G_\phi(\bar{\theta}) \delta_\phi \\ &\quad + \sqrt{N} G'_\beta(\theta^*) \widehat{\Sigma}_{nT}^{-1} G_\beta(\bar{\theta}) (\tilde{\beta}_{nT} - \beta_0) + o_p(1), \end{aligned} \quad (4.11)$$

where  $\bar{\theta}$  lies between  $\tilde{\theta}_{nT}$  and  $\theta^*$ . Note that  $\theta^* = \theta_0 + o_p(1)$  implies  $\bar{\theta} = \theta_0 + o_p(1)$ . By Lemma 1, we have  $B(\theta^*) = \mathcal{H} + o_p(1)$ , and  $G'(\theta^*) \widehat{\Sigma}_{nT} = \mathcal{G}' \Sigma_{nT} + o_p(1)$ . Then, from (4.10) and (4.11), we get the following fundamental result:

$$\begin{aligned} \sqrt{N} C_\psi(\tilde{\theta}_{nT}) &= [\mathcal{G}'_\psi \Sigma_{nT}^{-1} - \mathcal{H}_{\psi\beta} \mathcal{H}_\beta^{-1} \mathcal{G}'_\beta \Sigma_{nT}^{-1}] \frac{1}{\sqrt{N}} g_{nT}(\theta_0) \\ &\quad - [\mathcal{H}_\psi - \mathcal{H}_{\psi\beta} \mathcal{H}_\beta^{-1} \mathcal{H}_{\beta\psi}] \delta_\psi - [\mathcal{H}_{\psi\phi} - \mathcal{H}_{\psi\beta} \mathcal{H}_\beta^{-1} \mathcal{H}_{\beta\phi}] \delta_\phi + o_p(1). \end{aligned} \quad (4.12)$$

130 By Lemma 1, we have  $\frac{1}{\sqrt{N}} g_{nT}(\theta_0) \xrightarrow{d} N(0, \text{plim}_{n \rightarrow \infty} \Sigma_{nT})$ , and thus (4.12) implies that  
 $\sqrt{N} C_\psi(\tilde{\theta}_{nT}) \xrightarrow{d} N(-\mathcal{H}_{\psi\cdot\beta} \delta_\psi - \mathcal{H}_{\psi\phi\cdot\beta} \delta_\phi, \mathcal{H}_{\psi\cdot\beta})$ , where  $\mathcal{H}_{\psi\cdot\beta} = [\mathcal{H}_\psi - \mathcal{H}_{\psi\beta} \mathcal{H}_\beta^{-1} \mathcal{H}_{\beta\psi}]$ , and  
132  $\mathcal{H}_{\psi\phi\cdot\beta} = [\mathcal{H}_{\psi\phi} - \mathcal{H}_{\psi\beta} \mathcal{H}_\beta^{-1} \mathcal{H}_{\beta\phi}]$ . Hence,  $LM_\psi(\tilde{\theta}_{nT}) \xrightarrow{d} \chi_{k_\psi}^2(\vartheta_1)$  under  $H_A^\psi$  and  $H_A^\phi$ , where  
134  $\vartheta_1 = \delta'_\psi \mathcal{H}_{\psi\cdot\beta} \delta_\psi + \delta'_\psi \mathcal{H}_{\psi\phi\cdot\beta} \delta_\phi + \delta'_\phi \mathcal{H}'_{\psi\phi\cdot\beta} \delta_\psi + \delta'_\phi \mathcal{H}'_{\psi\phi\cdot\beta} \mathcal{H}_{\psi\cdot\beta}^{-1} \mathcal{H}_{\psi\phi\cdot\beta} \delta_\phi$  is the non-centrality parameter.<sup>7</sup> We provide the distributional results for  $LM_\psi(\tilde{\theta}_{nT})$  and its robust version in the following proposition.

136 **Proposition 2.** — Given our stated assumptions, the following results hold.

1. Under  $H_A^\psi$  and  $H_A^\phi$ , we have

$$LM_\psi(\tilde{\theta}_{nT}) \xrightarrow{d} \chi_{k_\psi}^2(\vartheta_1), \quad (4.13)$$

where  $\vartheta_1 = \delta'_\psi \mathcal{H}_{\psi\cdot\beta} \delta_\psi + \delta'_\psi \mathcal{H}_{\psi\phi\cdot\beta} \delta_\phi + \delta'_\phi \mathcal{H}'_{\psi\phi\cdot\beta} \delta_\psi + \delta'_\phi \mathcal{H}'_{\psi\phi\cdot\beta} \mathcal{H}_{\psi\cdot\beta}^{-1} \mathcal{H}_{\psi\phi\cdot\beta} \delta_\phi$ .

2. Under  $H_A^\psi$  and  $H_0^\phi$ , we have

$$LM_\psi(\tilde{\theta}_{nT}) \xrightarrow{d} \chi_{k_\psi}^2(\vartheta_2), \quad (4.14)$$

138 where  $\vartheta_2 = \delta'_\psi \mathcal{H}_{\psi\cdot\beta} \delta_\psi$ .

3. Under  $H_0^\psi$  and  $H_A^\phi$ , we have

$$LM_\psi(\tilde{\theta}_{nT}) \xrightarrow{d} \chi_{k_\psi}^2(\vartheta_3), \quad (4.15)$$

where  $\vartheta_3 = \delta'_\phi \mathcal{H}'_{\psi\phi\cdot\beta} \mathcal{H}_{\psi\cdot\beta}^{-1} \mathcal{H}_{\psi\phi\cdot\beta} \delta_\phi$ .

4. Let  $C_\psi^*(\tilde{\theta}_{nT}) = [C_\psi(\tilde{\theta}_{nT}) - B_{\psi\phi\cdot\beta}(\tilde{\theta}_{nT}) B_{\phi\cdot\beta}^{-1}(\tilde{\theta}_{nT}) C_\phi(\tilde{\theta}_{nT})]$  be the adjusted pseudo-score, where  $B_{\psi\phi\cdot\beta}(\tilde{\theta}_{nT}) = B_{\psi\phi}(\tilde{\theta}_{nT}) - B_{\psi\beta}(\tilde{\theta}_{nT}) B_\beta^{-1}(\tilde{\theta}_{nT}) B_{\beta\phi}(\tilde{\theta}_{nT})$ , and  $B_{\phi\cdot\beta}(\tilde{\theta}_{nT}) = B_\phi(\tilde{\theta}_{nT}) -$

<sup>7</sup>For the definition of non-centrality chi-square distribution, see Anderson (2003, p.81-82).

$B_{\phi\beta}(\tilde{\theta}_{nT})B_{\beta}^{-1}(\tilde{\theta}_{nT})B_{\beta\phi}(\tilde{\theta}_{nT})$ . Under  $H_0^{\psi}$  and irrespective of whether  $H_0^{\phi}$  or  $H_A^{\phi}$  holds, we have

$$LM_{\psi}^*(\tilde{\theta}_{nT}) = N C_{\psi}^{\star\prime}(\tilde{\theta}_{nT}) [B_{\psi\cdot\beta}(\tilde{\theta}_{nT}) - B_{\psi\phi\cdot\beta}(\tilde{\theta}_{nT})B_{\phi\cdot\beta}^{-1}(\tilde{\theta}_{nT})B'_{\psi\phi\cdot\beta}(\tilde{\theta}_{nT})]^{-1} C_{\psi}^{\star}(\tilde{\theta}_{nT}) \xrightarrow{d} \chi_{k_{\psi}}^2. \quad (4.16)$$

5. Under  $H_A^{\psi}$  and  $H_0^{\phi}$ , we have

$$LM_{\psi}^*(\tilde{\theta}_{nT}) \xrightarrow{d} \chi_{k_{\psi}}^2(\vartheta_4), \quad (4.17)$$

140 where  $\vartheta_4 = \delta'_{\psi}(\mathcal{H}_{\psi\cdot\beta} - \mathcal{H}_{\psi\phi\cdot\beta}\mathcal{H}_{\phi\cdot\beta}^{-1}\mathcal{H}'_{\psi\phi\cdot\beta})\delta_{\psi}$ .

*Proof.* See Section C.2. □

142 There are three important observations regarding to the results presented in Proposition 2. First, the one directional test has a non-central chi-square distribution when the alternative model is misspecified, i.e., when the alternative model includes  $\phi_0$ . The non-centrality parameter is 144  $\vartheta_3 = \delta'_{\phi}\mathcal{H}'_{\psi\phi\cdot\beta}\mathcal{H}_{\psi\cdot\beta}^{-1}\mathcal{H}_{\psi\phi\cdot\beta}\delta_{\phi}$ , which would be zero if and only if  $\mathcal{H}_{\psi\phi\cdot\beta} = 0$ . Second, the robust 146 test  $LM_{\psi}^*(\tilde{\theta}_{nT})$  has a central chi-square distribution even when the alternative model is locally misspecified. Finally,  $LM_{\psi}^*(\tilde{\theta}_{nT})$  has less asymptotic power than  $LM_{\psi}(\tilde{\theta}_{nT})$ , since  $\vartheta_2 - \vartheta_4 \geq 0$  148 under  $H_A^{\psi}$  and  $H_0^{\phi}$ .

150 Proposition 2 provides a template that can be used to determine the test statistics for the following hypotheses:

1. The null hypothesis for the contemporaneous spatial lag terms:  $H_0^{\lambda} : \lambda_0 = 0$  in the presence of  $\rho_0$  and  $\gamma_0$ . 152
2. The null hypothesis for the spatial lag terms at time  $t - 1$ :  $H_0^{\rho} : \rho_0 = 0$  in the presence of  $\lambda_0$  and  $\gamma_0$ . 154
3. The null hypothesis for the time lag term:  $H_0^{\gamma} : \gamma_0 = 0$  in the presence of  $\lambda_0$  and  $\rho_0$ .

In the following, we provide the test statistic for each hypothesis and leave the detailed derivations to Appendix B. We start with  $H_0^{\lambda} : \lambda_0 = 0$ . In the context of this hypothesis,  $\phi = (\rho', \gamma)'$ . Then, the one directional test can be written as

$$LM_{\lambda}(\tilde{\theta}_{nT}) = N C'_{\lambda}(\tilde{\theta}_{nT}) [B_{\lambda\cdot\beta}(\tilde{\theta}_{nT})]^{-1} C_{\lambda}(\tilde{\theta}_{nT}), \quad (4.18)$$

where  $B_{\lambda\cdot\beta}(\tilde{\theta}_{nT}) = B_{\lambda}(\tilde{\theta}_{nT}) - B_{\lambda\beta}(\tilde{\theta}_{nT})B_{\beta}^{-1}(\tilde{\theta}_{nT})B_{\beta\lambda}(\tilde{\theta}_{nT})$ . Then,  $LM_{\lambda}(\tilde{\theta}_{nT}) \xrightarrow{d} \chi_p^2(\vartheta_2)$  under  $H_A^{\lambda}$  and  $H_0^{\phi}$ ; and  $LM_{\lambda}(\tilde{\theta}_{nT}) \xrightarrow{d} \chi_p^2(\vartheta_3)$  under  $H_0^{\lambda}$  and  $H_A^{\phi}$ , where  $\vartheta_2 = \delta'_{\lambda}\mathcal{H}_{\lambda\cdot\beta}\delta_{\lambda}$  and  $\vartheta_3 = \delta'_{\phi}\mathcal{H}'_{\lambda\phi\cdot\beta}\mathcal{H}_{\lambda\cdot\beta}^{-1}\mathcal{H}_{\lambda\phi\cdot\beta}\delta_{\phi}$ . The robust version is stated as

$$LM_{\lambda}^*(\tilde{\theta}_{nT}) = N C_{\lambda}^{\star\prime}(\tilde{\theta}_{nT}) [B_{\lambda\cdot\beta}(\tilde{\theta}_{nT}) - B_{\lambda\phi\cdot\beta}(\tilde{\theta}_{nT})B_{\phi\cdot\beta}^{-1}(\tilde{\theta}_{nT})B'_{\lambda\phi\cdot\beta}(\tilde{\theta}_{nT})]^{-1} C_{\lambda}^{\star}(\tilde{\theta}_{nT}), \quad (4.19)$$

156 where  $C_{\lambda}^{\star}(\tilde{\theta}_{nT}) = [C_{\lambda}(\tilde{\theta}_{nT}) - B_{\lambda\phi\cdot\beta}(\tilde{\theta}_{nT})B_{\phi\cdot\beta}^{-1}(\tilde{\theta}_{nT})C_{\phi}(\tilde{\theta}_{nT})]$  is the adjusted score. Irrespective of whether  $H_0^{\phi}$  or  $H_A^{\phi}$  holds,  $LM_{\lambda}^*(\tilde{\theta}_{nT})$  has an asymptotic  $\chi_p^2$  distribution under  $H_0^{\lambda}$  by Propo- 158 sition 2. Finally, under  $H_A^{\lambda}$  and  $H_0^{\phi}$ , we have  $LM_{\lambda}^*(\tilde{\theta}_{nT}) \xrightarrow{d} \chi_p^2(\vartheta_4)$ , where  $\vartheta_4 = \delta'_{\lambda}(\mathcal{H}_{\lambda\cdot\beta} - \mathcal{H}_{\lambda\phi\cdot\beta}\mathcal{H}_{\phi\cdot\beta}^{-1}\mathcal{H}'_{\lambda\phi\cdot\beta})\delta_{\lambda}$ .

Next, we consider  $H_0^\rho : \rho_0 = 0$ . In the context of this hypothesis,  $\phi = (\lambda', \gamma)'$ . The one directional test can be written as

$$LM_\rho(\tilde{\theta}_{nT}) = N C'_\rho(\tilde{\theta}_{nT}) [B_{\rho\cdot\beta}(\tilde{\theta}_{nT})]^{-1} C_\rho(\tilde{\theta}_{nT}), \quad (4.20)$$

where  $B_{\rho\cdot\beta}(\tilde{\theta}_{nT}) = B_\rho(\tilde{\theta}_{nT}) - B_{\rho\beta}(\tilde{\theta}_{nT})B_\beta^{-1}(\tilde{\theta}_{nT})B_{\beta\rho}(\tilde{\theta}_{nT})$ . Proposition 2 implies that  $LM_\rho(\tilde{\theta}_{nT}) \xrightarrow{d} \chi_p^2(\vartheta_2)$  under  $H_A^\rho$  and  $H_0^\phi$ ; and  $LM_\rho(\tilde{\theta}_{nT}) \xrightarrow{d} \chi_p^2(\vartheta_3)$  under  $H_0^\rho$  and  $H_A^\phi$ , where  $\vartheta_2 = \delta'_\rho \mathcal{H}_{\rho\cdot\beta} \delta_\rho$  and  $\vartheta_3 = \delta'_\phi \mathcal{H}'_{\rho\phi\cdot\beta} \mathcal{H}_{\rho\cdot\beta}^{-1} \mathcal{H}_{\rho\phi\cdot\beta} \delta_\phi$ . The robust version of  $LM_\rho(\tilde{\theta}_{nT})$  is stated as

$$LM_\rho^*(\tilde{\theta}_{nT}) = N C_{\rho'}^*(\tilde{\theta}_{nT}) [B_{\rho\cdot\beta}(\tilde{\theta}_{nT}) - B_{\rho\phi\cdot\beta}(\tilde{\theta}_{nT})B_{\phi\cdot\beta}^{-1}(\tilde{\theta}_{nT})B'_{\rho\phi\cdot\beta}(\tilde{\theta}_{nT})]^{-1} C_\rho^*(\tilde{\theta}_{nT}), \quad (4.21)$$

160 where  $C_\rho^*(\tilde{\theta}_{nT}) = [C_\rho(\tilde{\theta}_{nT}) - B_{\rho\phi\cdot\beta}(\tilde{\theta}_{nT})B_{\phi\cdot\beta}^{-1}(\tilde{\theta}_{nT})C_\phi(\tilde{\theta}_{nT})]$ . The asymptotic null distribution of  $LM_\rho^*(\tilde{\theta}_{nT})$  is  $\chi_p^2$ , irrespective of whether  $H_0^\phi$  or  $H_A^\phi$  holds. Finally, under  $H_A^\rho$  and  $H_0^\phi$ , we have  
162  $LM_\rho^*(\tilde{\theta}_{nT}) \xrightarrow{d} \chi_p^2(\vartheta_4)$ , where  $\vartheta_4 = \delta'_\rho (\mathcal{H}_{\rho\cdot\beta} - \mathcal{H}_{\rho\phi\cdot\beta} \mathcal{H}_{\phi\cdot\beta}^{-1} \mathcal{H}'_{\rho\phi\cdot\beta}) \delta_\rho$ .

Finally, we consider  $H_0^\gamma : \gamma_0 = 0$ . Here, we have  $\phi = (\lambda', \rho)'$ . The one directional test can be written as

$$LM_\gamma(\tilde{\theta}_{nT}) = N C'_\gamma(\tilde{\theta}_{nT}) [B_{\gamma\cdot\beta}(\tilde{\theta}_{nT})]^{-1} C_\gamma(\tilde{\theta}_{nT}), \quad (4.22)$$

where  $B_{\gamma\cdot\beta}(\tilde{\theta}_{nT}) = B_\gamma(\tilde{\theta}_{nT}) - B_{\gamma\beta}(\tilde{\theta}_{nT})B_\beta^{-1}(\tilde{\theta}_{nT})B_{\beta\gamma}(\tilde{\theta}_{nT})$ . Then,  $LM_\gamma(\tilde{\theta}_{nT}) \xrightarrow{d} \chi_1^2(\vartheta_2)$  under  $H_A^\gamma$  and  $H_0^\phi$ ; and  $LM_\gamma(\tilde{\theta}_{nT}) \xrightarrow{d} \chi_1^2(\vartheta_3)$  under  $H_0^\gamma$  and  $H_A^\phi$ , where  $\vartheta_2 = \delta'_\gamma \mathcal{H}_{\gamma\cdot\beta} \delta_\gamma$  and  $\vartheta_3 = \delta'_\phi \mathcal{H}'_{\gamma\phi\cdot\beta} \mathcal{H}_{\gamma\cdot\beta}^{-1} \mathcal{H}_{\gamma\phi\cdot\beta} \delta_\phi$ . The robust version is stated as

$$LM_\gamma^*(\tilde{\theta}_{nT}) = N C_{\gamma'}^*(\tilde{\theta}_{nT}) [B_{\gamma\cdot\beta}(\tilde{\theta}_{nT}) - B_{\gamma\phi\cdot\beta}(\tilde{\theta}_{nT})B_{\phi\cdot\beta}^{-1}(\tilde{\theta}_{nT})B'_{\gamma\phi\cdot\beta}(\tilde{\theta}_{nT})]^{-1} C_\gamma^*(\tilde{\theta}_{nT}), \quad (4.23)$$

where  $C_\gamma^*(\tilde{\theta}_{nT}) = [C_\gamma(\tilde{\theta}_{nT}) - B_{\gamma\phi\cdot\beta}(\tilde{\theta}_{nT})B_{\phi\cdot\beta}^{-1}(\tilde{\theta}_{nT})C_\phi(\tilde{\theta}_{nT})]$ . The asymptotic null distribution of  
164  $LM_\gamma^*(\tilde{\theta}_{nT})$  is  $\chi_1^2$ , irrespective of whether  $H_0^\phi$  or  $H_A^\phi$  holds. Finally, under  $H_A^\gamma$  and  $H_0^\phi$ , we have  
 $LM_\gamma^*(\tilde{\theta}_{nT}) \xrightarrow{d} \chi_1^2(\vartheta_4)$ , where  $\vartheta_4 = \delta'_\gamma (\mathcal{H}_{\gamma\cdot\beta} - \mathcal{H}_{\gamma\phi\cdot\beta} \mathcal{H}_{\phi\cdot\beta}^{-1} \mathcal{H}'_{\gamma\phi\cdot\beta}) \delta_\gamma$ .

## 166 5 Monte Carlo Simulation

In this section, we describe the details of Monte Carlo design for our analysis. Our design is based on Lee and Yu (2014) and Yang (2015). For the model in (2.1), we will focus on the case where  $p = 1$ :

$$Y_{nt} = \lambda_0 W_n Y_{nt} + \gamma_0 Y_{n,t-1} + \rho_0 W_n Y_{n,t-1} + X_{nt} \beta_0 + \mathbf{c}_{n0} + \alpha_{t0} l_n + V_{nt}, \quad (5.1)$$

for  $t = 1, 2, \dots, T$ . We generate the weights matrix according to (i) Rook contiguity and (ii) Queen  
168 contiguity. The  $n$  spatial units are randomly permuted and allocated into a lattice of  $k \times m$  squares, where  $m \geq n$ . In the Rook contiguity,  $w_{ij,n} = 1$  if the spatial unit  $j$  is in a square that is adjacent  
170 (left/right/above or below) to the square of the spatial unit  $i$ . In the Queen contiguity,  $w_{ij,n} = 1$  if the spatial unit  $j$  is in a square that is adjacent to or shares a corner with the square of the spatial  
172 unit  $i$ . In both cases,  $W_n$  is row normalized.

We allow for two exogenous regressors. The first one is generated as  $X_{1,nt} = \Psi_n + 0.01 t l_n + U_{nt}$ ,  
174 where  $U_{nt} = 0.5 U_{n,t-1} + \varepsilon_{nt} + 0.5 \varepsilon_{n,t-1}$  and  $\varepsilon_{nt} \sim N(0_{n \times 1}, 2I_n)$ . Furthermore,  $\Psi_n = \Upsilon_n + 1/(T+m +$

1)  $\sum_{t=-m}^T \varepsilon_{nt}$ , where  $\Upsilon_n \sim N(0_{n \times 1}, I_n)$  and  $m = 20$ . Then,  $X_{nt} = (X_{1,nt}, W_n X_{2,nt})$  where  $X_{2,nt} \sim N(0_{n \times 1}, I_n)$ . We set  $\beta_0 = (1.2, 0.6)$ . For the individual effects, we let  $\mathbf{c}_{n0} = (1/T) \sum_{t=1}^T X_{1,nt}$ , and draw  $\alpha_{t0}$  from  $N(0, 1)$ . For the error term  $V_{i,nt}$ , we specify two cases: (i)  $V_{i,nt} \sim N(0, 1)$  and (ii)  $V_{i,nt} \sim \text{Gamma}(1, 1) - 1$ . The data generating process has  $21 + T$  periods and the last  $T + 1$  periods are used for estimation. For the sample size, we use the following  $n$  and  $T$  combinations:  $(n, T) = \{(100, 10), (20, 200)\}$ .<sup>8</sup>

Under the null model (i.e.,  $\lambda_0 = \gamma_0 = \rho_0 = 0$ ), (5.1) reduces to a two-way error model (2WE). We can employ seven different specifications for the alternative model. We choose to focus on the following four specifications as they are more common in empirical applications. The first specification is a dynamic panel data model with no spatial effects (DPD), i.e., when  $\lambda_0 = \rho_0 = 0$  and  $\gamma_0 \neq 0$  in (5.1). The second specification is a spatial static panel model (SSPD), i.e., when  $\lambda_0 \neq 0$  and  $\rho_0 = \gamma_0 = 0$  in (5.1). The third specification is a spatial dynamic panel data model with no spatial-time lag (SDPDW), i.e., when  $\rho_0 = 0$ ,  $\lambda_0 \neq 0$  and  $\gamma_0 \neq 0$  in (5.1). The final specification for the alternative modes is the spatial dynamic panel data model (SDPD), i.e., when  $\rho_0 \neq 0$ ,  $\lambda_0 \neq 0$  and  $\gamma_0 \neq 0$  in (5.1). Note that the first three alternative models can be considered as the null models for the one-directional tests and their robust counterparts in the following way: (i) the DPD model for  $\text{LM}_\rho$ ,  $\text{LM}_\rho^*$ ,  $\text{LM}_\lambda$  and  $\text{LM}_\lambda^*$ ; (ii) the SSP model for  $\text{LM}_\rho$ ,  $\text{LM}_\rho^*$ ,  $\text{LM}_\gamma$  and  $\text{LM}_\gamma^*$ ; (iii) the SDPDW model for  $\text{LM}_\rho$  and  $\text{LM}_\rho^*$ . We let  $\lambda_0$ ,  $\gamma_0$  and  $\rho_0$  take values from  $\{-0.3, -0.1, -0.05, 0.05, 0.1, 0.3\}$  for the alternative models. Hence, the DPD, SSPD, SDPDW and SDPD specifications yield respectively 6, 6, 16 and 216 combinations. Resampling is carried out for 5,000 times.

Table 1 summarizes the null hypotheses and the respective test statistics along with the source of misspecification in each hypothesis considered in the Monte Carlo study. For example, the source of misspecification for  $H_0 : \lambda_0 = 0$  is the presence of  $\rho_0$  and  $\gamma_0$  in the alternative model. All test statistics presented in Table 1 are computed by the estimates from the 2WE model. For the test statistics, we also need to specify the set of moment functions. The set of linear moments consists of  $Q_{nt} = (Y_{n,t-1}, W_n Y_{n,t-1}, W_n^2 Y_{n,t-1}, X_{n,t}^*, W_n X_{n,t}^*, W_n^2 X_{n,t}^*)$  for  $t = 1, 2, \dots, T - 1$ . For the quadratic moments, we employ  $P_{n1} = W_n - \text{tr}(W_n J_n)/(n - 1)J_n$  and  $P_{n2} = W_n^2 - \text{tr}(W_n^2 J_n)/(n - 1)J_n$ . Note that we do not consider the conditional tests that require a restricted GMME (see Proposition 1) for the computation of the test statistics. Here our aim is to compare the performance of the robust tests with their non-robust counterparts once the estimates of the simple 2WE model are available.

## 5.1 Results on Size Properties

A P value discrepancy plots is generated from the empirical distribution function (edf) of  $p$  values. To see how, let  $\tau$  denote a test statistic, and  $\tau_j$  for  $j = 1, \dots, \mathcal{R}$  be the  $\mathcal{R}$  realizations of  $\tau$  generated in a Monte Carlo experiment. Let  $F(x)$  denote the cumulative distribution function (cdf) of the asymptotic distribution of  $\tau$  evaluated at the level  $x$ . Then, the  $p$  value associated with  $\tau_j$ , denoted by  $p(\tau_j)$ , is given by  $p(\tau_j) = 1 - F(\tau_j)$ . An estimate of the cdf of  $p(\tau)$  can be constructed simply from the edf of  $p(\tau_j)$ . Consider a sequence of levels denoted by  $\{x_i\}$  for  $i = 1, \dots, m$  from the interval  $(0, 1)$ . Then, an estimate of the cdf of  $p(\tau)$  is given by  $\hat{F}(x_i) = \sum_{j=1}^{\mathcal{R}} \mathbf{1}(p(\tau_j) \leq x_i)/\mathcal{R}$ .<sup>9</sup>

The P value discrepancy plot is created by plotting  $\hat{F}(x_i) - x_i$  against  $x_i$  under the assumption that the true data generating process is characterized by the null hypothesis. To assess the

<sup>8</sup>For the sake of brevity, we only provide estimation results for  $(n, T) = (100, 10)$ .

<sup>9</sup>We choose the following sequence and focus on the levels smaller than or equal to 0.1:  $\{x_i\}_{i=1}^m = \{0.001 : 0.001 : 0.010 \quad 0.015 : 0.005 : 0.990 \quad 0.991 : 0.001 : 0.999\}$ .

Table 1: Summary of test statistics

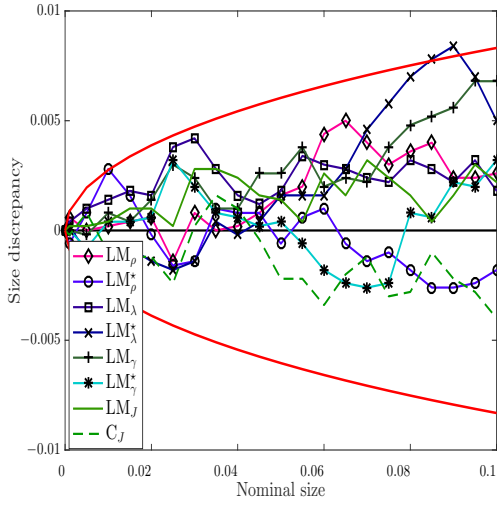
Null hypothesis	Parameter		Test statistic
	Spatial time lag: $\rho_0$	Time lag: $\gamma_0$	
$H_0 : \lambda_0 = 0$	Set to zero	Set to zero	$LM_\lambda$ in (4.18)
$H_0 : \lambda_0 = 0$	Unrestricted, not estimated	Unrestricted, not estimated	$LM_\lambda^*$ in (4.19)
	Contemporaneous spatial lag: $\lambda_0$	Time lag: $\gamma_0$	
$H_0 : \rho_0 = 0$	Set to zero	Set to zero	$LM_\rho$ in (4.20)
$H_0 : \rho_0 = 0$	Unrestricted, not estimated	Unrestricted, not estimated	$LM_\rho^*$ in (4.21)
	Contemporaneous spatial lag: $\lambda_0$	Spatial time lag: $\rho_0$	
$H_0 : \gamma_0 = 0$	Set to zero	Set to zero	$LM_\gamma$ in (4.22)
$H_0 : \gamma_0 = 0$	Unrestricted, not estimated	Unrestricted, not estimated	$LM_\gamma^*$ in (4.23)
$H_0 : \lambda_0 = 0, \rho_0 = 0, \gamma_0 = 0$	-	-	$LM_J$ in (4.6)
$H_0 : \lambda_0 = 0, \rho_0 = 0, \gamma_0 = 0$	-	-	$C_J$ in (4.8)

significance of discrepancies in a P value discrepancy plot, we construct a point-wise 95% confidence interval for a nominal size by using a normal approximation to the binomial distribution (Anselin et al. 1996). Let  $\alpha$  denote the nominal size at which the test is carried out. Using a normal approximation to the binomial distribution, a point-wise 95% confidence interval centered on  $\alpha$  would be given by  $\alpha \pm 1.96 [\alpha(1 - \alpha)/\mathcal{R}]^{1/2}$ , and thus it would include rejection rates between  $\alpha - 1.96 [\alpha(1 - \alpha)/\mathcal{R}]^{1/2}$  and  $\alpha + 1.96 [\alpha(1 - \alpha)/\mathcal{R}]^{1/2}$ . We use this approach to insert a 95% confidence interval in a P value discrepancy plot. In the discrepancy plots, the interval will be represented by the red solid lines.

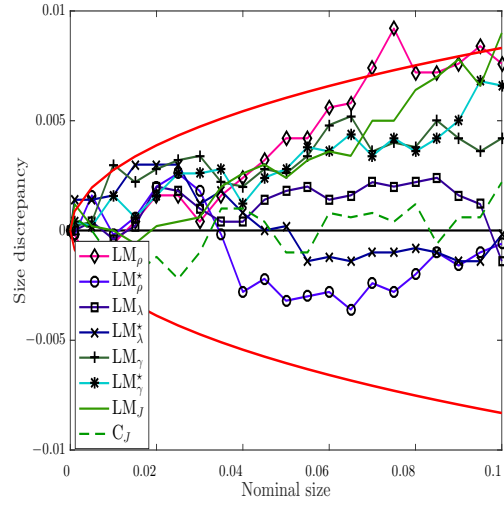
Table 2: Empirical sizes when  $H_0$ : The DPD model and  $(n, T) = (100, 10)$ 

$\gamma_0$	<i>Normal Distribution</i>				<i>Gamma Distribution</i>			
	$LM_\rho$	$LM_\rho^*$	$LM_\lambda$	$LM_\lambda^*$	$LM_\rho$	$LM_\rho^*$	$LM_\lambda$	$LM_\lambda^*$
	Rook							
-0.30	0.046	0.015	0.042	0.005	0.047	0.016	0.042	0.008
-0.10	0.044	0.038	0.042	0.039	0.040	0.042	0.041	0.037
-0.05	0.040	0.049	0.048	0.051	0.043	0.045	0.047	0.046
0.05	0.061	0.046	0.061	0.056	0.057	0.051	0.056	0.052
0.10	0.074	0.042	0.064	0.039	0.070	0.041	0.061	0.043
0.30	0.135	0.028	0.100	0.024	0.128	0.035	0.099	0.028
	Queen							
-0.30	0.063	0.020	0.053	0.012	0.062	0.018	0.049	0.011
-0.10	0.044	0.047	0.046	0.043	0.039	0.038	0.044	0.038
-0.05	0.049	0.053	0.051	0.048	0.044	0.048	0.042	0.044
0.05	0.055	0.046	0.058	0.051	0.062	0.049	0.055	0.050
0.10	0.075	0.050	0.060	0.050	0.070	0.045	0.061	0.043
0.30	0.099	0.012	0.062	0.017	0.083	0.015	0.051	0.020

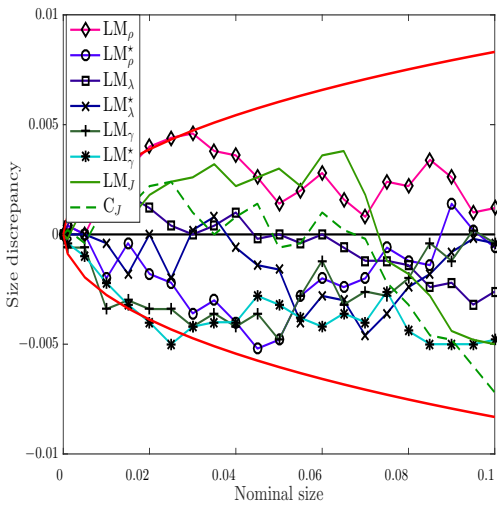
To save space, the size results based on the 2WE model will be presented through the P value discrepancy plots whereas the size results based on the DPD, SSPD and SDPDW models will be summarized in tables. Note that in our design we allow for 6 different values for  $\lambda_0$ ,  $\gamma_0$  and  $\rho_0$ , which would yield 216 P value discrepancy plots for each. Hence, when the null model is one of the DPD, SSPD and SDPDW models, we focus solely on the nominal size of 5% and provide size



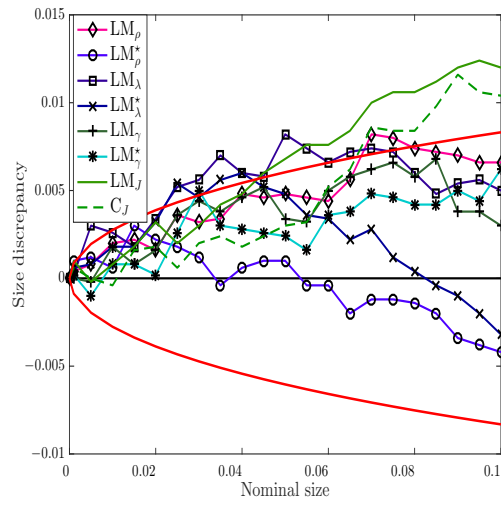
(a) Rook weight matrix and normal errors



(b) Rook weight matrix and non-normal errors



(c) Queen weight matrix and normal errors



(d) Queen weight matrix and non-normal errors

Figure 1: Size discrepancy plots when  $(n, T) = (100, 10)$ .

230 deviations at this level only. The general observations on the size properties of tests from Figure 1  
 and Tables 2 through 4 are listed as follows.

- 232 1. Figure 1 presents the size discrepancy plots when the null model is 2WE. The results show  
 that all tests have little size distortions and their size discrepancies generally lie inside the 95%  
 234 confidence interval. The size discrepancies are relatively larger in the case of queen weight  
 matrix and non-normal errors.
- 236 2. Table 2 provide some evidences on the magnitude of size distortions as a function of the size  
 of local misspecification of the alternative model, the DPD model. One would expect to see  
 238 robust versions of the one directional tests,  $LM_\rho^*$  and  $LM_\lambda^*$ , to perform better than  $LM_\rho$  and  
 $LM_\lambda$ , respectively, when the magnitude of misspecification is small. Overall, this seems to

- 240 be the case. For example, when the value of  $\gamma_0$  is 0.05 in absolute value in the true model,  
 242 the actual size of the robust tests are very close to the nominal size of 5%. However, as the  
 misspecification deteriorates, this property of the robust tests vanish as expected.
- 244 3. Similar results hold for Table 3 as well, the robust versions of the one directional tests,  $LM_\rho^*$   
 and  $LM_\gamma^*$ , perform better than  $LM_\rho$  and  $LM_\gamma$ , respectively, when  $\lambda_0$  deviates locally from  
 zero in the null model.
  - 246 4. Tables 4 and 5 confirms our previous findings:  $LM_\rho^*$  perform better than  $LM_\rho$ , when  $\lambda_0$  and  
 248  $\gamma_0$  deviate locally from zero. For example, in Table 4, when true values of  $\lambda_0$  and  $\gamma_0$  are 0.1,  
 the actual size of  $LM_\rho^*$  is 0.045 at the 5% level in the case of normal errors, whereas the actual  
 size of  $LM_\rho$  is 0.985.
  - 250 5. Recall that the robust tests use the residuals from the estimation of 2WE model and imple-  
 252 ments a correction on the test statistics for a local misspecification of the alternative model,  
 i.e., ignoring the spatial component(s). The bias in these residuals depends on the strength  
 254 of spatial dependence as well as on the connectedness of the weights matrix. Therefore, we  
 can expect poor performance for the robust tests as spatial parameters deviate from zero  
 substantially in the alternative model.
  - 256 6. Finally, Tables 2, 4 and 5 indicate that as the temporal dependence strengthens, i.e., the mis-  
 258 specification in  $\gamma_0$  gets larger in absolute value, the performance of the robust one-directional  
 tests deteriorates significantly relative to their non-robust counterparts. This is not surprising  
 260 in the sense that the bias in the residuals from the estimation of 2WE model increase as the  
 dependence over time strengthens.

Table 3: Empirical sizes when  $H_0$ : The SSPD model and  $(n, T) = (100, 10)$

$\lambda_0$	<i>Normal Distribution</i>				<i>Gamma Distribution</i>			
	$LM_\rho$	$LM_\rho^*$	$LM_\gamma$	$LM_\gamma^*$	$LM_\rho$	$LM_\rho^*$	$LM_\gamma$	$LM_\gamma^*$
	Rook							
-0.30	1.000	0.791	1.000	0.999	1.000	0.793	1.000	1.000
-0.10	0.913	0.051	0.334	0.184	0.913	0.048	0.335	0.168
-0.05	0.394	0.053	0.087	0.071	0.379	0.052	0.085	0.065
0.05	0.326	0.048	0.077	0.068	0.336	0.051	0.073	0.064
0.10	0.853	0.051	0.204	0.136	0.863	0.053	0.215	0.143
0.30	1.000	0.730	0.998	0.997	1.000	0.708	0.999	0.998
	Queen							
-0.30	0.994	0.134	0.604	0.431	0.997	0.144	0.614	0.451
-0.10	0.393	0.058	0.070	0.068	0.374	0.055	0.072	0.064
-0.05	0.134	0.052	0.056	0.057	0.132	0.047	0.049	0.049
0.05	0.171	0.046	0.073	0.063	0.187	0.045	0.060	0.054
0.10	0.550	0.053	0.103	0.071	0.539	0.055	0.116	0.073
0.30	0.999	0.202	0.972	0.970	0.999	0.195	0.972	0.969

## 5.2 Results on Power Properties

262 To investigate power properties of all tests, we use the approach described in Davidson and MacK-  
 innon (1998) to generate the size power curves against the actual size obtained under the cor-

Table 4: Empirical sizes when  $H_0$ : The SDPDW model and  $(n, T) = (100, 10)$ : Rook

$\lambda_0$	$\gamma_0$	<i>Normal Distribution</i>		<i>Gamma Distribution</i>	
		$LM_\rho$	$LM_\rho^*$	$LM_\rho$	$LM_\rho^*$
-0.30	-0.30	0.580	0.299	0.578	0.310
-0.30	-0.10	0.999	0.833	1.000	0.831
-0.30	-0.05	1.000	0.831	1.000	0.830
-0.30	0.05	1.000	0.772	1.000	0.778
-0.30	0.10	1.000	0.885	1.000	0.892
-0.30	0.30	1.000	1.000	1.000	1.000
-0.10	-0.30	0.120	0.045	0.118	0.043
-0.10	-0.10	0.466	0.039	0.466	0.039
-0.10	-0.05	0.734	0.046	0.751	0.048
-0.10	0.05	0.971	0.048	0.974	0.042
-0.10	0.10	0.990	0.047	0.987	0.049
-0.10	0.30	1.000	0.739	0.999	0.739
-0.05	-0.30	0.073	0.025	0.067	0.024
-0.05	-0.10	0.128	0.044	0.137	0.045
-0.05	-0.05	0.242	0.050	0.255	0.047
-0.05	0.05	0.515	0.051	0.502	0.048
-0.05	0.10	0.612	0.049	0.610	0.046
-0.05	0.30	0.819	0.219	0.835	0.215
0.05	-0.30	0.068	0.018	0.062	0.015
0.05	-0.10	0.121	0.046	0.115	0.036
0.05	-0.05	0.207	0.049	0.208	0.053
0.05	0.05	0.474	0.051	0.469	0.053
0.05	0.10	0.557	0.042	0.585	0.045
0.05	0.30	0.598	0.022	0.597	0.017
0.10	-0.30	0.133	0.031	0.134	0.035
0.10	-0.10	0.360	0.042	0.347	0.051
0.10	-0.05	0.639	0.053	0.639	0.056
0.10	0.05	0.956	0.054	0.957	0.055
0.10	0.10	0.985	0.045	0.985	0.046
0.10	0.30	0.990	0.151	0.991	0.157
0.30	-0.30	0.763	0.296	0.764	0.302
0.30	-0.10	0.976	0.746	0.970	0.768
0.30	-0.05	1.000	0.757	1.000	0.754
0.30	0.05	1.000	0.643	1.000	0.652
0.30	0.10	1.000	0.637	1.000	0.627
0.30	0.30	1.000	1.000	1.000	1.000

264 responding null hypothesis. Therefore, two experiments need to be carried out. First, the data  
generating process under the alternative hypothesis is used to generate the edf of p-values. We  
266 denote the resulting edf by  $\tilde{F}(x)$ . Second, the data generating process satisfies the null hypothesis,  
and as before  $\hat{F}(x)$  denotes the resulting edf of p-values. Then, a size-power curve is generated by  
268 plotting  $\tilde{F}(x_i)$  against  $\hat{F}(x_i)$  for  $i = 1, \dots, m$ . As stated in Davidson and MacKinnon (1998), the  
size-power curve avoids the size adjustments made to generate the power curves.



Table 5: Empirical sizes when  $H_0$ : The SDPDW model and  $(n, T) = (100, 10)$ : Queen

		<i>Normal Distribution</i>		<i>Gamma Distribution</i>	
$\lambda_0$	$\gamma_0$	$LM_\rho$	$LM_\rho^*$	$LM_\rho$	$LM_\rho^*$
-0.30	-0.30	0.223	0.021	0.227	0.017
-0.30	-0.10	0.670	0.125	0.662	0.118
-0.30	-0.05	0.935	0.153	0.934	0.153
-0.30	0.05	1.000	0.105	1.000	0.106
-0.30	0.10	1.000	0.067	1.000	0.065
-0.30	0.30	1.000	0.418	1.000	0.432
-0.10	-0.30	0.046	0.021	0.041	0.021
-0.10	-0.10	0.126	0.042	0.120	0.043
-0.10	-0.05	0.230	0.049	0.234	0.048
-0.10	0.05	0.541	0.045	0.533	0.050
-0.10	0.10	0.638	0.048	0.636	0.043
-0.10	0.30	0.675	0.021	0.670	0.019
-0.05	-0.30	0.043	0.020	0.045	0.020
-0.05	-0.10	0.058	0.039	0.062	0.042
-0.05	-0.05	0.092	0.048	0.094	0.047
-0.05	0.05	0.179	0.050	0.175	0.051
-0.05	0.10	0.221	0.053	0.210	0.049
-0.05	0.30	0.209	0.009	0.208	0.010
0.05	-0.30	0.121	0.024	0.117	0.021
0.05	-0.10	0.065	0.042	0.061	0.041
0.05	-0.05	0.105	0.045	0.114	0.043
0.05	0.05	0.264	0.050	0.274	0.050
0.05	0.10	0.344	0.049	0.364	0.042
0.05	0.30	0.477	0.049	0.484	0.047
0.10	-0.30	0.230	0.032	0.220	0.035
0.10	-0.10	0.157	0.044	0.153	0.042
0.10	-0.05	0.328	0.047	0.326	0.043
0.10	0.05	0.713	0.056	0.732	0.050
0.10	0.10	0.821	0.048	0.821	0.049
0.10	0.30	0.912	0.178	0.918	0.187
0.30	-0.30	0.866	0.028	0.858	0.030
0.30	-0.10	0.783	0.170	0.789	0.170
0.30	-0.05	0.977	0.194	0.974	0.204
0.30	0.05	1.000	0.231	1.000	0.241
0.30	0.10	1.000	0.350	1.000	0.351
0.30	0.30	1.000	1.000	1.000	1.000

270 For all our proposed tests, the power curves can be generated in several ways. For example,  
the power curves can be generated when the null model is the 2WE model, and the alternative can  
272 be one of the DPD, SSPD, SDPDW and SDPD model. We will refer to this as Case 1. However,  
this approach would yield several plots, for instance, 216 plots for the 2WE–SDPD combination.  
274 To save space, we instead summarize the results in Tables 6 through 8, where the level for all tests  
is 5%. As we mentioned in the Monte Carlo design, the DPD, SSPD and SDPDW models can be

276 considered as null models for one-directional tests and their robust counterparts. Therefore, we can  
 278 generate size power curves for these one directional tests, where the null model is one of the DPD,  
 SSPD and SDPDW models and the alternative model is one of the SDPDW and SDPD models.  
 We will refer to this as Case 2. For example, we could investigate the size power curves for  $LM_\lambda$   
 280 and  $LM_\lambda^*$  where the null model is the DPD model and the alternative model is SDPDW model.  
 Similarly, for  $LM_\lambda$  and  $LM_\lambda^*$ , the null of the DPD and the alternative of the SDPD would yield  
 282 another size power curve. We chose to present some representative cases in Figures 2 and 3.<sup>10</sup>

284 The general observations from Tables 6 through 8 on the power properties of our proposed tests  
 for Case 1 are listed as follows. To save space, we only present the normally distributed error case,  
 as the results for the gamma distributed error case are similar. Also, for the case of the SDPD  
 286 model, we focus on some representative tables.

Table 6: Power of tests when  $H_1$ : The DPD/SSPD model and  $H_0$ : The 2WE model

$\gamma_0/\lambda_0$	$LM_\rho$	$LM_\rho^*$	$LM_\lambda$	$LM_\lambda^*$	$LM_\gamma$	$LM_\gamma^*$	$LM_J$	$C_J$
H <sub>1</sub> : The DPD model								
-0.30	0.046	0.015	0.042	0.005	1.000	1.000	1.000	1.000
-0.10	0.044	0.038	0.042	0.039	0.550	0.536	0.376	0.374
-0.05	0.040	0.049	0.048	0.051	0.178	0.171	0.114	0.113
0.05	0.061	0.046	0.061	0.056	0.236	0.231	0.149	0.144
0.10	0.074	0.042	0.064	0.039	0.634	0.616	0.454	0.454
0.30	0.135	0.028	0.100	0.024	1.000	1.000	1.000	1.000
H <sub>1</sub> : The SSPD model								
-0.30	1.000	0.791	1.000	1.000	1.000	0.999	1.000	1.000
-0.10	0.913	0.051	0.993	0.810	0.334	0.184	0.975	0.973
-0.05	0.394	0.053	0.600	0.303	0.087	0.071	0.443	0.431
0.05	0.326	0.048	0.593	0.343	0.077	0.068	0.426	0.421
0.10	0.853	0.051	0.992	0.844	0.204	0.136	0.965	0.962
0.30	1.000	0.730	1.000	1.000	0.998	0.997	1.000	1.000

- 288 1. Table 6 shows that the joint test statistics and the one directional test statistics,  $LM_\gamma$ ,  $LM_\gamma^*$   
 in the case of the DPD model and  $LM_\lambda$ ,  $LM_\lambda^*$  in the case of the SSPD model, have desirable  
 power.<sup>11</sup>
- 290 2. In Table 6, the robust versions of the one directional tests generally perform similar to their  
 292 non-robust counterparts. However, as the value of  $\gamma_0$  increases in the DPD model for example,  
 we see that the rejection frequencies of  $LM_\rho^*$  remain low whereas  $LM_\rho$  over rejects the true null,  
 294 confirming the (over) size problem in Table 2. A similar finding applies to  $LM_\lambda^*$ . Therefore, in  
 case of temporal dependence in the data generating process, the robust tests are preferable.  
 In the case of the SSPD model in Table 6,  $LM_\gamma^*$  and  $LM_\rho^*$  report relatively smaller rejection  
 296 frequencies and hence perform better than the non-robust counterparts. Again, in case of  
 spatial dependence in the data generating process, the robust tests are preferable.

<sup>10</sup>We only present results based on the rook weight matrix for the power analysis. The results based on the queen weight matrix are available upon request.

<sup>11</sup>Note that the one directional tests and their robust counterparts for  $\lambda$  and  $\rho$  should have lower rejection frequencies for the case where  $H_1$ : The DPD model and  $H_0$ : The 2WE model. Similarly, the one directional tests and their robust counterparts for  $\gamma$  and  $\rho$  should report lower rejection frequencies for the case where  $H_1$ : The SSPD model and  $H_0$ : The 2WE model.

- 298 3. Table 7 reveals similar findings. The joint test statistics and the one directional test statistics,  
 300  $LM_\gamma$ ,  $LM_\gamma^*$  and  $LM_\lambda$ ,  $LM_\lambda^*$ , have desirable power.  $LM_\rho^*$ 's rejection frequency remains low for  
 302 smaller deviations of  $\lambda_0$  and  $\gamma_0$  from zero, whereas  $LM_\rho$  over rejects the true null, confirming  
 the (over) size problem in Table 4. Therefore, in case of spatial and temporal dependence in  
 the data generating process, the robust tests are preferable.
4. Tables 8, 9 and 10 shows that all one directional tests and the joint tests have proper power.  
 304 The non-robust tests have higher power relative to their robust counterparts in some cases  
 but the differences are generally negligible.

306 For all our proposed tests, the power curves can be generated in several ways in Case 2. First,  
 one can obviously consider the 2WE model as the null model and the alternative can be one of the  
 308 DPD, SSP, SDPDW and SDPD models. We will not generate size power curves for these cases as  
 we already summarized the results in Tables 6 through 10. Furthermore, for the one directional  
 310 tests and their robust versions, one of the DPD, SSPD and SDPDW models can be the null model  
 and one of the SDPDW and SDPD models as the alternative model. For example, we can generate  
 312 a size power curve for  $LM_\lambda$  and  $LM_\lambda^*$  using the DPD model as the null model and the SDPDW  
 model as the alternative. Another size power curve for  $LM_\lambda$  and  $LM_\lambda^*$  can be obtained from the  
 314 DPD model as the null model and the SDPD model as the alternative.<sup>12</sup>

In Figures 2 and 3, the lines with the red color correspond to the non-robust one directional  
 316 test whereas the lines with blue color correspond to their robust counterparts. Different markers  
 are used to identify varying true values of the spatial parameter in the corresponding alternative  
 318 model. The general observations on the power properties of our proposed tests are listed as follows.

---

<sup>12</sup>The experiments based on the gamma distributed errors are not presented, because results are similar to the  
 experiments based on the normally distributed errors.

Table 7: Power of tests when  $H_1$ : The SDPDW model and  $H_0$ : The 2WE model

$\lambda_0$	$\gamma_0$	$LM_\rho$	$LM_\rho^*$	$LM_\lambda$	$LM_\lambda^*$	$LM_\gamma$	$LM_\gamma^*$	$LM_J$	$C_J$
-0.30	-0.30	0.580	0.299	1.000	1.000	0.979	0.201	1.000	1.000
-0.30	-0.10	0.999	0.833	1.000	1.000	0.964	0.984	1.000	1.000
-0.30	-0.05	1.000	0.831	1.000	1.000	1.000	0.999	1.000	1.000
-0.30	0.05	1.000	0.772	1.000	0.995	1.000	1.000	1.000	1.000
-0.30	0.10	1.000	0.885	1.000	0.992	1.000	1.000	1.000	1.000
-0.30	0.30	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
-0.10	-0.30	0.120	0.045	0.946	0.877	1.000	1.000	1.000	1.000
-0.10	-0.10	0.466	0.039	0.983	0.936	0.274	0.283	0.978	0.980
-0.10	-0.05	0.734	0.046	0.990	0.901	0.088	0.075	0.962	0.963
-0.10	0.05	0.971	0.048	0.995	0.628	0.754	0.525	0.989	0.987
-0.10	0.10	0.990	0.047	0.999	0.483	0.965	0.864	0.998	0.998
-0.10	0.30	1.000	0.739	1.000	0.945	1.000	1.000	1.000	1.000
-0.05	-0.30	0.073	0.025	0.404	0.212	1.000	1.000	1.000	1.000
-0.05	-0.10	0.128	0.044	0.512	0.394	0.481	0.477	0.646	0.663
-0.05	-0.05	0.242	0.050	0.553	0.374	0.128	0.134	0.466	0.473
-0.05	0.05	0.515	0.051	0.651	0.230	0.380	0.296	0.607	0.582
-0.05	0.10	0.612	0.049	0.703	0.162	0.777	0.687	0.827	0.808
-0.05	0.30	0.819	0.219	0.857	0.395	1.000	1.000	1.000	1.000
0.05	-0.30	0.068	0.018	0.457	0.246	1.000	1.000	1.000	1.000
0.05	-0.10	0.121	0.046	0.531	0.411	0.495	0.469	0.652	0.675
0.05	-0.05	0.207	0.049	0.545	0.373	0.154	0.143	0.457	0.473
0.05	0.05	0.474	0.051	0.613	0.244	0.307	0.271	0.566	0.547
0.05	0.10	0.557	0.042	0.629	0.161	0.726	0.685	0.789	0.774
0.05	0.30	0.598	0.022	0.657	0.169	1.000	1.000	1.000	1.000
0.10	-0.30	0.133	0.031	0.949	0.881	1.000	0.999	1.000	1.000
0.10	-0.10	0.360	0.042	0.986	0.948	0.333	0.286	0.979	0.984
0.10	-0.05	0.639	0.053	0.986	0.914	0.082	0.077	0.957	0.961
0.10	0.05	0.956	0.054	0.992	0.706	0.617	0.481	0.984	0.981
0.10	0.10	0.985	0.045	0.995	0.520	0.912	0.828	0.995	0.994
0.10	0.30	0.990	0.151	0.998	0.761	1.000	1.000	1.000	1.000
0.30	-0.30	0.763	0.296	1.000	1.000	0.998	0.228	1.000	1.000
0.30	-0.10	0.976	0.746	1.000	1.000	0.671	0.902	1.000	1.000
0.30	-0.05	1.000	0.757	1.000	1.000	0.961	0.976	1.000	1.000
0.30	0.05	1.000	0.643	1.000	1.000	1.000	1.000	1.000	1.000
0.30	0.10	1.000	0.637	1.000	0.999	1.000	1.000	1.000	1.000
0.30	0.30	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 8: Power of tests when  $H_1$ :The SDPD model and  $H_0$ : The 2WE model

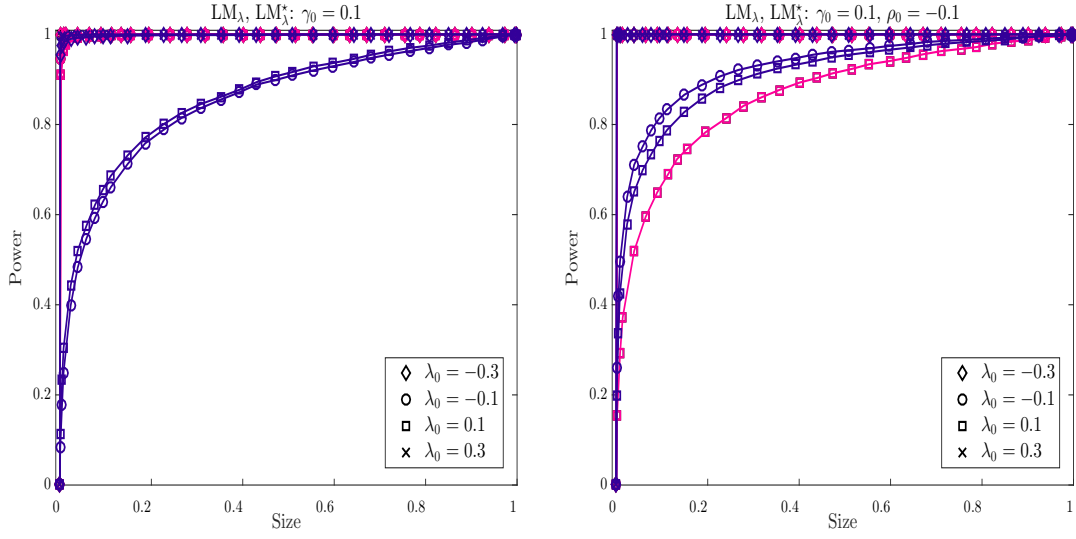
<i>Normal distribution</i>										
$\lambda_0$	$\gamma_0$	$\rho_0$	$LM_\rho$	$LM_\rho^*$	$LM_\lambda$	$LM_\lambda^*$	$LM_\gamma$	$LM_\gamma^*$	$LM_J$	$C_J$
0.05	-0.30	-0.30	1.000	0.964	0.631	0.627	1.000	1.000	1.000	1.000
0.05	-0.30	-0.10	0.861	0.198	0.464	0.020	1.000	1.000	1.000	1.000
0.05	-0.30	-0.05	0.390	0.040	0.430	0.081	1.000	1.000	1.000	1.000
0.05	-0.30	0.05	0.188	0.138	0.478	0.486	1.000	1.000	1.000	1.000
0.05	-0.30	0.10	0.688	0.464	0.496	0.737	1.000	1.000	1.000	1.000
0.05	-0.30	0.30	1.000	0.999	0.503	1.000	1.000	1.000	1.000	1.000
0.05	-0.10	-0.30	1.000	0.991	0.424	0.068	0.929	0.930	1.000	1.000
0.05	-0.10	-0.10	0.446	0.375	0.137	0.167	0.830	0.574	0.796	0.789
0.05	-0.10	-0.05	0.073	0.121	0.287	0.276	0.668	0.509	0.629	0.640
0.05	-0.10	0.05	0.564	0.129	0.758	0.510	0.413	0.486	0.852	0.859
0.05	-0.10	0.10	0.938	0.383	0.917	0.617	0.406	0.533	0.981	0.981
0.05	-0.10	0.30	1.000	0.987	1.000	0.951	0.553	0.891	1.000	1.000
0.05	-0.05	-0.30	1.000	0.991	0.764	0.027	0.547	0.669	1.000	1.000
0.05	-0.05	-0.10	0.292	0.380	0.098	0.220	0.462	0.210	0.530	0.519
0.05	-0.05	-0.05	0.049	0.141	0.262	0.304	0.261	0.156	0.354	0.362
0.05	-0.05	0.05	0.700	0.134	0.824	0.440	0.111	0.141	0.773	0.778
0.05	-0.05	0.10	0.966	0.377	0.964	0.466	0.145	0.185	0.968	0.969
0.05	-0.05	0.30	1.000	0.976	1.000	0.816	0.308	0.588	1.000	1.000
0.05	0.05	-0.30	1.000	0.965	0.997	0.029	0.759	0.098	1.000	1.000
0.05	0.05	-0.10	0.146	0.332	0.069	0.265	0.118	0.166	0.277	0.277
0.05	0.05	-0.05	0.098	0.129	0.221	0.277	0.178	0.235	0.298	0.289
0.05	0.05	0.05	0.883	0.115	0.922	0.186	0.512	0.340	0.876	0.866
0.05	0.05	0.10	0.991	0.322	0.991	0.157	0.633	0.341	0.987	0.986
0.05	0.05	0.30	1.000	0.841	1.000	0.553	0.944	0.225	1.000	1.000
0.05	0.10	-0.30	1.000	0.873	1.000	0.040	0.989	0.490	1.000	1.000
0.05	0.10	-0.10	0.157	0.238	0.084	0.218	0.490	0.497	0.470	0.464
0.05	0.10	-0.05	0.136	0.098	0.204	0.219	0.559	0.610	0.543	0.538
0.05	0.10	0.05	0.929	0.094	0.940	0.129	0.859	0.727	0.964	0.959
0.05	0.10	0.10	0.996	0.218	0.997	0.112	0.931	0.743	0.997	0.996
0.05	0.10	0.30	1.000	0.574	1.000	0.608	0.999	0.712	1.000	1.000
0.05	0.30	-0.30	1.000	0.173	1.000	0.495	1.000	1.000	1.000	1.000
0.05	0.30	-0.10	0.690	0.037	0.368	0.028	1.000	1.000	1.000	1.000
0.05	0.30	-0.05	0.125	0.015	0.120	0.035	1.000	1.000	1.000	1.000
0.05	0.30	0.05	0.985	0.121	0.983	0.507	1.000	1.000	1.000	1.000
0.05	0.30	0.10	1.000	0.330	1.000	0.796	1.000	1.000	1.000	1.000
0.05	0.30	0.30	1.000	0.452	1.000	0.972	1.000	1.000	1.000	1.000

Table 9: Power of tests when  $H_1$ : The SDPD model and  $H_0$ : The 2WE model

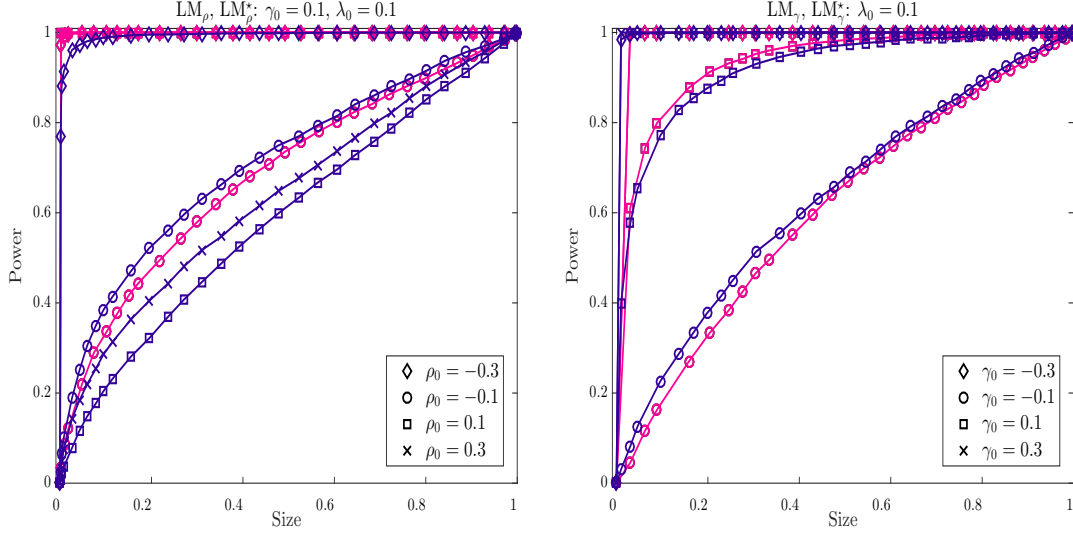
$\lambda_0$	$\gamma_0$	$\rho_0$	$LM_\rho$	$LM_\rho^*$	$LM_\lambda$	$LM_\lambda^*$	$LM_\gamma$	$LM_\gamma^*$	$LM_J$	$C_J$
0.10	-0.30	-0.30	1.000	0.883	0.976	0.137	1.000	1.000	1.000	1.000
0.10	-0.30	-0.10	0.925	0.114	0.954	0.406	1.000	1.000	1.000	1.000
0.10	-0.30	-0.05	0.556	0.016	0.948	0.685	1.000	1.000	1.000	1.000
0.10	-0.30	0.05	0.121	0.191	0.956	0.966	1.000	1.000	1.000	1.000
0.10	-0.30	0.10	0.579	0.574	0.963	0.993	0.999	1.000	1.000	1.000
0.10	-0.30	0.30	1.000	1.000	0.959	1.000	0.995	1.000	1.000	1.000
0.10	-0.10	-0.30	1.000	0.990	0.210	0.077	0.998	0.927	1.000	1.000
0.10	-0.10	-0.10	0.228	0.385	0.815	0.779	0.867	0.405	0.950	0.956
0.10	-0.10	-0.05	0.069	0.154	0.935	0.890	0.619	0.317	0.944	0.952
0.10	-0.10	0.05	0.848	0.113	0.997	0.963	0.176	0.286	0.995	0.995
0.10	-0.10	0.10	0.991	0.340	0.999	0.975	0.159	0.339	1.000	1.000
0.10	-0.10	0.30	1.000	0.975	1.000	0.996	0.336	0.729	1.000	1.000
0.10	-0.05	-0.30	1.000	0.992	0.140	0.138	0.920	0.706	1.000	1.000
0.10	-0.05	-0.10	0.084	0.417	0.758	0.809	0.522	0.117	0.846	0.854
0.10	-0.05	-0.05	0.167	0.168	0.926	0.883	0.218	0.081	0.879	0.887
0.10	-0.05	0.05	0.954	0.101	0.998	0.917	0.133	0.087	0.995	0.995
0.10	-0.05	0.10	0.999	0.308	1.000	0.926	0.250	0.115	1.000	1.000
0.10	-0.05	0.30	1.000	0.928	1.000	0.968	0.647	0.373	1.000	1.000
0.10	0.05	-0.30	0.997	0.989	0.645	0.310	0.271	0.078	0.995	0.994
0.10	0.05	-0.10	0.186	0.353	0.648	0.759	0.087	0.242	0.684	0.672
0.10	0.05	-0.05	0.664	0.150	0.928	0.753	0.290	0.360	0.892	0.880
0.10	0.05	0.05	0.999	0.091	1.000	0.627	0.836	0.560	0.999	0.999
0.10	0.05	0.10	1.000	0.214	1.000	0.554	0.926	0.598	1.000	1.000
0.10	0.05	0.30	1.000	0.503	1.000	0.919	0.998	0.512	1.000	1.000
0.10	0.10	-0.30	0.999	0.972	0.896	0.305	0.782	0.257	0.999	0.998
0.10	0.10	-0.10	0.290	0.250	0.595	0.650	0.400	0.588	0.771	0.757
0.10	0.10	-0.05	0.793	0.094	0.919	0.599	0.685	0.730	0.950	0.942
0.10	0.10	0.05	1.000	0.059	1.000	0.459	0.980	0.872	1.000	1.000
0.10	0.10	0.10	1.000	0.117	1.000	0.474	0.995	0.894	1.000	1.000
0.10	0.10	0.30	1.000	0.183	1.000	0.961	1.000	0.908	1.000	1.000
0.10	0.30	-0.30	1.000	0.206	1.000	0.058	1.000	1.000	1.000	1.000
0.10	0.30	-0.10	0.126	0.019	0.311	0.187	1.000	1.000	1.000	1.000
0.10	0.30	-0.05	0.702	0.028	0.891	0.419	1.000	1.000	1.000	1.000
0.10	0.30	0.05	1.000	0.519	1.000	0.956	1.000	1.000	1.000	1.000
0.10	0.30	0.10	1.000	0.808	1.000	0.994	1.000	1.000	1.000	1.000
0.10	0.30	0.30	1.000	0.877	1.000	1.000	1.000	1.000	1.000	1.000

Table 10: Power of tests when  $H_1$ : The SDPD model and  $H_0$ : The 2WE model

<i>Normal distribution</i>										
$\lambda_0$	$\gamma_0$	$\rho_0$	$LM_\rho$	$LM_\rho^*$	$LM_\lambda$	$LM_\lambda^*$	$LM_\gamma$	$LM_\gamma^*$	$LM_J$	$C_J$
0.30	-0.30	-0.30	1.000	0.766	1.000	0.978	1.000	0.927	1.000	1.000
0.30	-0.30	-0.10	0.999	0.656	1.000	1.000	1.000	0.349	1.000	1.000
0.30	-0.30	-0.05	0.977	0.496	1.000	1.000	1.000	0.269	1.000	1.000
0.30	-0.30	0.05	0.321	0.138	1.000	1.000	0.941	0.233	1.000	1.000
0.30	-0.30	0.10	0.327	0.088	1.000	1.000	0.638	0.235	1.000	1.000
0.30	-0.30	0.30	1.000	0.895	1.000	1.000	0.820	0.450	1.000	1.000
0.30	-0.10	-0.30	1.000	1.000	1.000	1.000	1.000	0.123	1.000	1.000
0.30	-0.10	-0.10	0.296	0.964	1.000	1.000	0.605	0.764	1.000	1.000
0.30	-0.10	-0.05	0.718	0.908	1.000	1.000	0.291	0.865	1.000	1.000
0.30	-0.10	0.05	1.000	0.519	1.000	1.000	0.960	0.913	1.000	1.000
0.30	-0.10	0.10	1.000	0.272	1.000	1.000	0.998	0.920	1.000	1.000
0.30	-0.10	0.30	1.000	0.175	1.000	1.000	1.000	0.892	1.000	1.000
0.30	-0.05	-0.30	0.985	1.000	1.000	1.000	1.000	0.117	1.000	1.000
0.30	-0.05	-0.10	0.734	0.962	1.000	1.000	0.284	0.927	1.000	1.000
0.30	-0.05	-0.05	0.975	0.901	1.000	1.000	0.646	0.967	1.000	1.000
0.30	-0.05	0.05	1.000	0.530	1.000	1.000	0.999	0.982	1.000	1.000
0.30	-0.05	0.10	1.000	0.303	1.000	1.000	1.000	0.983	1.000	1.000
0.30	-0.05	0.30	1.000	0.447	1.000	0.999	1.000	0.985	1.000	1.000
0.30	0.05	-0.30	0.377	0.998	1.000	1.000	0.940	0.373	1.000	1.000
0.30	0.05	-0.10	1.000	0.863	1.000	1.000	0.941	0.998	1.000	1.000
0.30	0.05	-0.05	1.000	0.779	1.000	1.000	0.998	0.999	1.000	1.000
0.30	0.05	0.05	1.000	0.551	1.000	1.000	1.000	1.000	1.000	1.000
0.30	0.05	0.10	1.000	0.589	1.000	0.999	1.000	1.000	1.000	1.000
0.30	0.05	0.30	1.000	0.994	1.000	1.000	1.000	1.000	1.000	1.000
0.30	0.10	-0.30	0.280	0.988	1.000	1.000	0.531	0.666	1.000	1.000
0.30	0.10	-0.10	1.000	0.659	1.000	1.000	0.998	1.000	1.000	1.000
0.30	0.10	-0.05	1.000	0.621	1.000	1.000	1.000	1.000	1.000	1.000
0.30	0.10	0.05	1.000	0.747	1.000	0.999	1.000	1.000	1.000	1.000
0.30	0.10	0.10	1.000	0.891	1.000	0.999	1.000	1.000	1.000	1.000
0.30	0.10	0.30	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.30	0.30	-0.30	0.516	0.674	0.999	1.000	1.000	1.000	1.000	1.000
0.30	0.30	-0.10	1.000	0.895	1.000	1.000	1.000	1.000	1.000	1.000
0.30	0.30	-0.05	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000
0.30	0.30	0.05	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.30	0.30	0.10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.30	0.30	0.30	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000



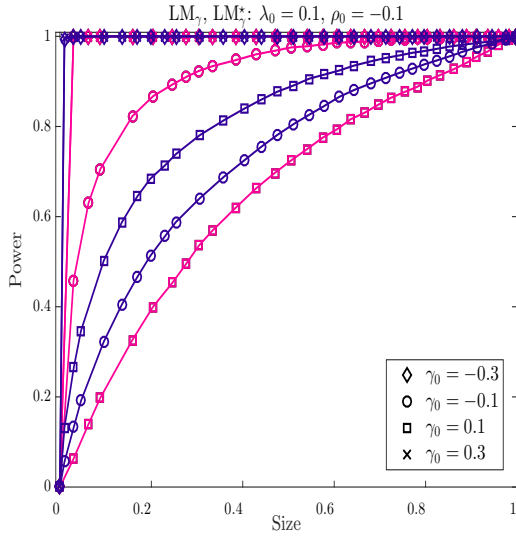
(a)  $H_0$ : The DPD model,  $H_1$ : The SDPDW model      (b)  $H_0$ : The DPD model,  $H_1$ : The SDPD model



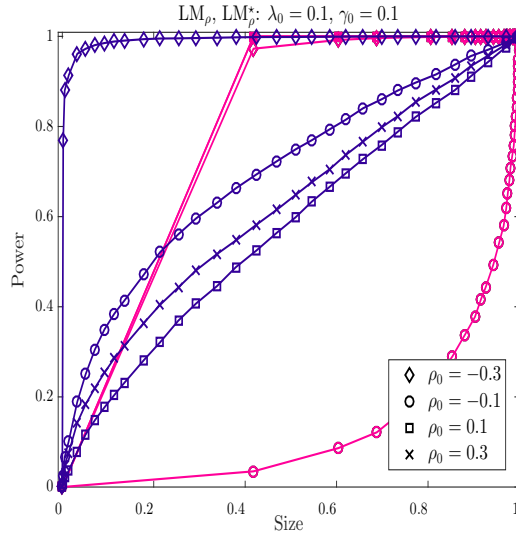
(c)  $H_0$ : The DPD model,  $H_1$ : The SDPD model      (d)  $H_0$ : The SSPD model,  $H_1$ : The SDPDW model

Figure 2: Size-power curves

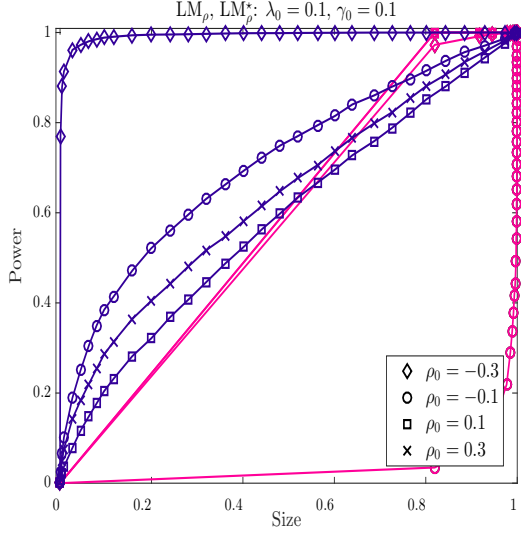




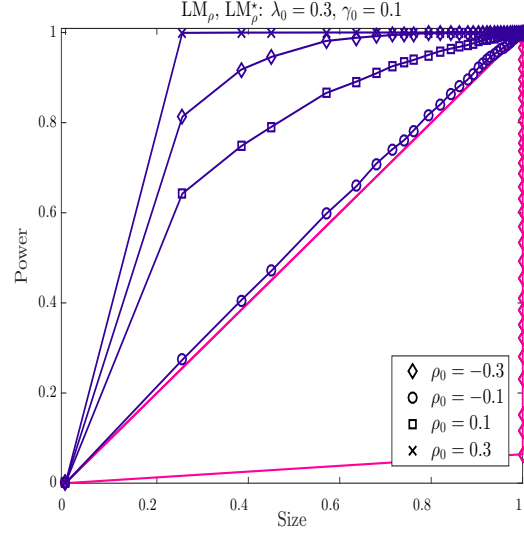
(a)  $H_0$ : The SSPD model,  $H_1$ : The SDPD model



(b)  $H_0$ : The SSPD model,  $H_1$ : The SDPD model



(c)  $H_0$ : The SDPDW model,  $H_1$ : The SDPD model



(d)  $H_0$ : The SDPDW model,  $H_1$ : The SDPD model

Figure 3: Size-power curves

- 320 1. In Figure 2(a), the null model is the DPD model and the alternative model is the SDPDW  
 322 model. Both  $LM_\lambda$  and  $LM_\lambda^*$  has satisfactory power. For lower values of  $\lambda_0$ ,  $LM_\lambda^*$  is less  
 324 powerful than  $LM_\lambda$ . In Figure 2(b), the null model is the DPD model and the alternative  
 326 model is the SDPD model. Generally,  $LM_\lambda^*$  is less powerful than  $LM_\lambda$  except for the case  
 328 where  $\lambda_0 = 0.1$ .
- 324 2. In Figure 2(c), the null model is the DPD model and the alternative model is again the  
 326 SDPD model.  $LM_\rho^*$  is slightly less powerful than  $LM_\rho$  except for the case where  $\rho_0 = -0.1$ .  
 328 In Figure 2(d), the null model is the SSPD model and the alternative model is again the  
 330 SDPDW model.  $LM_\gamma^*$  and  $LM_\gamma$  behave similarly and both lack power when  $\gamma_0 = -0.1$ .
- 328 3. In Figure 3(a), the null model is the SSPD model and the alternative model is again the  
 330 SDPD model. Generally,  $LM_\gamma^*$  and  $LM_\gamma$  behave similarly. We see that when  $\gamma_0 = 0.1$ ,  $LM_\gamma^*$   
 332 is more powerful than  $LM_\gamma$ . But this picture reverses when  $\gamma_0 = -0.1$ .
- 332 4. In Figure 3(b), the null model is the SSPD model and the alternative model is again the  
 334 SDPD model. It confirms the results on the one directional tests of  $\rho_0$  from Table 3.  $LM_\rho$   
 336 over rejects when the true model involves dependence over space and time. Furthermore,  
 338 when spatial time lag coefficient is small on the negative side,  $LM_\rho$  suffers from positive size  
 340 distortion and lack of power. Surprisingly though,  $LM_\rho^*$  lack power when  $\rho_0 = 0.3$ .
- 336 5. In Figures 3(c) and 3(d), the null model is the SDPDW model and the alternative model is the  
 338 SDPD model. It confirms the results on the one directional tests of  $\rho_0$  from Table 4. Clearly,  
 340  $LM_\rho$  over rejects when the true model involves dependence over space and time. Again, we  
 see that  $LM_\rho^*$  lack power when  $\rho_0 = 0.3$ . But, it does not suffer from size distortions unless  
 the misspecification in the alternative becomes larger.

## 6 Conclusion

342 In this paper, we introduce the robust LM tests within the GMM framework for a spatial dynamic  
 344 panel data model. These tests are robust in the sense that their asymptotic distributions under the  
 346 null hypothesis are still a central chi-square distribution when the alternative model is misspecified.  
 On the other hand, when the alternative model is misspecified, the asymptotic null distributions  
 348 of the standard LM tests deviate from the central chi-square distributions. Hence, the robust tests  
 obtain asymptotically the correct size. We derive the asymptotic distributions of our proposed tests  
 350 under the null and the local alternative hypotheses. These tests can be used to test the presence of  
 the contemporaneous dependence over space, dependence over time and spatial time dependence.

352 One attractive feature of our proposed tests is that their test statistics are easy to compute and  
 only require the estimates from a two-way error model. Therefore, our proposed tests can easily  
 354 be made available for the practical applications by using the standard statistical softwares. In a  
 Monte Carlo study, we investigate the size and power properties of our proposed tests. Our results  
 356 shows that the robust tests have good finite sample properties and would be useful for the detection  
 of the source of dependence in a spatial dynamic panel data model. The simulation results, hence,  
 358 confirm our analytical results that the robust tests are valid, when the alternative models locally  
 deviate from the the true data generating process.

# Appendix

## A A Useful Lemma

**Lemma 1.** — Under our stated assumptions, the following results hold.

1.  $\frac{1}{N} \mathbb{E} \left( g_{nT}(\theta_0) g'_{nT}(\theta_0) \right) = \Sigma_{nT} + o(1)$  and  $\widehat{\Sigma}_{nT} = \Sigma_{nT} + o_p(1)$ , where  $\widehat{\Sigma}_{nT}$  and  $\Sigma_{nT}$  are stated in the main text.
2.  $G(\widehat{\theta}_{nT}) = D_{nT} + R_{nT} + O\left(\frac{1}{\sqrt{nT}}\right)$ , where  $D_{nT}$  is  $O(1)$ ,  $R_{nT}$  is  $O\left(\frac{1}{T}\right)$  and  $\widehat{\theta}_{nT}$  is any consistent estimator of  $\theta_0$ .
3.  $G(\widehat{\theta}_{nT}) \widehat{\Sigma}_{nT} G(\widehat{\theta}_{nT}) = (D_{nT} + R_{nT})' \Sigma_{nT} (D_{nT} + R_{nT}) + o_p(1)$ , where  $\widehat{\theta}_{nT}$  is any consistent estimator of  $\theta_0$ .
4. Let  $a_{nT}$  be a  $k_a \times (m + q)$  non-stochastic matrix. Then

$$\frac{1}{\sqrt{N}} a_{nT} g_{nT}(\theta_0) \xrightarrow{d} N\left(0, \text{plim}_{n \rightarrow \infty} a_{nT} \Sigma_{nT} a'_{nT}\right) \quad (\text{A.1})$$

*Proof.* See Lee and Yu (2014). □

## B Expressions for Test Statistics

In this section, we provide explicit expressions for the elements of test statistics. Let the  $j$ th column of  $G_a(\theta)$  be denoted by  $G_a(\theta)[:, j]$ . We start with  $G(\theta) = (G_\lambda(\theta), G_\gamma(\theta), G_\rho(\theta), G_\beta(\theta))$ , where

$$G_\lambda(\theta)[:, j] = -\frac{1}{N} \begin{pmatrix} \mathbf{Y}_{n,T-1}^{*'} \mathbf{W}'_{nj,T-1} \mathbf{J}_{n,T-1} \mathbf{P}_{n1,T-1}^s \mathbf{J}_{n,T-1} \mathbf{V}_{n,T-1}^*(\theta) \\ \mathbf{Y}_{n,T-1}^{*'} \mathbf{W}'_{nj,T-1} \mathbf{J}_{n,T-1} \mathbf{P}_{n2,T-1}^s \mathbf{J}_{n,T-1} \mathbf{V}_{n,T-1}^*(\theta) \\ \vdots \\ \mathbf{Y}_{n,T-1}^{*'} \mathbf{W}'_{nj,T-1} \mathbf{J}_{n,T-1} \mathbf{P}_{nm,T-1}^s \mathbf{J}_{n,T-1} \mathbf{V}_{n,T-1}^*(\theta) \\ \mathbf{Q}_{n,T-1} \mathbf{J}_{n,T-1} \mathbf{W}_{nj,T-1} \mathbf{Y}_{n,T-1}^* \end{pmatrix}. \quad (\text{B.1})$$

$$G_\gamma(\theta) = -\frac{1}{N} \begin{pmatrix} \mathbf{V}_{n,T-1}^{*'}(\theta) \mathbf{J}_{n,T-1} \mathbf{P}_{n1,T-1}^s \mathbf{J}_{n,T-1} \mathbf{Y}_{n,T-1}^{(*,-1)} \\ \mathbf{V}_{n,T-1}^{*'}(\theta) \mathbf{J}_{n,T-1} \mathbf{P}_{n2,T-1}^s \mathbf{J}_{n,T-1} \mathbf{Y}_{n,T-1}^{(*,-1)} \\ \vdots \\ \mathbf{V}_{n,T-1}^{*'}(\theta) \mathbf{J}_{n,T-1} \mathbf{P}_{nm,T-1}^s \mathbf{J}_{n,T-1} \mathbf{Y}_{n,T-1}^{(*,-1)} \\ \mathbf{Q}'_{n,T-1} \mathbf{J}_{n,T-1} \mathbf{Y}_{n,T-1}^{(*,-1)} \end{pmatrix}. \quad (\text{B.2})$$

$$G_\rho(\theta)[:, j] = -\frac{1}{N} \begin{pmatrix} \mathbf{V}_{n,T-1}^{*'}(\theta) \mathbf{J}_{n,T-1} \mathbf{P}_{n1,T-1}^s \mathbf{J}_{n,T-1} \mathbf{W}_{nj,T-1} \mathbf{Y}_{n,T-1}^{(*,-1)} \\ \mathbf{V}_{n,T-1}^{*'}(\theta) \mathbf{J}_{n,T-1} \mathbf{P}_{n2,T-1}^s \mathbf{J}_{n,T-1} \mathbf{W}_{nj,T-1} \mathbf{Y}_{n,T-1}^{(*,-1)} \\ \vdots \\ \mathbf{V}_{n,T-1}^{*'}(\theta) \mathbf{J}_{n,T-1} \mathbf{P}_{nm,T-1}^s \mathbf{J}_{n,T-1} \mathbf{W}_{nj,T-1} \mathbf{Y}_{n,T-1}^{(*,-1)} \\ \mathbf{Q}'_{n,T-1} \mathbf{J}_{n,T-1} \mathbf{W}_{nj,T-1} \mathbf{Y}_{n,T-1}^{(*,-1)} \end{pmatrix}. \quad (\text{B.3})$$

$$G_\beta(\theta) = -\frac{1}{N} \begin{pmatrix} \mathbf{V}_{n,T-1}'(\theta) \mathbf{J}_{n,T-1} \mathbf{P}_{n1,T-1}^s \mathbf{J}_{n,T-1} \mathbf{X}_{n,T-1}^* \\ \mathbf{V}_{n,T-1}'(\theta) \mathbf{J}_{n,T-1} \mathbf{P}_{n2,T-1}^s \mathbf{J}_{n,T-1} \mathbf{X}_{n,T-1}^* \\ \vdots \\ \mathbf{V}_{n,T-1}'(\theta) \mathbf{J}_{n,T-1} \mathbf{P}_{nm,T-1}^s \mathbf{J}_{n,T-1} \mathbf{X}_{n,T-1}^* \\ \mathbf{Q}'_{n,T-1} \mathbf{J}_{n,T-1} \mathbf{X}_{n,T-1}^* \end{pmatrix}. \quad (\text{B.4})$$

Using the inverse of the partitioned matrix formula (Amemiya 1985, p.460), we have

$$\begin{aligned} \widehat{\Sigma}_{nT}^{-1} &= \begin{pmatrix} \frac{1}{N} \left[ \widehat{\sigma}^4 \Delta_{nm,T} + (\widehat{\mu}_4 - 3\widehat{\sigma}^4) \omega'_{nm,T} \omega_{nm,T} \right] & 0_{m \times q} \\ 0_{q \times m} & \widehat{\sigma}^2 \frac{1}{N} \mathbf{Q}'_{n,T-1} \mathbf{J}_{n,T-1} \mathbf{Q}_{n,T-1} \end{pmatrix}^{-1} \\ &= \begin{pmatrix} O_{11} & O_{12} \\ O_{21} & O_{22} \end{pmatrix}, \end{aligned} \quad (\text{B.5})$$

where  $O_{11} = N[\widehat{\sigma}^4 \Delta_{nm,T} + (\widehat{\mu}_4 - 3\widehat{\sigma}^4) \omega'_{nm,T} \omega_{nm,T}]^{-1}$ ,  $O_{12} = O'_{21} = 0_{m \times q}$ , and  $O_{22} = \frac{N}{\widehat{\sigma}^2} [\mathbf{Q}'_{n,T-1} \mathbf{J}_{n,T-1} \mathbf{Q}_{n,T-1}]^{-1}$ . The component of  $C(\theta)$  are given by

$$1. \quad C_\lambda(\theta) = G'_\lambda(\theta) \begin{pmatrix} O_{11} & O_{12} \\ O_{21} & O_{22} \end{pmatrix} \bar{g}_{nT}(\theta), \quad C_\gamma(\theta) = G'_\gamma(\theta) \begin{pmatrix} O_{11} & O_{12} \\ O_{21} & O_{22} \end{pmatrix} \bar{g}_{nT}(\theta) \quad (\text{B.6})$$

$$2. \quad C_\rho(\theta) = G'_\rho(\theta) \begin{pmatrix} O_{11} & O_{12} \\ O_{21} & O_{22} \end{pmatrix} \bar{g}_{nT}(\theta), \quad C_\beta(\theta) = G'_\beta(\theta) \begin{pmatrix} O_{11} & O_{12} \\ O_{21} & O_{22} \end{pmatrix} \bar{g}_{nT}(\theta) \quad (\text{B.7})$$

The components of  $B(\theta)$  are defined in below.

$$\begin{aligned} 1. & B_\lambda(\theta) = G'_\lambda(\theta) \begin{pmatrix} O_{11} & O_{12} \\ O_{21} & O_{22} \end{pmatrix} G_\lambda, \quad B_{\lambda\rho}(\theta) = B'_{\rho\lambda}(\theta) = G'_\lambda(\theta) \begin{pmatrix} O_{11} & O_{12} \\ O_{21} & O_{22} \end{pmatrix} G_\rho \\ 2. & B_{\lambda\gamma}(\theta) = B'_{\gamma\lambda}(\theta) = G'_\lambda(\theta) \begin{pmatrix} O_{11} & O_{12} \\ O_{21} & O_{22} \end{pmatrix} G_\gamma, \quad B_{\lambda\beta}(\theta) = B'_{\beta\lambda}(\theta) = G'_\lambda(\theta) \begin{pmatrix} O_{11} & O_{12} \\ O_{21} & O_{22} \end{pmatrix} G_\beta \\ 3. & B_\rho(\theta) = G'_\rho(\theta) \begin{pmatrix} O_{11} & O_{12} \\ O_{21} & O_{22} \end{pmatrix} G_\rho, \quad B_{\rho\gamma}(\theta) = B'_{\gamma\rho}(\theta) = G'_\rho(\theta) \begin{pmatrix} O_{11} & O_{12} \\ O_{21} & O_{22} \end{pmatrix} G_\gamma \\ 4. & B_{\rho\beta}(\theta) = B'_{\beta\rho}(\theta) = G'_\rho(\theta) \begin{pmatrix} O_{11} & O_{12} \\ O_{21} & O_{22} \end{pmatrix} G_\beta, \quad B_\gamma(\theta) = G'_\gamma(\theta) \begin{pmatrix} O_{11} & O_{12} \\ O_{21} & O_{22} \end{pmatrix} G_\gamma \\ 5. & B_{\gamma\beta}(\theta) = B'_{\beta\gamma}(\theta) = G'_\gamma(\theta) \begin{pmatrix} O_{11} & O_{12} \\ O_{21} & O_{22} \end{pmatrix} G_\beta, \quad B_\beta(\theta) = G'_\beta(\theta) \begin{pmatrix} O_{11} & O_{12} \\ O_{21} & O_{22} \end{pmatrix} G_\beta. \end{aligned}$$

**Expressions for  $H_0^\lambda : \lambda_0 = 0$ :**

$$C_\lambda^*(\tilde{\theta}_{nT}) = [C_\lambda(\tilde{\theta}_{nT}) - B_{\lambda\phi\beta}(\tilde{\theta}_{nT}) B_{\phi\beta}^{-1}(\tilde{\theta}_{nT}) C_\phi(\tilde{\theta}_{nT})], \quad (\text{B.8})$$

372 where  $\phi = (\rho', \gamma)'$ ,  $C_\phi(\tilde{\theta}_{nT}) = (C'_\rho(\tilde{\theta}_{nT}), C'_\gamma(\tilde{\theta}_{nT}))'$ , and

$$B_{\lambda\phi\cdot\beta}(\tilde{\theta}_{nT}) = B_{\lambda\phi}(\tilde{\theta}_{nT}) - B_{\lambda\beta}(\tilde{\theta}_{nT})B_{\beta}^{-1}(\tilde{\theta}_{nT})B_{\beta\phi}(\tilde{\theta}_{nT}) \quad (\text{B.9})$$

$$= (B_{\lambda\rho}(\tilde{\theta}_{nT}), B_{\lambda\gamma}(\tilde{\theta}_{nT})) - B_{\lambda\beta}(\tilde{\theta}_{nT})B_{\beta}^{-1}(\tilde{\theta}_{nT})(B_{\beta\rho}(\tilde{\theta}_{nT}), B_{\beta\gamma}(\tilde{\theta}_{nT}))$$

$$B_{\phi\cdot\beta}(\tilde{\theta}_{nT}) = B_{\phi}(\tilde{\theta}_{nT}) - B_{\phi\beta}(\tilde{\theta}_{nT})B_{\beta}^{-1}(\tilde{\theta}_{nT})B_{\beta\phi}(\tilde{\theta}_{nT})$$

$$= \begin{bmatrix} B_{\rho}(\tilde{\theta}_{nT}) & B_{\rho\gamma}(\tilde{\theta}_{nT}) \\ B_{\gamma\rho}(\tilde{\theta}_{nT}) & B_{\rho}(\tilde{\theta}_{nT}) \end{bmatrix} - \begin{bmatrix} B_{\rho\beta}(\tilde{\theta}_{nT}) \\ B_{\gamma\beta}(\tilde{\theta}_{nT}) \end{bmatrix} B_{\beta}^{-1}(\tilde{\theta}_{nT}) \begin{bmatrix} B_{\beta\rho}(\tilde{\theta}_{nT}), B_{\beta\gamma}(\tilde{\theta}_{nT}) \end{bmatrix}.$$

$$(\text{B.10})$$

**Expressions for  $H_0^{\rho} : \rho_0 = 0$ :**

$$C_{\rho}^*(\tilde{\theta}_{nT}) = [C_{\rho}(\tilde{\theta}_{nT}) - B_{\rho\phi\cdot\beta}(\tilde{\theta}_{nT})B_{\phi\cdot\beta}^{-1}(\tilde{\theta}_{nT})C_{\phi}(\tilde{\theta}_{nT})], \quad (\text{B.11})$$

where  $\phi = (\lambda', \gamma)'$ ,  $C_{\phi}(\tilde{\theta}_{nT}) = (C'_{\lambda}(\tilde{\theta}_{nT}), C'_{\gamma}(\tilde{\theta}_{nT}))'$ , and

$$B_{\rho\phi\cdot\beta}(\tilde{\theta}_{nT}) = B_{\rho\phi}(\tilde{\theta}_{nT}) - B_{\rho\beta}(\tilde{\theta}_{nT})B_{\beta}^{-1}(\tilde{\theta}_{nT})B_{\beta\phi}(\tilde{\theta}_{nT}) \quad (\text{B.12})$$

$$= (B_{\rho\lambda}(\tilde{\theta}_{nT}), B_{\rho\gamma}(\tilde{\theta}_{nT})) - B_{\rho\beta}(\tilde{\theta}_{nT})B_{\beta}^{-1}(\tilde{\theta}_{nT})(B_{\beta\lambda}(\tilde{\theta}_{nT}), B_{\beta\gamma}(\tilde{\theta}_{nT})),$$

$$B_{\phi\cdot\beta}(\tilde{\theta}_{nT}) = B_{\phi}(\tilde{\theta}_{nT}) - B_{\phi\beta}(\tilde{\theta}_{nT})B_{\beta}^{-1}(\tilde{\theta}_{nT})B_{\beta\phi}(\tilde{\theta}_{nT})$$

$$= \begin{bmatrix} B_{\lambda}(\tilde{\theta}_{nT}) & B_{\lambda\gamma}(\tilde{\theta}_{nT}) \\ B_{\gamma\lambda}(\tilde{\theta}_{nT}) & B_{\gamma}(\tilde{\theta}_{nT}) \end{bmatrix} - \begin{bmatrix} B_{\lambda\beta}(\tilde{\theta}_{nT}) \\ B_{\gamma\beta}(\tilde{\theta}_{nT}) \end{bmatrix} B_{\beta}^{-1}(\tilde{\theta}_{nT}) \begin{bmatrix} B_{\beta\lambda}(\tilde{\theta}_{nT}), B_{\beta\gamma}(\tilde{\theta}_{nT}) \end{bmatrix}.$$

$$(\text{B.13})$$

**Expressions for  $H_0^{\gamma} : \gamma_0 = 0$ :**

$$C_{\gamma}^*(\tilde{\theta}_{nT}) = [C_{\gamma}(\tilde{\theta}_{nT}) - B_{\gamma\phi\cdot\beta}(\tilde{\theta}_{nT})B_{\phi\cdot\beta}^{-1}(\tilde{\theta}_{nT})C_{\phi}(\tilde{\theta}_{nT})], \quad (\text{B.14})$$

where  $\phi = (\lambda', \rho)'$ ,  $C_{\phi}(\tilde{\theta}_{nT}) = (C'_{\lambda}(\tilde{\theta}_{nT}), C'_{\rho}(\tilde{\theta}_{nT}))'$ , and

$$B_{\gamma\phi\cdot\beta}(\tilde{\theta}_{nT}) = B_{\gamma\phi}(\tilde{\theta}_{nT}) - B_{\gamma\beta}(\tilde{\theta}_{nT})B_{\beta}^{-1}(\tilde{\theta}_{nT})B_{\beta\phi}(\tilde{\theta}_{nT}) \quad (\text{B.15})$$

$$= (B_{\gamma\lambda}(\tilde{\theta}_{nT}), B_{\gamma\rho}(\tilde{\theta}_{nT})) - B_{\gamma\beta}(\tilde{\theta}_{nT})B_{\beta}^{-1}(\tilde{\theta}_{nT})(B_{\beta\lambda}(\tilde{\theta}_{nT}), B_{\beta\rho}(\tilde{\theta}_{nT})),$$

$$B_{\phi\cdot\beta}(\tilde{\theta}_{nT}) = B_{\phi}(\tilde{\theta}_{nT}) - B_{\phi\beta}(\tilde{\theta}_{nT})B_{\beta}^{-1}(\tilde{\theta}_{nT})B_{\beta\phi}(\tilde{\theta}_{nT})$$

$$= \begin{bmatrix} B_{\lambda}(\tilde{\theta}_{nT}) & B_{\lambda\rho}(\tilde{\theta}_{nT}) \\ B_{\rho\lambda}(\tilde{\theta}_{nT}) & B_{\rho}(\tilde{\theta}_{nT}) \end{bmatrix} - \begin{bmatrix} B_{\lambda\beta}(\tilde{\theta}_{nT}) \\ B_{\rho\beta}(\tilde{\theta}_{nT}) \end{bmatrix} B_{\beta}^{-1}(\tilde{\theta}_{nT}) \begin{bmatrix} B_{\beta\lambda}(\tilde{\theta}_{nT}), B_{\beta\rho}(\tilde{\theta}_{nT}) \end{bmatrix}.$$

$$(\text{B.16})$$

## C Proofs of Propositions

### 374 C.1 Proof of Proposition 1

Let  $g_{nT}(\theta)$  denote the  $m+q$  dimensional vector of empirical moments such that  $m+q \geq 2p+k_x+1$ .

376 Define the OGMME  $\hat{\theta}_{nT} = \operatorname{argmin}_{\theta} g'_{nT}(\theta) \tilde{\Sigma}_{nT}^{-1} g_{nT}(\theta)$ , where  $\tilde{\Sigma}_{nT}$  is a consistent estimate of  $\Sigma_{nT}$  by Lemma 1. By the implicit function theorem, the set of  $k_r$  restrictions on  $\theta_0$  can also be stated as

378  $h(\xi_0) = \theta_0$ , where  $h : \mathbb{R}^{\bar{q}} \rightarrow \mathbb{R}^{2p+k_x+1}$  is continuously differentiable,  $\xi_0$  contains the free parameters, and  $\bar{q} = 2p+k_x+1-k_r$ . Define  $\hat{\xi}_{nT} = \operatorname{argmin}_{\xi} g'_{nT}(h(\xi)) \hat{\Sigma}_{nT}^{-1} g_{nT}(h(\xi))$ . Then, we have  $\hat{\theta}_{c,nT} =$

380  $h(\widehat{\xi}_{nT})$  as the constrained OGMME of  $\theta_0$ . Let  $\tilde{\xi}_{nT}$  denote a  $\sqrt{N}$ -consistent estimate of  $\xi_0$ .

For notational simplicity, denote  $G_\theta = \frac{1}{N} \frac{\partial g_{nT}(h(\xi))}{\partial \theta'}$ ,  $\tilde{G}_\theta = \frac{1}{N} \frac{\partial g_{nT}(h(\tilde{\xi}_{nT}))}{\partial \theta'}$ ,  $G_\xi = \frac{1}{N} \frac{\partial g_{nT}(h(\xi))}{\partial \xi'}$ ,  
 382  $\tilde{G}_\xi = \frac{1}{N} \frac{\partial g_{nT}(h(\tilde{\xi}_{nT}))}{\partial \xi'}$ , and  $\tilde{g}_{nT} = g_{nT}(h(\tilde{\xi}_{nT}))$ . By Lemma 1, we have  $\text{plim}_{n,T \rightarrow \infty} \tilde{G}_\theta = G_\theta$ ,  
 $\text{plim}_{n,T \rightarrow \infty} \tilde{G}_\xi = G_\xi$ , where  $\mathcal{G}_\xi = \text{plim}_{n,T \rightarrow \infty} \frac{1}{N} \frac{\partial g_{nT}(h(\xi_0))}{\partial \xi'}$ .

In the following, we first establish the null asymptotic distribution of  $C(\alpha)$  test and then that of  $LM$ . Our proof for the null asymptotic distribution of  $C(\alpha)$  test is similar to the one provided by Lee and Yu (2012b). Let

$$\begin{aligned} \mathcal{T}_{nT}^*(\xi) &= \frac{1}{N} \frac{\partial g'_{nT}(h(\xi))}{\partial \theta} \left[ I_{m+q} - \tilde{\Sigma}_{nT}^{-1} \frac{1}{N} \frac{\partial g_{nT}(h(\xi))}{\partial \xi'} \right. \\ &\quad \times \left. \left( \frac{1}{N} \frac{\partial g'_{nT}(h(\xi))}{\partial \xi} \tilde{\Sigma}_{nT}^{-1} \frac{1}{N} \frac{\partial g_{nT}(h(\xi))}{\partial \xi'} \right)^{-1} \frac{1}{N} \frac{\partial g'_{nT}(h(\xi))}{\partial \xi} \right] \times \tilde{\Sigma}_{nT}^{-1} \frac{1}{\sqrt{N}} g_{nT}(h(\xi)) \\ &= G'_\theta \left[ I_{m+q} - \tilde{\Sigma}_{nT}^{-1} G'_\xi (G'_\xi \tilde{\Sigma}_{nT}^{-1} G'_\xi)^{-1} G'_\xi \right] \tilde{\Sigma}_{nT}^{-1} \frac{1}{\sqrt{N}} g_{nT}(h(\xi)). \end{aligned} \quad (\text{C.1})$$

**Claim 1.** — Let  $\mathcal{A}_{nT}$  be any sequence of  $(2p + k_x + 1) \times \bar{q}$  constant matrices. Define the following class of functions

$$\mathcal{T}_{nT}(\mathcal{A}_{nT}, \xi) = (\mathcal{G}'_\theta + \mathcal{A}_{nT} \mathcal{G}'_\xi) \Sigma_{nT}^{-1} \frac{1}{\sqrt{N}} g_{nT}(h(\xi)).$$

Then,

$$\frac{1}{\sqrt{N}} \text{E} \left( \frac{\partial \mathcal{T}_{nT}(\mathcal{A}_{nT}, \xi_0)}{\partial \xi'} \right) = \frac{1}{\sqrt{N}} \text{E} \left( \mathcal{T}_{nT}(\mathcal{A}_{nT}, \xi_0) g'_{nT}(\theta_0) \Sigma_{nT}^{-1} \mathcal{G}_\xi \right) + o(1).$$

*Proof.* Note that

$$\frac{\partial \mathcal{T}_{nT}(\mathcal{A}_{nT}, \xi)}{\partial \xi'} = (\mathcal{G}'_\theta + \mathcal{A}_{nT} \mathcal{G}'_\xi) \Sigma_{nT}^{-1} \frac{1}{\sqrt{N}} \frac{\partial g_{nT}(h(\xi))}{\partial \xi'}.$$

By Lemma 1, we have

$$\frac{1}{\sqrt{N}} \text{E} \left( \frac{\partial \mathcal{T}_{nT}(\mathcal{A}_{nT}, \xi_0)}{\partial \xi'} \right) = (\mathcal{G}'_\theta + \mathcal{A}_{nT} \mathcal{G}'_\xi) \Sigma_{nT}^{-1} \mathcal{G}_\xi + o(1).$$

Now, write down

$$\begin{aligned} \frac{1}{\sqrt{N}} \text{E} \left( \mathcal{T}_{nT}(\mathcal{A}_{nT}, \xi_0) g'_{nT}(\theta_0) \Sigma_{nT}^{-1} \mathcal{G}_\xi \right) &= (\mathcal{G}'_\theta + \mathcal{A}_{nT} \mathcal{G}'_\xi) \Sigma_{nT}^{-1} \frac{1}{N} \text{E} \left( g_{nT}(h(\xi_0)) g'_{nT}(\theta_0) \right) \Sigma_{nT}^{-1} \mathcal{G}_\xi \\ &= (\mathcal{G}'_\theta + \mathcal{A}_{nT} \mathcal{G}'_\xi) \Sigma_{nT}^{-1} \mathcal{G}_\xi + o(1), \end{aligned} \quad (\text{C.2})$$

384 where we use the fact that  $\frac{1}{N} \text{E} (g_{nT}(h(\phi_0)) g'_{nT}(\theta_0)) = \Sigma_{nT} + o(1)$  (see Lemma 1).  $\square$

**Claim 2.** — There exists a unique  $\mathcal{A}_{nT}^*$  in the class including  $\mathcal{A}_{nT}$  such that

$$\frac{1}{\sqrt{N}} \text{E} (\mathcal{T}_{nT}(\mathcal{A}_{nT}^*, \xi_0) g'_{nT}(\theta_0) \Sigma_{nT}^{-1} \mathcal{G}_\xi) = o(1),$$

where  $\mathcal{A}_{nT}^* = -\mathcal{G}'_\theta \Sigma_{nT}^{-1} \mathcal{G}_\xi (G'_\xi \Sigma_{nT}^{-1} G'_\xi)^{-1}$ .

386 *Proof.* The result follows from setting (C.2) to zero and solving it for  $\mathcal{A}_{nT}$ . □

**Claim 3.** — For any  $\sqrt{N}$ -consistent estimate of  $\tilde{\xi}_{nT}$  of  $\xi_0$ , we have  $\mathcal{T}_{nT}(\mathcal{A}_{nT}^*, \tilde{\xi}_{nT}) =$   
 388  $\mathcal{T}_{nT}(\mathcal{A}_{nT}^*, \xi_0) + o_p(1)$ .

*Proof.* By assumption  $\tilde{\xi}_{nT}$  is a  $\sqrt{N}$ -consistent estimator. Hence  $\sqrt{N}(\tilde{\xi}_{nT} - \xi_0) = O_p(1)$ . By the mean value theorem, we obtain

$$\mathcal{T}_{nT}(\mathcal{A}_{nT}, \tilde{\xi}_{nt}) = \mathcal{T}_{nT}(\mathcal{A}_{nT}, \xi_0) + \frac{1}{\sqrt{N}} \frac{\partial \mathcal{T}_{nT}(\mathcal{A}_{nT}, \bar{\xi}_{nT})}{\partial \xi'} \sqrt{N}(\tilde{\xi}_{nt} - \xi_0)$$

where  $\bar{\xi}_{nT}$  lies between  $\tilde{\xi}_{nt}$  and  $\xi_0$ . By  $\bar{\xi}_{nT} \xrightarrow{p} \xi_0$  and Lemma 1, we obtain

$$\begin{aligned} \frac{1}{\sqrt{N}} \frac{\partial \mathcal{T}_{nT}(\mathcal{A}_{nT}, \bar{\xi}_{nT})}{\partial \xi'} - \frac{1}{\sqrt{N}} \left( \frac{\partial \mathcal{T}_{nT}(\mathcal{A}_{nT}, \xi_0)}{\partial \xi'} \right) \\ = (\mathcal{G}'_{\theta} + \mathcal{A}_{n,T} \mathcal{G}'_{\xi}) \Sigma_{nT}^{-1} \times \underbrace{\left( \frac{1}{N} \frac{\partial g_{nT}(h(\bar{\xi}_{nT}))}{\partial \xi'} - \mathcal{G}_{\xi} \right)}_{o_p(1)} + o_p(1) = o_p(1). \end{aligned}$$

Replacing  $\mathcal{A}_{nT}$  with  $\mathcal{A}_{nT}^*$  in the mean value expansion and noting from Claim 2 that  
 390  $\frac{1}{\sqrt{N}} \mathbf{E} \left( \frac{\partial \mathcal{T}_{nT}(\mathcal{A}_{nT}^*, \xi_0)}{\partial \xi'} \right) = o(1)$ , we obtain the desired result. □

**Claim 4.** — At any  $\sqrt{N}$ -consistent estimate  $\tilde{\xi}_{nT}$ ,  $\mathcal{T}_{nT}^*(\tilde{\xi}_{nT}) - \mathcal{T}_{nT}(\mathcal{A}_{nT}^*, \tilde{\xi}_{nT}) = o_p(1)$  and  
 392  $\mathcal{T}_{nT}^*(\tilde{\xi}_{nT}) = \mathcal{T}_{nT}(\mathcal{A}_{nT}^*, \xi_0) + o_p(1)$ .

*Proof.* Let  $\mathcal{B}_{nT}(\tilde{\xi}_{nT}) = \tilde{G}'_{\theta} [\mathbf{I}_{m+q} - \tilde{\Sigma}_{nT}^{-1} \tilde{G}_{\xi} (\tilde{G}'_{\xi} \tilde{\Sigma}_{nT}^{-1} \tilde{G}_{\xi})^{-1} \tilde{G}'_{\xi}] \tilde{\Sigma}_{nT}^{-1}$  and  $\mathcal{B}_{nT}^* = \mathcal{G}'_{\theta} [\mathbf{I}_{m+q} -$   
 $\Sigma_{nT}^{-1} \mathcal{G}_{\xi} (\mathcal{G}'_{\xi} \Sigma_{nT}^{-1} \mathcal{G}_{\xi})^{-1} \mathcal{G}'_{\xi}] \Sigma_{nT}^{-1}$ . Then, it follows that  $\mathcal{T}_{nT}^*(\tilde{\xi}_{nT}) - \mathcal{T}_{nT}(\mathcal{A}_{nT}^*, \tilde{\xi}_{nT}) = [\mathcal{B}_{nT}(\tilde{\xi}_{nT}) -$   
 $\mathcal{B}_{nT}^*] \frac{1}{\sqrt{N}} g_{nT}(h(\tilde{\xi}_{nT}))$ . By Lemma 1,  $[\mathcal{B}_{nT}(\tilde{\xi}_{nT}) - \mathcal{B}_{nT}^*] = o_p(1)$ . By the mean value theorem,

$$\begin{aligned} \frac{1}{\sqrt{N}} g_{nT}(h(\tilde{\xi}_{nT})) &= \frac{1}{\sqrt{N}} g_{nT}(h(\xi_0)) + \frac{\partial g_{nT}(h(\bar{\xi}_{nT}))}{\partial \xi'} \frac{1}{\sqrt{N}} (\tilde{\xi}_{nT} - \xi_0) \\ &= \frac{1}{\sqrt{N}} g_{nT}(h(\xi_0)) + \frac{1}{N} \frac{\partial g_{nT}(h(\bar{\xi}_{nT}))}{\partial \xi'} \sqrt{N} (\tilde{\xi}_{nT} - \xi_0). \end{aligned}$$

Since (i)  $\sqrt{N}(\tilde{\xi}_{nT} - \xi_0) = O_p(1)$ , (ii)  $\frac{1}{N} \frac{\partial g_{nT}(h(\bar{\xi}_{nT}))}{\partial \xi'} = \mathcal{G}_{\xi} + o_p(1)$  by  $\bar{\xi}_{nT} \xrightarrow{p} \xi_0$  and Lemma 1,  
 394 and (iii)  $\frac{1}{\sqrt{N}} g_{nT}(h(\xi_0)) = O_p(1)$ , and  $\frac{1}{\sqrt{N}} g_{nT}(h(\tilde{\xi}_{nT})) = O_p(1)$  by Lemma 1. Hence,  $\mathcal{T}_{nT}^*(\tilde{\xi}_{nT}) -$   
 $\mathcal{T}_{nT}(\mathcal{A}_{nT}^*, \tilde{\xi}_{nT}) = o_p(1)$ . Then, by Claim 3, we have  $\mathcal{T}_{nT}^*(\tilde{\xi}_{nT}) = \mathcal{T}_{nT}(\mathcal{A}_{nT}^*, \xi_0) + o_p(1)$ . □

**Claim 5.** — Under  $H_0$ , the random variable  $\mathcal{T}_{nT}(\mathcal{A}_{nT}^*, \xi_0)$  has zero mean and variance  $\Omega =$   
 $\text{plim}_{n,T \rightarrow \infty} \Omega_{nT}$ , where  $\Omega_{nT} = \mathcal{G}'_{\theta} [\Sigma_{nT}^{-1} - \Sigma_{nT}^{-1} \mathcal{G}_{\xi} (\mathcal{G}'_{\xi} \Sigma_{nT}^{-1} \mathcal{G}_{\xi})^{-1} \mathcal{G}'_{\xi} \Sigma_{nT}^{-1}] \mathcal{G}_{\theta}$  with rank  $k_r$ . Furthermore,  
 398  $\mathcal{T}_{nT}(\mathcal{A}_{nT}^*, \xi_0) \xrightarrow{d} N(0, \Omega)$ .

*Proof.* Note that  $\mathcal{G}_{\theta}$  has full rank  $2p + k_x + 1$ . Hence,  $\mathcal{G}'_{\theta} \Sigma_{nT}^{-1} \mathcal{G}_{\theta}$  is a positive definite matrix which can be cholesky decomposed as  $L_{nT} L'_{nT}$ , where  $L_{nT}$  is invertible. Further, since  $\frac{1}{N} \frac{\partial g_{nT}(h(\xi_0))}{\partial \xi'} =$

$\frac{1}{N} \frac{\partial g_{nT}(\theta_0)}{\partial \theta'} \frac{\partial h(\xi_0)}{\partial \xi'}$ , we have  $\mathcal{G}_\xi = \mathcal{G}_\theta H_{nT}$ , where  $H_{nT} = \frac{\partial h(\xi_0)}{\partial \xi'}$ . Then,  $\mathcal{T}_{nT}(\mathcal{A}_{nT}^*, \xi_0)$  can be written as

$$\begin{aligned} \mathcal{T}_{nT}(\mathcal{A}_{nT}^*, \xi_0) &= [I_{2p+k_x+1} - \mathcal{G}'_\theta \Sigma_{nT}^{-1} \mathcal{G}_\theta H_{nT} (H'_{nT} \mathcal{G}'_\theta \Sigma_{nT}^{-1} \mathcal{G}_\theta H_{nT})^{-1} H'_{nT}] \mathcal{G}'_\theta \Sigma_{nT}^{-1} \frac{1}{\sqrt{N}} g_{nT}(\theta_0) \\ &= L_{nT} \mathbf{M}_{L'H} L_{nT}^{-1} \mathcal{G}'_\theta \Sigma_{nT}^{-1} \frac{1}{\sqrt{N}} g_{nT}(\theta_0) \end{aligned} \quad (\text{C.3})$$

where  $\mathbf{M}_{L'H} = I_{2p+k_x+1} - \mathbf{P}_{L'H}$  and  $\mathbf{P}_{L'H} = L'_{nT} H_{nT} (H'_{nT} L_{nT} L'_{nT} H_{nT})^{-1} H'_{nT} L_{nT}$ . Note that  $\mathbf{M}_{L'H}$  is idempotent with its rank equal to  $2p + k_x + 1 - \bar{q} = k_r$ . Then,

$$\begin{aligned} \text{Var}[\mathcal{T}_{nT}(\mathcal{A}_{nT}^*, \xi_0)] &= \text{plim}_{n,T \rightarrow \infty} L_{nT} \mathbf{M}_{L'H} L_{nT}^{-1} L_{nT} L'_{nT} L_{nT}^{-1} \mathbf{M}_{L'H} L'_{nT} \\ &= \text{plim}_{n,T \rightarrow \infty} L_{nT} \mathbf{M}_{L'H} L'_{nT} = \text{plim}_{n,T \rightarrow \infty} \Omega_{nT} \end{aligned}$$

where  $\Omega_{nT}$  is singular with rank  $k_r$ . By Lemma 1,  $\frac{1}{\sqrt{N}} g_{nT}(\theta_0) \xrightarrow{d} N(0, \text{plim}_{n \rightarrow \infty} \Sigma_{nT})$ . Hence,

$$400 \quad \mathcal{T}_{nT}(\mathcal{A}_{nT}^*, \xi_0) \xrightarrow{d} N(0, \Omega). \quad \square$$

**Claim 6.** — Denote  $C^*(\alpha) = \mathcal{T}_{nT}(\mathcal{A}_{nT}^*, \xi_0)' \Omega_{nT}^- \mathcal{T}_{nT}(\mathcal{A}_{nT}^*, \xi_0)$ , where  $\Omega_{nT}^-$  is the generalized inverse of  $\Omega_{nT}$ .

*Proof.* It follows from Claim 5 that  $C^*(\alpha) \xrightarrow{A} \chi_{k_r}^2$ . Note that  $\Omega_{nT} = L_{nT} \mathbf{M}_{L'H} L'_{nT}$  and the generalized inverse of  $\mathbf{M}_{L'H}$  is itself, then  $\Omega_{nT}^- = L_{nT}^{-1} \mathbf{M}_{L'H} L_{nT}$ . It follows from (C.3)

$$\begin{aligned} C^*(\alpha) &= N \frac{1}{\sqrt{N}} g'_{nT}(\theta_0) \Sigma_{nT}^{-1} \mathcal{G}'_\theta L_{nT}^{-1} \mathbf{M}_{L'H} L'_{nT} \Omega_{nT}^- L_{nT} \mathbf{M}_{L'H} L_{nT}^{-1} \mathcal{G}'_\theta \Sigma_{nT}^{-1} \frac{1}{\sqrt{N}} g_{nT}(\theta_0) \\ &= \frac{1}{N} g'_{nT}(\theta_0) \Sigma_{nT}^{-1} \mathcal{G}'_\theta L_{nT}^{-1} \mathbf{M}_{L'H} L'_{nT} \Omega_{nT}^- L_{nT} \mathbf{M}_{L'H} L_{nT}^{-1} \mathcal{G}'_\theta \Sigma_{nT}^{-1} g_{nT}(\theta_0) \\ &= \frac{1}{N} g'_{nT}(\theta_0) \Sigma_{nT}^{-1} \mathcal{G}'_\theta L_{nT}^{-1} \mathbf{M}_{L'H} L'_{nT} \mathcal{G}'_\theta \Sigma_{nT}^{-1} g_{nT}(\theta_0). \end{aligned} \quad (\text{C.4})$$

Note that

$$\begin{aligned} L_{nT}^{-1} \mathbf{M}_{L'H} L'_{nT} &= (L_{nT} L'_{nT})^{-1} - H_{nT} (H'_{nT} \mathcal{G}'_\theta \Sigma_{nT}^{-1} \mathcal{G}_\theta H_{nT})^{-1} H'_{nT} \\ &= (\mathcal{G}'_\theta \Sigma_{nT}^{-1} \mathcal{G}_\theta)^{-1} - H_{nT} (\mathcal{G}'_\xi \Sigma_{nT}^{-1} \mathcal{G}_\xi)^{-1} H'_{nT} \end{aligned} \quad (\text{C.5})$$

Then, it follows from (C.4) and (C.5) that

$$\begin{aligned} C^*(\alpha) &= \frac{1}{N} g'_{nT}(\theta_0) \Sigma_{nT}^{-1} \mathcal{G}'_\theta (\mathcal{G}'_\theta \Sigma_{nT}^{-1} \mathcal{G}_\theta)^{-1} \mathcal{G}'_\theta \Sigma_{nT}^{-1} g_{nT}(\theta_0) \\ &\quad - \frac{1}{N} g'_{nT}(\theta_0) \Sigma_{nT}^{-1} \mathcal{G}'_\xi (\mathcal{G}'_\xi \Sigma_{nT}^{-1} \mathcal{G}_\xi)^{-1} \mathcal{G}'_\xi \Sigma_{nT}^{-1} g_{nT}(\theta_0). \end{aligned} \quad (\text{C.6})$$

404 **Claim 7.** — The test statistic can be written as  $C(\alpha) = \mathcal{T}_{nT}(\tilde{\xi}_{nT})^* \tilde{\Omega}_{nT}^- \mathcal{T}_{nT}^*(\tilde{\xi}_{nT})$ , where  $\tilde{\Omega}_{nT} = \tilde{G}'_\theta [\tilde{\Sigma}_{nT}^{-1} - \tilde{\Sigma}_{nT}^{-1} \tilde{G}'_\xi (\tilde{G}'_\xi \tilde{\Sigma}_{nT}^{-1} \tilde{G}_\xi)^{-1} \tilde{G}'_\xi \tilde{\Sigma}_{nT}^{-1}] \tilde{G}_\theta$ . Under  $H_0$ , it follows that  $C(\alpha) \xrightarrow{d} \chi_{k_r}^2$ .

406 *Proof.* By Lemma 2,  $\tilde{\Omega}_{nT}^- - \Omega_{nT}^- = o_p(1)$ . Furthermore, by Claim 4  $\mathcal{T}_{nT}^*(\tilde{\xi}_{nT}) = \mathcal{T}_{nT}(\mathcal{A}_{nT}^*, \xi_0) + o_p(1)$ . Hence,  $C(\alpha) - C^*(\alpha) = o_p(1)$  by continuous mapping theorem. Then, the asymptotic equivalence  
408 (White (2001, Lemma 4.7, p.67)) and Claim 4 yield the desired result.  $\square$



Now we will establish the null asymptotic distribution of  $LM$  test. Recall that the test statistic is

$$LM = N C'(\hat{\theta}_{nT,r}) B^{-1}(\hat{\theta}_{nT,r}) C(\hat{\theta}_{nT,r}). \quad (\text{C.7})$$

Let  $\widetilde{LM} = \sqrt{N} C'(\hat{\theta}_{nT,r}) \mathcal{H}^{-1} \sqrt{N} C(\hat{\theta}_{nT,r})$ . Under  $H_0 : r(\theta_0) = 0$ , we have  $LM = \widetilde{LM} + o_p(1)$  by Lemma 1 and  $\hat{\theta}_{nT,r} = \theta_0 + o_p(1)$ . Now consider the limiting behavior of  $\sqrt{N} C(\hat{\theta}_{nT,r})$ . By the mean value theorem, we have

$$\begin{aligned} \sqrt{N} C(\hat{\theta}_{nT,r}) &= \sqrt{N} C(\theta_0) - \mathcal{G}'(\bar{\theta}) \widehat{\Sigma}_{nT} \mathcal{G}(\bar{\theta}) \times \sqrt{N} (\hat{\theta}_{nT,r} - \theta_0) \\ &= \sqrt{N} C(\theta_0) - \mathcal{H} \times \sqrt{N} (\hat{\theta}_{nT,r} - \theta_0) + o_p(1). \end{aligned} \quad (\text{C.8})$$

To evaluate (C.8), we need to consider the limiting behavior of  $\sqrt{N}(\hat{\theta}_{nT,r} - \theta_0)$ . The result derived for the limiting behavior of constrained GMME in Hall (2004, Lemma 5.4, p.167) can be considered for our case. It can be shown that

$$\sqrt{N} (\hat{\theta}_{nT,r} - \theta_0) = [\mathcal{H}^{-1} - \mathcal{H}^{-1} R' (R \mathcal{H}^{-1} R')^{-1} R \mathcal{H}^{-1}] \sqrt{N} C(\theta_0) + o_p(1), \quad (\text{C.9})$$

where  $R = R(\theta_0) = \frac{\partial r(\theta_0)}{\partial \theta'}$ . Substituting (C.9) into (C.8) yields

$$\sqrt{N} C(\hat{\theta}_{nT,r}) = R' (R \mathcal{H}^{-1} R')^{-1} R \mathcal{H}^{-1} \sqrt{N} C(\theta_0) + o_p(1). \quad (\text{C.10})$$

Substituting (C.10) into  $\widetilde{LM}$  yields  $\widetilde{LM} = \sqrt{N} C'(\theta_0) \mathcal{H}^{-1} R' (R \mathcal{H}^{-1} R')^{-1} R \mathcal{H}^{-1} \sqrt{N} C(\theta_0) + o_p(1)$ .

410 By Lemma 1, we have  $R \mathcal{H}^{-1} \sqrt{N} C(\theta_0) \xrightarrow{d} N(0, R \mathcal{H}^{-1} R')$ , which implies that  $\widetilde{LM} \xrightarrow{d} \chi_{k_r}^2$ . Then, the desired results follows from the asymptotic equivalence of  $\widetilde{LM}$  and  $LM$ .

## 412 C.2 Proof of Proposition 2

The first three results follows directly from  $LM_\psi(\tilde{\theta}_{nT}) \xrightarrow{d} \chi_{k_\psi}^2(\vartheta_1)$  under  $H_A^\psi$  and  $H_A^\phi$ , where  $\vartheta_1 = \delta'_\psi \mathcal{H}_{\psi \cdot \beta} \delta_\psi + \delta'_\psi \mathcal{H}_{\psi \phi \cdot \beta} \delta_\phi + \delta'_\phi \mathcal{H}'_{\psi \phi \cdot \beta} \delta_\psi + \delta'_\phi \mathcal{H}'_{\psi \phi \cdot \beta} \mathcal{H}_{\psi \cdot \beta}^{-1} \mathcal{H}_{\psi \phi \cdot \beta} \delta_\phi$  is the non-centrality parameter. Here, we will prove the last two results. For this purpose, we consider the distribution of  $\mathbf{C}_{\psi\phi}(\tilde{\theta}_{nT}) = (C'_\psi(\tilde{\theta}_{nT}), C'_\phi(\tilde{\theta}_{nT}))'$ . The first order Taylor expansions of  $\mathbf{C}_{\psi\phi}(\tilde{\theta}_{nT})$  and  $C_\beta(\tilde{\theta}_{nT})$  around  $\theta^* = (\beta'_0, \psi'_0 + \delta'_\psi/\sqrt{N}, \phi'_0 + \delta'_\phi/\sqrt{N})'$  are given by

$$\begin{aligned} \sqrt{N} \mathbf{C}_{\psi\phi}(\tilde{\theta}_{nT}) &= \sqrt{N} \mathbf{C}_{\psi\phi}(\theta^*) - \mathbf{G}'_{\psi\phi}(\theta^*) \widehat{\Sigma}_{nT}^{-1} \mathbf{G}_{\psi\phi}(\bar{\theta}) (\delta'_\psi, \delta'_\phi)' \\ &\quad + \sqrt{N} \mathbf{G}'_{\psi\phi}(\theta^*) \widehat{\Sigma}_{nT}^{-1} G_\beta(\bar{\theta}) (\tilde{\beta}_{nT} - \beta_0) + o_p(1), \end{aligned} \quad (\text{C.11})$$

$$\begin{aligned} \sqrt{N} C_\beta(\tilde{\theta}_{nT}) &= \sqrt{N} C_\beta(\theta^*) - G'_\beta(\theta^*) \widehat{\Sigma}_{nT}^{-1} \mathbf{G}_{\psi\phi}(\bar{\theta}) (\delta'_\psi, \delta'_\phi)' \\ &\quad + \sqrt{N} G'_\beta(\theta^*) \widehat{\Sigma}_{nT}^{-1} G_\beta(\bar{\theta}) (\tilde{\beta}_{nT} - \beta_0) + o_p(1), \end{aligned} \quad (\text{C.12})$$

where  $\mathbf{G}_{\psi\phi}(\theta) = (G_\psi(\theta), G_\phi(\theta))$ . Then, using (C.11) and (C.12), we obtain

$$\begin{aligned} \sqrt{N} \mathbf{C}_{\psi\phi}(\tilde{\theta}_{nT}) &= (\mathcal{G}'_{\psi\phi} \Sigma_{nT}^{-1}, \mathbf{H}_{\psi\phi,\beta} \mathcal{H}_\beta^{-1} \mathcal{G}'_{\beta} \Sigma_{nT}^{-1}) \frac{1}{\sqrt{N}} g_{nT}(\theta^*) \\ &\quad - \begin{pmatrix} \mathcal{H}_{\psi\cdot\beta} & \mathcal{H}_{\psi\phi\cdot\beta} \\ \mathcal{H}_{\phi\psi\cdot\beta} & \mathcal{H}_{\phi\cdot\beta} \end{pmatrix} \begin{pmatrix} \delta_\psi \\ \delta_\phi \end{pmatrix} + o_p(1), \end{aligned} \quad (\text{C.13})$$

where  $\mathcal{G}_{\psi\phi} = (\mathcal{G}_\psi, \mathcal{G}_\phi)$ ,  $\mathbf{H}_{\psi\phi,\beta} = (\mathcal{H}'_{\psi\beta}, \mathcal{H}'_{\phi\beta})'$ , and  $\mathcal{H}_{\phi\psi\cdot\beta} = \mathcal{H}_{\phi\psi} - \mathcal{H}_{\phi\beta} \mathcal{H}_\beta^{-1} \mathcal{H}_{\beta\psi}$ . Using Lemma 1, we can determine the distribution of  $\sqrt{N} \mathbf{C}_{\psi\phi}(\tilde{\theta}_{nT})$  under  $H_0^\psi$  and  $H_A^\phi$  from (C.13). Then,

$$\sqrt{N} \mathbf{C}_{\psi\phi}(\tilde{\theta}_{nT}) \xrightarrow{d} N\left(-\begin{pmatrix} \mathcal{H}_{\psi\phi\cdot\beta} \delta_\phi \\ \mathcal{H}_{\phi\cdot\beta} \delta_\phi \end{pmatrix}, \begin{pmatrix} \mathcal{H}_{\psi\cdot\beta} & \mathcal{H}_{\psi\phi\cdot\beta} \\ \mathcal{H}_{\phi\psi\cdot\beta} & \mathcal{H}_{\phi\cdot\beta} \end{pmatrix}\right). \quad (\text{C.14})$$

The result in (C.14) can be used to determine the distribution of  $\sqrt{N} C_\psi^*(\tilde{\theta}_{nT}) =$   
 $414 \quad (I, -\mathcal{H}_{\psi\phi\cdot\beta} \mathcal{H}_{\phi\cdot\beta}^{-1}) \sqrt{N} \mathbf{C}_{\psi\phi}(\tilde{\theta}_{nT}) + o_p(1)$ . Hence  $\sqrt{N} [C_\psi(\tilde{\theta}_{nT}) - \mathcal{H}_{\psi\phi\cdot\beta} \mathcal{H}_{\phi\cdot\beta}^{-1} C_\phi(\tilde{\theta}_{nT})] \xrightarrow{d} N(0, \mathcal{H}_{\psi\cdot\beta} - \mathcal{H}_{\psi\phi\cdot\beta} \mathcal{H}_{\phi\cdot\beta}^{-1} \mathcal{H}_{\phi\psi\cdot\beta})$ . Then, this last result and Lemma 1 yield the desired result.

Using (C.13) and Lemma 1, we can also determine the distribution of  $\sqrt{N} C_\psi^*(\tilde{\theta}_{nT})$  under  $H_A^\psi$  and  $H_0^\phi$  for the asymptotic power analysis. We have

$$\sqrt{N} C_\psi^*(\tilde{\theta}_{nT}) \xrightarrow{d} N\left(-(\mathcal{H}_{\psi\cdot\beta} - \mathcal{H}_{\psi\phi\cdot\beta} \mathcal{H}_{\phi\cdot\beta}^{-1} \mathcal{H}'_{\psi\phi\cdot\beta}) \delta_\psi, \mathcal{H}_{\psi\cdot\beta} - \mathcal{H}_{\psi\phi\cdot\beta} \mathcal{H}_{\phi\cdot\beta}^{-1} \mathcal{H}'_{\psi\phi\cdot\beta}\right). \quad (\text{C.15})$$

Therefore,  $LM_\psi^*(\tilde{\theta}_{nT}) \xrightarrow{d} \chi_{k_\psi}^2(\vartheta_4)$ , where  $\vartheta_4 = \delta'_\psi (\mathcal{H}_{\psi\cdot\beta} - \mathcal{H}_{\psi\phi\cdot\beta} \mathcal{H}_{\phi\cdot\beta}^{-1} \mathcal{H}'_{\psi\phi\cdot\beta}) \delta_\psi$ . It follows that  
 $416 \quad \vartheta_2 - \vartheta_4 \geq 0$ . This result indicates that  $LM_\psi^*(\tilde{\theta}_{nT})$  has less asymptotic power than  $LM_\psi(\tilde{\theta}_{nT})$  when  
 $418 \quad$  there is no local misspecification.

## References

- Amemiya, Takeshi (1985). *Advanced Econometrics*. Cambridge: Harvard University Press.
- Anderson, T.W. (2003). *An Introduction to Multivariate Statistical Analysis*. 3rd ed. Wiley Series in Probability and Statistics. Wiley-Interscience.
- Anselin, Luc et al. (1996). “Simple diagnostic tests for spatial dependence”. In: *Regional Science and Urban Economics* 26.1, pp. 77–104.
- Anselin, Luc, Julie Le Gallo, and Hubert Jayet (2008). “Spatial Panel Econometrics”. In: *The Econometrics of Panel Data*. Ed. by Laszlo Matyas et al. Vol. 46. Advanced Studies in Theoretical and Applied Econometrics. Springer Berlin Heidelberg.
- Baltagi, Badi. (2008). *Econometric Analysis of Panel Data*. John Wiley & Sons.
- Baltagi, Badi H. and Zhenlin Yang (2013). “Standardized LM tests for spatial error dependence in linear or panel regressions”. In: *The Econometrics Journal* 16.1.
- Baltagi, Badi H., Seuck Heun Song, and Won Koh (2003). “Testing panel data regression models with spatial error correlation”. In: *Journal of Econometrics* 117.1, pp. 123–150.
- Baltagi, Badi H. et al. (2007). “Testing for serial correlation, spatial autocorrelation and random effects using panel data”. In: *Journal of Econometrics* 140.1, pp. 5–51.
- Bera, Anil K. and Yannis Biliias (2001). “Rao’s score, Neyman’s  $C(\alpha)$  and Silvey’s LM tests: an essay on historical developments and some new results”. In: *Journal of Statistical Planning and Inference* 97.1, pp. 9–44.
- Bera, Anil K. and Mann J. Yoon (1993). “Specification Testing with Locally Misspecified Alternatives”. In: *Econometric Theory* 9.4.
- Bera, Anil K., Gabriel Montes-Rojas, and Walter Sosa-Escudero (2010). “General Specification Testing with Locally Misspecified Models”. In: *Econometric Theory* 26 (06).
- Breusch, T. S. and A. R. Pagan (1980). “The Lagrange Multiplier Test and its Applications to Model Specification in Econometrics”. In: *The Review of Economic Studies* 47.1.
- Davidson, Russell and James G. MacKinnon (1987). “Implicit Alternatives and the Local Power of Test Statistics”. English. In: *Econometrica* 55.6.
- (1998). “Graphical Methods for Investigating the Size and Power of Hypothesis Tests”. In: *The Manchester School* 66.1, pp. 1–26.
- Debary, Nicolas and Cem Ertur (2010). “Testing for spatial autocorrelation in a fixed effects panel data model”. In: *Regional Science and Urban Economics* 40.6.
- Elhorst, J. Paul (2005). “Unconditional Maximum Likelihood Estimation of Linear and Log-Linear Dynamic Models for Spatial Panels”. In: *Geographical Analysis* 37.1.
- Hall, Alastair R. (2004). *Generalized Method of Moments*. Advanced Texts in Econometrics. Oxford University Press.
- Hsiao, Cgeng. (2014). *Analysis of Panel Data*. Analysis of Panel Data. Cambridge University Press.
- Kapoor, Mudit, Harry H. Kelejian, and Ingmar R. Prucha (2007). “Panel data models with spatially correlated error components”. In: *Journal of Econometrics* 140.1, pp. 97–130.
- Kelejian, Harry H. and Ingmar R. Prucha (2010). “Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances”. In: *Journal of Econometrics* 157, pp. 53–67.
- Lee, Lung-fei and Jihai Yu (2010a). “A Spatial Dynamic Panel Data Model with both Time and Individual Fixed Effects”. In: *Econometric Theory* 26 (02), pp. 564–597.
- (2010b). “Some recent developments in spatial panel data models”. In: *Regional Science and Urban Economics* 40.5.
- (2011). “A Unified Transformation Approach for the Estimation of Spatial Dynamic Panel Data Models: Stability, Spatial Cointegration and Explosive Roots”. In: *Handbook of Empirical Eco-*

- 466 *nomics and Finance (Statistics: A Series of Textbooks and Monographs)*. Ed. by Aman Ullah  
and David E.A. Giles. Statistics: Textbooks and Monographs. Chapman & Hall/CRC Taylor &  
468 Francis Group.
- Lee, Lung-fei and Jihai Yu (2012a). “QML Estimation of Spatial Dynamic Panel Data Models with  
470 Time Varying Spatial Weights Matrices”. In: *Spatial Economic Analysis* 7.1, pp. 31–74.
- (2012b). “The  $C(\alpha)$ -type gradient test for spatial dependence in spatial autoregressive models”.  
472 In: *Letters in Spatial and Resource Sciences* 5.3.
- (2014). “Efficient GMM estimation of spatial dynamic panel data models with fixed effects”. In:  
474 *Journal of Econometrics* 180.2.
- (2015). “Estimation of fixed effects panel regression models with separable and non-separable  
476 space-time filters”. In: *Journal of Econometrics* 184.1, pp. 174 –192.
- Newey, Whitney K. and Kenneth D. West (1987). “Hypothesis Testing with Efficient Method of  
478 Moments Estimation”. In: *International Economic Review* 28.3, pp. 777–787.
- Parent, Olivier and James P. LeSage (2011). “A space–time filter for panel data models containing  
480 random effects”. In: *Computational Statistics & Data Analysis* 55.1.
- Saikkonen, Pentti (1989). “Asymptotic relative efficiency of the classical test statistics under mis-  
482 specification”. In: *Journal of Econometrics* 42.3, pp. 351 –369.
- Su, Liangjun and Zhenlin Yang (2015). “QML estimation of dynamic panel data models with spatial  
484 errors”. In: *Journal of Econometrics* 185.1, pp. 230 –258.
- White, Halbert (2001). *Asymptotic Theory for Econometricians*. Economic theory, econometrics,  
486 and mathematical economics. Academic Press.
- Yang, Zhenlin (2015). *Unified M-Estimation of Fixed-Effects Spatial Dynamic Models with Short  
488 Panels*. Research Collection School of Economics 1783. Singapore Management University.
- Yu, Jihai and Lung-fei Lee (2010). “Estimation of Unit root Spatial Dynamic Panel Data Models”.  
490 In: *Econometric Theory* 26 (05).
- Yu, Jihai, Robert de Jong, and Lung fei Lee (2008). “Quasi-maximum likelihood estimators for  
492 spatial dynamic panel data with fixed effects when both  $n$  and  $T$  are large”. In: *Journal of  
Econometrics* 146.1, pp. 118 –134.
- 494 — (2012). “Estimation for spatial dynamic panel data with fixed effects: The case of spatial coin-  
tegration”. In: *Journal of Econometrics* 167.1, pp. 16 –37.