GMM Gradient Tests for Spatial Dynamic Panel Data Models

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

2017

GMM Gradient Tests for Spatial Dynamic Panel Data Models*

Süleyman Taşpınar† Osman Doğan‡ Anil K. Bera§

April 12, 2017

Abstract

In this study, we formulate adjusted gradient tests when the alternative model used to construct tests deviates from the true data generating process for a spatial dynamic panel data (SDPD) model. Following Bera et al. (2010), we introduce these adjusted gradient tests along with their standard counterparts within a generalized method of moments 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 a high order SDPD model. These adjusted tests have two advantages: (i) their null asymptotic distribution is a central chi-squared distribution irrespective of the mis-specified alternative model, and (ii) their test statistics are computationally simple and require only the ordinary least-squares estimates from a non-spatial two-way panel data model. We investigate the finite sample size and power properties of these tests through a Monte Carlo study. Our results indicates that the adjusted gradient tests have good finite sample properties. Finally, using an application from the empirical growth literature we complement our findings.

JEL-Classification: C13, C21, C31.
Keywords: Spatial Dynamic Panel Data Model, SDPD, GMM, Robust LM Tests, GMM Gradient Tests, Inference.

*We are grateful to the Editor and two anonymous referees for many pertinent comments and constructive suggestions. However, we retain the responsibility of any remaining shortcomings of the paper. This research was supported, in part, under National Science Foundation Grants CNS-0958379, CNS-0855217, ACI-1126113 and the City University of New York High Performance Computing Center at the College of Staten Island.

†Economics Program, Queens College, The City University of New York, United States, email: staspinar@qc.cuny.edu.

‡Economics Program, University of Illinois, Illinois, United States, email: odogan10@gmail.com.

§Economics Program, University of Illinois, Illinois, United States, email: abera@illinois.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 generalized method of moments (GMM) gradient tests, adjusted GMM gradient tests and a Neyman's $C(\alpha)$ test, which may be viewed as a conditional gradient test, see Bera and Bilias (2001), to test hypotheses about the parameters of the time lag term, spatial time lag terms and contemporaneous spatial lag terms.

In the literature, model specifications and estimation strategies, including the maximum likelihood (ML), GMM and Bayesian methods, receive considerably more attention than specification testing and other forms of hypothesis tests for the SDPD model. For some recent surveys on spatial panel data models, see Anselin et al. (2008), Elhorst (2010a,b, 2014) and Lee and Yu (2010b, 2015b). Yu et al. (2008, 2012), Yu and Lee (2010) and Lee and Yu (2010a, 2011, 2012b, 2016) 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 (quasi) ML estimator (QMLE) suggested in these studies may have asymptotic bias contingent on the asymptotic setup assumed for the growth rate of the number of individuals and the number of time periods. For example, the QMLE in Yu et al. (2008) has asymptotic bias when the number of individuals are asymptotically proportional to number of time periods, i.e., when $n/T \to k < \infty$. The limiting distributions of bias corrected versions suggested in Yu et al. (2008) are only properly centered when T grows relatively fast compare to $n^{1/3}$. For the asymptotic scenario where the cross-sectional dimension is large and the time dimension is fixed, Elhorst (2005) and Su and Yang (2015) consider the ML approach for the dynamic panel data models that have spatial autoregressive processes in the disturbance term. Lee and Yu (2015a) suggest a QMLE to a static two-way panel data model that has disturbances with dynamic and spatial correlations which might be spatially stable or unstable. The resulting QMLE is consistent and has a properly centered asymptotic normal distribution regardless of whether T is large or not, and whether the process defined for the disturbance term is stable or not.

Other estimation methods have also been used to estimate spatial panel data models. For example, Parent and LeSage (2010, 2011, 2012) and Han et al. (2016) consider the Bayesian MCMC method for panel data models that accommodate spatial dependence across space and time. 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. Korniotis (2010) suggests a hybrid of the least-squares dummy variable estimator and the instrumental variable (IV) estimator for a dynamic panel data model that has a spatial time lag term. Yang (2015b) proposes an M-estimator for an extended SDPD model when n is large and T can be fixed or large. The proposed M-estimator is robust in the sense that it is free from the specification of the initial conditions and it allows for the disturbance term to be non-normal.

To date, focus has been on specification testing for the cross-sectional and static spatial panel data models (Anselin 1988; Anselin 2001; Anselin et al. 1996; Baltagi and Li 2001; Baltagi and Yang 2013; Baltagi et al. 2003, 2007; Cliff and Ord 1972; Debarsy and Ertur 2010; Kelejian and Prucha 2001; Kelejian and Robinson 2004; Moran 1950; Robinson 2008; Yang 2010). The Moran I test is one of the most widely used test for spatial dependence. It does not require a specific specification for the alternative model and is simply formulated from the normalized quadratic form of the variables to be tested for spatial dependence. Cliff and Ord (1972) generalizes the Moran I test for testing spatial dependence in the disturbance terms of a classical linear regression model. Kelejian and

Prucha (2001) introduce a central limit theorem (CLT) that can be used to establish the asymptotic distribution of the Moran I test under certain regularity conditions. When the alternative model is specified as a spatial model, the preferred approach for testing is often the Rao's score tests (or the Lagrange multiplier (LM) tests), because their formulations require the estimation of the null models only, circumventing the estimation issues associated with the alternative models. Burridge (1980) shows that the Rao's score test formulated from the alternative model that has a spatial autoregressive or a spatial moving process is equivalent to the Moran I statistic. Anselin (1988) and Anselin et al. (1996) derive LM and adjusted LM statistics for cross-sectional spatial autoregressive models. Baltagi et al. (2003), Baltagi et al. (2007), Debarsy and Ertur (2010) and Baltagi and Yang (2013) consider LM tests for spatial panel data models. The refinement methods including bootstrap and Edgeworth expansions are also considered to improve the finite sample properties of test statistics. Among others, see for example, Fingleton and Le Gallo (2008), Burridge and Fingleton (2010), Yang (2015a), Jin and Lee (2015), Robinson and Rossi (2014, 2015a,b) and Taspinar et al. (2016).

In this study, we propose GMM-based tests for an SDPD model that has higher orders of the 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 non-linear restrictions on the parameter vector. We also consider a $C(\alpha)$ test within the GMM framework for the same model. While the computation of the GMM-gradient test requires estimate of the optimal restricted GMM estimator, computation of the $C(\alpha)$ test statistic requires only a consistent estimate of the parameter vector. For both tests, we provide analytical results for their asymptotic distributions within the context of our high order SDPD model.

Within the ML framework, Davidson and MacKinnon (1987), Saikkonen (1989) and Bera and Yoon (1993) show that the conventional LM tests are not robust to local mis-specifications in the alternative models. That is, the conventional LM tests have non-central chi-squared distributions 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 conventional GMM-gradient test is a non-central chi-squared distribution when the alternative model deviates from the true data generating process. In such a case, the conventional LM and GMM-gradient tests 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-square distributions irrespective of the local deviations of the alternative models from the true data generating process.

Following Bera et al. (2010), we construct adjusted GMM-gradient tests for an high order 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. 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 high order SDPD model, we analytically show the asymptotic distribution of robust tests under both the null and local alternative hypotheses. Our suggested test statistics are valid for the asymptotic case where the the number of individuals is large and the number of time period can be large or fixed. We investigate the size and power properties of our suggested robust tests through a Monte Carlo simulation. The simulation results are in line with our theoretical findings and indicate that the robust tests have good size and power properties. Through an empirical illustration from the macro growth literature, we complement the findings of our Monte Carlo study on the finite sample properties of the proposed robust tests.

The rest of this paper is organized as follows. Section 2 presents the higher order SDPD model and its assumptions. Section 3 lays out the details of the GMM 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 design and presents the results. Section 6 presents the empirical illustration. Section 7 ends the paper with concluding remarks. Some of the technical derivations and simulation results are relegated to an appendix.

2 The Model Specification and Assumptions

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

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

for $t = 1, 2, \ldots, T$, where $Y_{nt} = (y_{1t}, y_{2t}, \ldots, y_{nt})^{'}$ is the $n \times 1$ vector of a dependent variable, X_n is the $n \times k_x$ matrix of non-stochastic exogenous variables with a matching parameter vector β_0, and $V_{nt} = (v_{1t}, \ldots, v_{nt})^{'}$ is the $n \times 1$ vector of disturbances (or innovations). The spatial lags of 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 $j = 1, \ldots, p$. Here, W_{nj}s are the $n \times n$ spatial weight matrices of known constants with zero diagonal elements, $\lambda_0 = (\lambda_{10}, \ldots, \lambda_{p0})^{'}$ and $\rho_0 = (\rho_{10}, \ldots, \rho_{p0})^{'}$ are the spatial autoregressive parameters. The individual fixed effects are denoted by $\mathbf{c}_{n0} = (c_{1,0}, \ldots, c_{n,0})^{'}$ and the time fixed 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, we assume that Y_{n0} is observable. Let Θ be the parameter space of the model. In order to distinguish the true parameter vector from other possible values in Θ, we state the model with the true parameter vector $\theta_0 = \left(\lambda_0^{'}, \delta_0^{'}\right)^{'}$, where $\delta_0 = \left(\gamma_0, \rho_0^{'}, \beta_0^{'}\right)^{'}$. Furthermore, for notational simplicity we let $S_n(\lambda) = \left(I_n - \sum_{j=1}^{p} \lambda_j W_{nj}\right)$, $S_n = S_n(\lambda_0)$, $A_n = S_n^{-1}\left(\gamma_0 I_n + \sum_{j=1}^{p} \rho_j W_{nj}\right)$, $G_{nj}(\lambda) = W_{nj} S_n^{-1}(\lambda)$, and $G_{nj} = G_{nj}(\lambda_0)$.

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 orthonormal eigenvector matrix $\left[F_{T,T-1}, \frac{1}{\sqrt{T}} l_T\right]$ of $J_T = \left(I_T - \frac{1}{T} l_T l_T^{'}\right)$, where $F_{T,T-1}$ is the $T \times (T-1)$ eigenvectors matrix corresponding to the eigenvalue one and l_T is the $T \times 1$ vector of ones corresponding to the eigenvalue zero.[1] This orthonormal transformation can be applied by writing the model in an $n \times T$ system. Hence, the dependent variable is transformed as $\left[Y_{n1}, Y_{n2}, \ldots, Y_{nT}\right] \times F_{T,T-1} = \left[Y_{n1}^{*}, Y_{n2}^{*}, \ldots, Y_{n,T-1}^{*}\right]$, and also $\left[Y_{n0}, Y_{n1}, \ldots, Y_{n,T-1}\right] \times F_{T,T-1} = \left[Y_{n0}^{(*,-1)}, Y_{n1}^{(*,-1)}, \ldots, Y_{n,T-2}^{(*,-1)}\right]$. Similarly, $\left[X_{nj,1}, X_{nj,2}, \ldots, X_{nj,T}\right] \times F_{T,T-1} = \left[X_{nj,1}^{*}, X_{nj,2}^{*}, \ldots, X_{nj,T-1}^{*}\right]$ for $j = 1, \ldots, k_x$, $\left[V_{n1}, V_{n2}, \ldots, V_{nT}\right] \times F_{T,T-1} = \left[V_{n1}^{*}, V_{n2}^{*}, \ldots, V_{n,T-1}^{*}\right]$, and $\left[\alpha_{10}, \alpha_{20}, \ldots, \alpha_{T0}\right] \times F_{T,T-1} = \left[\alpha_{10}^{*}, \alpha_{20}^{*}, \ldots, \alpha_{T-1,0}^{*}\right]$. Since the column of $\left[F_{T,T-1}, \frac{1}{\sqrt{T}} l_T\right]$ are orthonormal, we have $[\mathbf{c}_{n0}, \mathbf{c}_{n0}, \ldots, \mathbf{c}_{n0}] \times F_{T,T-1} = 0_{n\times(T-1)}$. Thus, the transformed model does not include the individual fixed effects and can be written as

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

for $t = 1, \ldots, T-1$. Note that the effective sample size of transformed model in (2.2) is $N = n(T-1)$. We consider the forward orthogonal difference (FOD) transformation for

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

the orthonormal transformation. Hence, the terms in (2.2) can be explicitly stated as $V_{nt}^* = \left(\frac{T-t}{T-t+1}\right)^{1/2}\left[V_{nt} - \frac{1}{T-t}\sum_{h=t+1}^{T} V_{nh}\right]$, $Y_{n,t-1}^{(*,-1)} = \left(\frac{T-t}{T-t+1}\right)^{1/2}\left[Y_{n,t-1} - \frac{1}{T-t}\sum_{h=t}^{T-1} Y_{nh}\right]$, and the others terms are defined similarly. Let $\mathbf{V}_{n,T-1}^* = \left(V_{n1}^{*'}, \ldots, V_{n,T-1}^{*'}\right)'$. Then, $\mathrm{Var}\left(\mathbf{V}_{n,T-1}^*\right) = \left(F_{T,T-1}' \otimes I_n\right)\mathrm{E}\left(\mathbf{V}_{nT}\mathbf{V}_{nT}'\right)\left(F_{T,T-1} \otimes I_n\right) = \sigma_0^2 I_N$ 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 $t = 1, \ldots, T-1$,

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

The consistency and asymptotic normality of the GMME of θ_0 are established under Assumptions 1 through 5.[2]

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

Assumption 2. *The spatial weight matrix W_{nj} is uniformly bounded in row and column sums in absolute value for $j = 1, \ldots, p$, 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 λ in a compact parameter space.*

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

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

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

3 The GMM Estimation Approach

The methodology we use to construct our test statistics for various hypotheses in Section 4 is closely related to the GMM estimation method suggested by Lee and Yu (2014) for (2.3). Therefore, in this section, we summarize the GMM estimation approach for (2.3) under both large T and finite T scenarios by following Lee and Yu (2014). The model in (2.3) indicates that instrumental variables (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 $Z_{nt}^* = \left[Y_{n,t-1}^{(*,-1)}, W_{n1} Y_{n,t-1}^{(*,-1)}, \ldots, W_{np} Y_{n,t-1}^{(*,-1)}, X_{nt}^*\right]$, $\mathbf{J}_{n,T-1} = I_{T-1} \otimes J_n$, and $\mathbf{V}_{n,T-1}^*(\theta) = \left(V_{n1}^{*'}(\theta), \ldots, V_{n,T-1}^{*'}(\theta)\right)'$ where $V_{nt}^*(\theta) = S_n(\lambda) Y_{nt}^* - Z_{nt}^* \delta - \alpha_t^* l_n$. We consider the

[2] 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

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

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 $\operatorname{tr}(P_{nj}J_n)=0$ for $j=1,\ldots,m$, and $\mathbf{Q}_{n,T-1}=\left(Q_{n1}^{'},\ldots,Q_{n,T-1}^{'}\right)^{'}$ is the $N\times q$ liner IV matrix such that $q\geq k_x+2p+1$.[3] Under Assumptions 1-5, 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 $O\left(\frac{1}{T}\right)$.[4]

Let $\mathrm{vec}_D(\cdot)$ 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 $\mathrm{E}\left(g_{nT}^{'}(\theta_0)\,g_{nT}(\theta_0)\right)$, which can be approximated by

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

where $\omega_{nm,T}=\left[\mathrm{vec}_D(\mathbf{J}_{n,T-1}\mathbf{P}_{n1,T-1}\mathbf{J}_{n,T-1}),\ldots,\mathrm{vec}_D(\mathbf{J}_{n,T-1}\mathbf{P}_{nm,T-1}\mathbf{J}_{n,T-1})\right]$, $\Delta_{nm,T}=\left[\mathrm{vec}(\mathbf{J}_{n,T-1}\mathbf{P}_{n1,T-1}^{'}\mathbf{J}_{n,T-1}),\ldots,\mathrm{vec}(\mathbf{J}_{n,T-1}\mathbf{P}_{nm,T-1}^{'}\mathbf{J}_{n,T-1})\right]^{'}\times\left[\mathrm{vec}(\mathbf{J}_{n,T-1}\mathbf{P}_{n1,T-1}^{s}\mathbf{J}_{n,T-1}),\ldots,\mathrm{vec}(\mathbf{J}_{n,T-1}\mathbf{P}_{nm,T-1}^{s}\mathbf{J}_{n,T-1})\right]$, and $A_n^s=A_n+A_n^{'}$ for any square matrix A_n.

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

$$\widehat{\theta}_{nT}=\mathrm{argmin}_{\theta\in\Theta}\,g_{nT}^{'}(\theta)\,\widehat{\Sigma}_{nT}^{-1}g_{nT}(\theta) \tag{3.3}$$

Under Assumptions 1 - 5, Lee and Yu (2014) show that when both T and n tend to infinity[5]:

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

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

The optimal GMME defined in (3.3) has the advantages in terms of (i) computational burden, (ii) asymptotic bias and efficiency, and (iii) applicability to a wide range of models. The optimal

[3] Here, we give a general argument with respect to the quadratic and linear moment matrices. We state a set of these matrices in Section 5.

[4] 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).

[5] Lee and Yu (2014) state the identification conditions. Here, we simply assume that the parameter vector is identified.

GMME has a computational advantage over the likelihood based estimators, i.e., the MLE and the Bayesian estimators, because it does not require the computation of the determinant of a Jacobian matrix in the likelihood function, which is especially inconvenient when n is large or the model has high order spatial lags. Moreover, the optimal GMME has no asymptotic bias under either finite or large T case, and it can be more efficient than the QMLE when the disturbance terms are simply i.i.d (Lee and Yu 2014). The QMLE requires large T and has a bias of order $O(1/T)$ when T grows with n proportionally (Lee and Yu 2010a; Yu et al. 2008). When T is small, we need to model the initial observations to formulate an appropriate likelihood function that can lead to a consistent (quasi) MLE (Elhorst 2010a; Su and Yang 2015). Finally, the optimal GMME can be considered for models specified with both row-normalized and non-row-normalized spatial weight matrices (Lee and Yu 2014).

4 The GMM Gradient Tests

In this section, we consider the GMM gradient statistics (or the LM statistics) for linear and non-linear hypothesis in the GMM estimation framework described in the previous section. Let $r:\mathbb{R}^{2p+k_x+1}\to\mathbb{R}^{k_r}$ 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 on θ_0 denoted by the null hypothesis $H_0: r(\theta_0)=0$. Define $\widehat{\theta}_{nT,r}=\text{argmax}_{\{\theta:r(\theta)=0\}}\,\mathcal{Q}_n$, where $\mathcal{Q}_n=g^{'}_{nT}(\theta)\,\widehat{\Sigma}^{-1}_{nT}g_{nT}(\theta)$, as a restricted (or constrained) optimal GMME.

In order to give a general argument, consider the following partition of $\theta=\left(\beta^{'},\psi^{'},\phi^{'}\right)^{'}$, where ψ and ϕ are, respectively, $k_\psi\times1$ and $k_\phi\times1$ vectors such that $k_\psi+k_\phi=2p+1$. In the context of our model, ψ and ϕ can be any combinations of the remaining parameters, namely, $\{\lambda,\gamma,\rho\}$. Let $G_a=\frac{1}{N}\frac{\partial g_{nT}(\theta)}{\partial a^{'}}$, $C_a=G^{'}_a(\theta)\,\widehat{\Sigma}^{-1}_{nT}\bar{g}_{nT}(\theta)$, where $a\in\{\beta,\psi,\phi\}$ and $\bar{g}_{nT}=\frac{1}{N}g_{nT}$. Define $G(\theta)=\left(G_\beta(\theta),G_\psi(\theta),G_\phi(\theta)\right)$, and $C(\theta)=\left(C^{'}_\beta(\theta),C^{'}_\psi(\theta),C^{'}_\phi(\theta)\right)^{'}$. Then, the standard LM test statistic for $H_0:r(\theta_0)=0$ is defined in the following way (Newey and West 1987):

$$LM=N\,C^{'}\big(\widehat{\theta}_{nT,r}\big)B^{-1}\big(\widehat{\theta}_{nT,r}\big)C\big(\widehat{\theta}_{nT,r}\big). \tag{4.1}$$

A similar test is the $C(\alpha)$ test.[6] This test is designed to deal with the nuisance parameters when testing the parameter of main interest (Bera and Bilias 2001). Lee and Yu (2012c) 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 high order SDPD model. By the implicit function theorem, the set of k_r restrictions on θ_0 can also be stated as $h(\xi_0)=\theta_0$, where $h:\mathbb{R}^{\bar{q}}\to\mathbb{R}^{2p+k_x+1}$ is continuously differentiable, ξ_0 contains the free parameters, and $\bar{q}=2p+k_x+1-k_r$. Define $\widehat{\xi}_{nT}=\text{argmin}_\xi\, g^{'}_{nT}(h(\xi))\,\widehat{\Sigma}^{-1}_{nT}g_{nT}(h(\xi))$. Then, we have $\widehat{\theta}_{nT,r}=h\big(\widehat{\xi}_{nT}\big)$. Let $\tilde{\xi}_{nT}$ be a consistent estimate of ξ_0. Denote $G_\xi(\theta)=\frac{1}{N}\frac{\partial g_{nT}(\theta)}{\partial\xi^{'}}$, $C_\xi(\theta)=G^{'}_\xi(\theta)\,\widehat{\Sigma}^{-1}_{nT}\bar{g}_{nT}(\theta)$, and $B_\xi(\theta)=G^{'}_\xi(\theta)\,\widehat{\Sigma}^{-1}_{nT}G_\xi(\theta)$. Following the formulation suggested by Breusch and Pagan (1980), we state the $C(\alpha)$ test statistic in the following way

$$C(\alpha)=N\left[C^{'}\big(h(\tilde{\xi}_{nT})\big)B^{-1}\big(h(\tilde{\xi}_{nT})\big)C\big(h(\tilde{\xi}_{nT})\big)-C^{'}_\xi\big(h(\tilde{\xi}_{nT})\big)B^{-1}_\xi\big(h(\tilde{\xi}_{nT})\big)C_\xi\big(h(\tilde{\xi}_{nT})\big)\right]. \tag{4.2}$$

In (4.2), it is important to note that $\tilde{\xi}_{nT}$ can be any consistent estimator. In the case where $\tilde{\xi}_{nT}$ is an

[6]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.

optimal GMME, the $C(\alpha)$ statistic reduces to LM statistic, since $C_\xi\left(h(\tilde{\xi}_{nT})\right) = 0$ by definition.[7] The asymptotic distributions of $C(\alpha)$ and LM are given in the following proposition.

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

$$LM \xrightarrow{d} \chi^2_{k_r}, \quad \text{and} \quad C(\alpha) \xrightarrow{d} \chi^2_{k_r}. \tag{4.3}$$

Proof. See Section C.1. □

The asymptotic argument used in the proof of Proposition 1 is based on Lemma 1. Since the asymptotic results in Lemma 1 are valid under both fixed T and large T cases, our test results are valid under both cases.[8]

Next, we consider the following joint null hypothesis:

$$H_0 : \lambda_0 = 0,\, \rho_0 = 0,\, \gamma_0 = 0, \quad H_A : \text{At least one parameter is not equal to zero.} \tag{4.4}$$

Under the joint null hypothesis, the model reduces to a two-way non-spatial panel data model which can be estimated by an OLS type estimator (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 $\widehat{\theta}_{nT}$ be any other consistent estimator of θ_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 $\vartheta = \left(\lambda', \rho', \gamma\right)'$, $B(\theta) = G'(\theta)\,\widehat{\Sigma}^{-1}_{nT} G(\theta)$ and consider the following partition of $B(\theta)$:

$$B(\theta) = \begin{pmatrix} \underbrace{B_\beta(\theta)}_{k_x \times k_x} & \underbrace{B_{\beta\psi}(\theta)}_{k_x \times k_\psi} & \underbrace{B_{\beta\phi}(\theta)}_{k_x \times k_\phi} \\ \underbrace{B_{\psi\beta}(\theta)}_{k_\psi \times k_x} & \underbrace{B_\psi(\theta)}_{k_\psi \times k_\psi} & \underbrace{B_{\psi\phi}(\theta)}_{k_\psi \times k_\phi} \\ \underbrace{B_{\phi\beta}(\theta)}_{k_\phi \times k_x} & \underbrace{B_{\phi\psi}(\theta)}_{k_\phi \times k_\psi} & \underbrace{B_\phi(\theta)}_{k_\phi \times k_\phi} \end{pmatrix}, \tag{4.5}$$

where ψ and ϕ can be any combinations of $\{\lambda, \gamma, \rho\}$. Then, the LM test statistic for the joint null hypothesis can be expressed as

$$LM_J(\tilde{\theta}_{nT}) = N\, C'_\vartheta(\tilde{\theta}_{nT}) \left[B_{\vartheta\cdot\beta}\left(\tilde{\theta}_{nT}\right)\right]^{-1} C_\vartheta(\tilde{\theta}_{nT}), \tag{4.6}$$

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

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

[7]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 Bilias (2001).

[8]Note that Lemma 1 is stated for a large T case. In the case of a finite T case, $\text{plim}_{n,T\to\infty}$ should be replaced with $\text{plim}_{n\to\infty}$ in Lemma 1.

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

$$C_J(\alpha) = N\left[C'\big(\widehat{\theta}_{nT}\big)B^{-1}\big(\widehat{\theta}_{nT}\big)C\big(\widehat{\theta}_{nT}\big) - C'_\beta\big(\widehat{\theta}_{nT}\big)B_\beta^{-1}\big(\widehat{\theta}_{nT}\big)C_\beta\big(\widehat{\theta}_{nT}\big)\right]. \tag{4.8}$$

The properties of the LM test can be investigated under a sequence of local alternatives (Bera and Bilias 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 there is local parametric misspecification in the alternative model used to formulate the test statistics. We consider similar robust LM tests within the context of our model. In order to give a general result, we consider the following $LM_\psi\big(\tilde{\theta}_{nT}\big)$ statistic for $H_0^\psi : \psi_0 = \psi_\star$ when $H_0^\phi : \phi_0 = \phi_\star$ holds.

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

where $B_{\psi\cdot\beta}\big(\tilde{\theta}_{nT}\big) = B_\psi\big(\tilde{\theta}_{nT}\big) - B_{\psi\beta}\big(\tilde{\theta}_{nT}\big)B_\beta^{-1}\big(\tilde{\theta}_{nT}\big)B_{\beta\psi}\big(\tilde{\theta}_{nT}\big)$ and $\tilde{\theta}_{nT} = \big(\tilde{\beta}',\, \psi'_\star,\, \phi'_\star\big)'$ is the constrained optimal GMME. We investigate the asymptotic distribution of LM_ψ under the sequences of local alternatives $H_A^\psi : \psi_0 = \psi_\star + \delta_\psi/\sqrt{N}$, and $H_A^\phi : \phi_0 = \phi_\star + \delta_\phi/\sqrt{N}$, where $\big(\psi'_\star,\, \phi'_\star\big)'$ is the vector of hypothesized values under the null, and δ_ψ and δ_ϕ are bounded vectors. To this purpose, let $\mathcal{G}_a = \operatorname{plim}_{n,T\to\infty} \frac{1}{N}\frac{\partial g_{nT}(\theta_0)}{\partial a'}$ for $a \in \{\beta, \psi, \phi\}$. Define $\mathcal{G} = \big(\mathcal{G}_\beta,\, \mathcal{G}_\psi,\, \mathcal{G}_\phi\big)$ and $\mathcal{H} = \operatorname{plim}_{n,T\to\infty}\big(D_{nT} + R_{nt}\big)'\widehat{\Sigma}_{nT}^{-1}\big(D_{nT} + R_{nt}\big)$.[9] We consider the following partition of $\mathcal{H}$:

$$\mathcal{H} = \begin{pmatrix} \underbrace{\mathcal{H}_\beta}_{k_x\times k_x} & \underbrace{\mathcal{H}_{\beta\psi}}_{k_x\times k_\psi} & \underbrace{\mathcal{H}_{\beta\phi}}_{k_x\times k_\phi} \\ \underbrace{\mathcal{H}_{\psi\beta}}_{k_\psi\times k_x} & \underbrace{\mathcal{H}_\psi}_{k_\psi\times k_\psi} & \underbrace{\mathcal{H}_{\psi\phi}}_{k_\psi\times k_\phi} \\ \underbrace{\mathcal{H}_{\phi\beta}}_{k_\psi\times k_x} & \underbrace{\mathcal{H}_{\phi\psi}}_{k_\phi\times k_\psi} & \underbrace{\mathcal{H}_\phi}_{k_\phi\times k_\phi} \end{pmatrix}. \tag{4.10}$$

The distribution of (4.9), under H_A^ψ and H_A^ϕ, can be investigated from the first order Taylor expansions of pseudo-scores $C_\psi\big(\tilde{\theta}_{nT}\big)$ and $C_\beta\big(\tilde{\theta}_{nT}\big)$ around $\theta_0 = \big(\beta'_0,\, \psi'_0,\, \phi'_0\big)'$ under the local alternative hypotheses H_A^ψ and H_A^ϕ. These expansions can be written as

$$\begin{aligned}\sqrt{N}\, C_\psi(\tilde{\theta}_{nT}) &= \sqrt{N}\, C_\psi(\theta_0) - G'_\psi(\theta_0)\widehat{\Sigma}_{nT}^{-1}G_\psi(\bar{\theta})\delta_\psi - G'_\psi(\theta_0)\widehat{\Sigma}_{nT}^{-1}G_\phi(\bar{\theta})\delta_\phi \\ &\quad + \sqrt{N}\, G'_\psi(\theta_0)\widehat{\Sigma}_{nT}^{-1}G_\beta(\bar{\theta})\big(\tilde{\beta}_{nT} - \beta_0\big) + o_p(1),\end{aligned} \tag{4.11}$$

$$\begin{aligned}\sqrt{N}\, C_\beta(\tilde{\theta}_{nT}) &= \sqrt{N}\, C_\beta(\theta_0) - G'_\beta(\theta_0)\widehat{\Sigma}_{nT}^{-1}G_\psi(\bar{\theta})\delta_\psi - G'_\beta(\theta_0)\widehat{\Sigma}_{nT}^{-1}G_\phi(\bar{\theta})\delta_\phi \\ &\quad + \sqrt{N}\, G'_\beta(\theta_0)\widehat{\Sigma}_{nT}^{-1}G_\beta(\bar{\theta})\big(\tilde{\beta}_{nT} - \beta_0\big) + o_p(1),\end{aligned} \tag{4.12}$$

where $\bar{\theta}$ lies between $\tilde{\theta}_{nT}$ and θ_0, which implies that $\bar{\theta} = \theta_0 + o_p(1)$. By Lemma 1, we have $B(\theta_0) = \mathcal{H} + o_p(1)$, and $G'(\theta_0)\widehat{\Sigma}_{nT} = \mathcal{G}'\Sigma_{nT} + o_p(1)$. Then, from (4.11) and (4.12), we get the

[9] Note that we use the large T case to define $\mathcal{G}$ and $\mathcal{H}$. For the fixed T case, $\operatorname{plim}_{n,T\to\infty}$ should be replaced with $\operatorname{plim}_{n\to\infty}$ in these definitions.

following fundamental result:

$$\sqrt{N}\, C_\psi\big(\tilde{\theta}_{nT}\big) = \big[\mathcal{G}_\psi' \Sigma_{nT}^{-1} - \mathcal{H}_{\psi\beta}\mathcal{H}_\beta^{-1}\mathcal{G}_\beta'\Sigma_{nT}^{-1}\big]\frac{1}{\sqrt{N}}\, g_{nT}(\theta_0) \\ - \big[\mathcal{H}_\psi - \mathcal{H}_{\psi\beta}\mathcal{H}_\beta^{-1}\mathcal{H}_{\beta\psi}\big]\delta_\psi - \big[\mathcal{H}_{\psi\phi} - \mathcal{H}_{\psi\beta}\mathcal{H}_\beta^{-1}\mathcal{H}_{\beta\phi}\big]\delta_\phi + o_p(1). \tag{4.13}$$

By Lemma 1, we have $\frac{1}{\sqrt{N}} g_{nT}(\theta_0) \xrightarrow{d} N\big(0, \mathrm{plim}_{n,T\to\infty}\Sigma_{nT}\big)$, and thus (4.13) implies that

$$\sqrt{N}\, C_\psi\big(\tilde{\theta}_{nT}\big) \xrightarrow{d} N\big(-\mathcal{H}_{\psi\cdot\beta}\delta_\psi - \mathcal{H}_{\psi\phi\cdot\beta}\delta_\phi,\, \mathcal{H}_{\psi\cdot\beta}\big), \tag{4.14}$$

where $\mathcal{H}_{\psi\cdot\beta} = \big[\mathcal{H}_\psi - \mathcal{H}_{\psi\beta}\mathcal{H}_\beta^{-1}\mathcal{H}_{\beta\psi}\big]$, and $\mathcal{H}_{\psi\phi\cdot\beta} = \big[\mathcal{H}_{\psi\phi} - \mathcal{H}_{\psi\beta}\mathcal{H}_\beta^{-1}\mathcal{H}_{\beta\phi}\big]$. Hence, $LM_\psi\big(\tilde{\theta}_{nT}\big) \xrightarrow{d} \chi^2_{k_\psi}(\vartheta_1)$ under H_A^ψ and H_A^ϕ, 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.[10] This result indicates that $LM_\psi\big(\tilde{\theta}_{nT}\big)$ has a non-central chi-square distribution under H_0^ψ and H_A^ϕ, and therefore the test statistic leads to an over-rejection of the null hypothesis (Davidson and MacKinnon 1987; Saikkonen 1989). Following Bera et al. (2010), we can adjust the asymptotic mean of $\sqrt{N}\, C_\psi\big(\tilde{\theta}_{nT}\big)$ such that it has a multivariate normal distribution centered at a zero vector under H_0^ψ and H_A^ϕ. Let $C_\psi^\star\big(\tilde{\theta}_{nT}\big)$ be the resulting adjusted score, and $LM_\psi^\star\big(\tilde{\theta}_{nT}\big)$ be the robust test statistic formulated with $C_\psi^\star\big(\tilde{\theta}_{nT}\big)$. The adjustment method suggested in Bera et al. (2010) consists of two steps. In the first step, the asymptotic distribution of $\sqrt{N}C_\phi\big(\tilde{\theta}_{nT}\big)$, under H_0^ψ and H_A^ϕ, is used to adjust the asymptotic mean of $\sqrt{N}C_\psi\big(\tilde{\theta}_{nT}\big)$. This process yields the adjusted score $\sqrt{N}C_\psi^\star\big(\tilde{\theta}_{nT}\big)$ that has zero asymptotic mean. In the second step, the asymptotic variance of $\sqrt{N}C_\psi^\star\big(\tilde{\theta}_{nT}\big)$ is determined to formulate $LM_\psi^\star\big(\tilde{\theta}_{nT}\big)$.

In the following proposition, we summarize results on the asymptotic distributions of $LM_\psi\big(\tilde{\theta}_{nT}\big)$ and $LM_\psi^\star\big(\tilde{\theta}_{nT}\big)$.

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

1. *Under H_A^ψ and H_A^ϕ, we have*

$$LM_\psi\big(\tilde{\theta}_{nT}\big) \xrightarrow{d} \chi^2_{k_\psi}(\vartheta_1), \tag{4.15}$$

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^ψ and H_0^ϕ, we have*

$$LM_\psi\big(\tilde{\theta}_{nT}\big) \xrightarrow{d} \chi^2_{k_\psi}(\vartheta_2), \tag{4.16}$$

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

3. *Under H_0^ψ and H_A^ϕ, we have*

$$LM_\psi\big(\tilde{\theta}_{nT}\big) \xrightarrow{d} \chi^2_{k_\psi}(\vartheta_3), \tag{4.17}$$

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

[10] For the definition of non-central chi-squared distribution, see Anderson (2003, pp.81-82).

4. *Let* $C^{\star}_{\psi}\big(\tilde{\theta}_{nT}\big) = \big[C_{\psi}\big(\tilde{\theta}_{nT}\big) - B_{\psi\phi\cdot\beta}\big(\tilde{\theta}_{nT}\big)B^{-1}_{\phi\cdot\beta}\big(\tilde{\theta}_{nT}\big)C_{\phi}\big(\tilde{\theta}_{nT}\big)\big]$ *be the adjusted pseudo-score, where* $B_{\psi\phi\cdot\beta}\big(\tilde{\theta}_{nT}\big) = B_{\psi\phi}\big(\tilde{\theta}_{nT}\big) - B_{\psi\beta}\big(\tilde{\theta}_{nT}\big)B^{-1}_{\beta}\big(\tilde{\theta}_{nT}\big)B_{\beta\phi}\big(\tilde{\theta}_{nT}\big)$*, and* $B_{\phi\cdot\beta}\big(\tilde{\theta}_{nT}\big) = B_{\phi}\big(\tilde{\theta}_{nT}\big) - B_{\phi\beta}\big(\tilde{\theta}_{nT}\big)B^{-1}_{\beta}\big(\tilde{\theta}_{nT}\big)B_{\beta\phi}\big(\tilde{\theta}_{nT}\big)$*. Under* H^{ψ}_0 *and irrespective of whether* H^{ϕ}_0 *or* H^{ϕ}_A *holds, we have*

$$\begin{aligned} LM^{\star}_{\psi}\big(\tilde{\theta}_{nT}\big) &= N\, C^{\star\prime}_{\psi}\big(\tilde{\theta}_{nT}\big)\big[B_{\psi\cdot\beta}\big(\tilde{\theta}_{nT}\big) - B_{\psi\phi\cdot\beta}\big(\tilde{\theta}_{nT}\big)B^{-1}_{\phi\cdot\beta}\big(\tilde{\theta}_{nT}\big)B^{\prime}_{\psi\phi\cdot\beta}\big(\tilde{\theta}_{nT}\big)\big]^{-1}C^{\star}_{\psi}\big(\tilde{\theta}_{nT}\big) \\ &\xrightarrow{d} \chi^2_{k_{\psi}}. \end{aligned} \tag{4.18}$$

5. *Under* H^{ψ}_A *and irrespective of whether* H^{ϕ}_0 *or* H^{ϕ}_A *holds, we have*

$$LM^{\star}_{\psi}\big(\tilde{\theta}_{nT}\big) \xrightarrow{d} \chi^2_{k_{\psi}}\left(\vartheta_4\right), \tag{4.19}$$

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

Proof. See Section C.2. □

There are three important observations regarding to the results presented in Proposition 2. First, the standard one directional test $LM_{\psi}\big(\tilde{\theta}_{nT}\big)$ has a non-central chi-square distribution when the alternative model is misspecified, i.e., when the alternative model includes ϕ_0. The non-centrality parameter is $\vartheta_3 = \delta^{\prime}_{\phi}\mathcal{H}^{\prime}_{\psi\phi\cdot\beta}\mathcal{H}^{-1}_{\psi\cdot\beta}\mathcal{H}_{\psi\phi\cdot\beta}\delta_{\phi}$, which would be zero if and only if $\mathcal{H}_{\psi\phi\cdot\beta} = 0$. Second, the robust test $LM^{\star}_{\psi}\big(\tilde{\theta}_{nT}\big)$ has a central chi-square distribution even when the alternative model is locally misspecified. Finally, $LM^{\star}_{\psi}\big(\tilde{\theta}_{nT}\big)$ has less asymptotic power than $LM_{\psi}\big(\tilde{\theta}_{nT}\big)$, since $\vartheta_2 - \vartheta_4 \geq 0$ under H^{ψ}_A and H^{ϕ}_0.

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: $H^{\lambda}_0 : \lambda_0 = 0$ in the presence of ρ_0 and γ_0.

2. The null hypothesis for the spatial lag terms at time $t-1$: $H^{\rho}_0 : \rho_0 = 0$ in the presence of λ_0 and γ_0.

3. The null hypothesis for the time lag term: $H^{\gamma}_0 : \gamma_0 = 0$ in the presence of λ_0 and ρ_0.

In the following, we provide the test statistic for each hypothesis and leave the detailed derivations to Appendix B. We start with $H^{\lambda}_0 : \lambda_0 = 0$. In the context of this hypothesis, we have $\psi = \lambda$ and $\phi = \big(\rho^{\prime}, \gamma\big)^{\prime}$. Then, the one directional test can be written as

$$LM_{\lambda}\big(\tilde{\theta}_{nT}\big) = N\, C^{\prime}_{\lambda}\big(\tilde{\theta}_{nT}\big)\big[B_{\lambda\cdot\beta}\big(\tilde{\theta}_{nT}\big)\big]^{-1}C_{\lambda}\big(\tilde{\theta}_{nT}\big), \tag{4.20}$$

where $B_{\lambda\cdot\beta}\big(\tilde{\theta}_{nT}\big) = B_{\lambda}\big(\tilde{\theta}_{nT}\big) - B_{\lambda\beta}\big(\tilde{\theta}_{nT}\big)B^{-1}_{\beta}\big(\tilde{\theta}_{nT}\big)B_{\beta\lambda}\big(\tilde{\theta}_{nT}\big)$. The robust version is stated as

$$LM^{\star}_{\lambda}\big(\tilde{\theta}_{nT}\big) = N\, C^{\star\prime}_{\lambda}\big(\tilde{\theta}_{nT}\big)\big[B_{\lambda\cdot\beta}\big(\tilde{\theta}_{nT}\big) - B_{\lambda\phi\cdot\beta}\big(\tilde{\theta}_{nT}\big)B^{-1}_{\phi\cdot\beta}\big(\tilde{\theta}_{nT}\big)B^{\prime}_{\lambda\phi\cdot\beta}\big(\tilde{\theta}_{nT}\big)\big]^{-1}C^{\star}_{\lambda}\big(\tilde{\theta}_{nT}\big), \tag{4.21}$$

where $C^{\star}_{\lambda}\big(\tilde{\theta}_{nT}\big) = \big[C_{\lambda}\big(\tilde{\theta}_{nT}\big) - B_{\lambda\phi\cdot\beta}\big(\tilde{\theta}_{nT}\big)B^{-1}_{\phi\cdot\beta}\big(\tilde{\theta}_{nT}\big)C_{\phi}\big(\tilde{\theta}_{nT}\big)\big]$ is the adjusted score. The following corollary summarizes results on the asymptotic distributions of these statistics.

Corollary 1. *Our stated assumptions ensure the following results.*

1. *$LM_\lambda(\tilde{\theta}_{nT}) \xrightarrow{d} \chi_p^2(\vartheta_5)$ under H_A^λ and H_0^ϕ, and $LM_\lambda(\tilde{\theta}_{nT}) \xrightarrow{d} \chi_p^2(\vartheta_6)$ under H_0^λ and H_A^ϕ, where $\vartheta_5 = \delta_\lambda' \mathcal{H}_{\lambda\cdot\beta}\delta_\lambda$ and $\vartheta_6 = \delta_\phi' \mathcal{H}_{\lambda\phi\cdot\beta}' \mathcal{H}_{\lambda\cdot\beta}^{-1} \mathcal{H}_{\lambda\phi\cdot\beta}\delta_\phi$.*

2. *Under H_0^λ and irrespective of whether H_0^ϕ or H_A^ϕ holds, $LM_\lambda^\star(\tilde{\theta}_{nT})$ has an asymptotic χ_p^2 distribution. Under H_A^λ and irrespective of whether H_0^ϕ or H_A^ϕ holds, we have $LM_\lambda^\star(\tilde{\theta}_{nT}) \xrightarrow{d} \chi_p^2(\vartheta_7)$, where $\vartheta_7 = \delta_\lambda'\big(\mathcal{H}_{\lambda\cdot\beta} - \mathcal{H}_{\lambda\phi\cdot\beta}\mathcal{H}_{\phi\cdot\beta}^{-1}\mathcal{H}_{\lambda\phi\cdot\beta}'\big)\delta_\lambda$.*

Proof. See Section C.3. □

Next, we consider $H_0^\rho : \rho_0 = 0$. In terms of notation in Proposition 2, we have $\psi = \rho$ and $\phi = (\lambda', \gamma)'$ for this hypothesis. The standard one directional test statistic can be written as

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

where $B_{\rho\cdot\beta}(\tilde{\theta}_{nT}) = B_\rho(\tilde{\theta}_{nT}) - B_{\rho\beta}(\tilde{\theta}_{nT}) B_\beta^{-1}(\tilde{\theta}_{nT}) B_{\beta\rho}(\tilde{\theta}_{nT})$. The robust version of $LM_\rho(\tilde{\theta}_{nT})$ is stated as

$$LM_\rho^\star(\tilde{\theta}_{nT}) = N\, C_\rho^{\star\prime}(\tilde{\theta}_{nT})\big[B_{\rho\cdot\beta}(\tilde{\theta}_{nT}) - B_{\rho\phi\cdot\beta}(\tilde{\theta}_{nT}) B_{\phi\cdot\beta}^{-1}(\tilde{\theta}_{nT}) B_{\rho\phi\cdot\beta}'(\tilde{\theta}_{nT})\big]^{-1} C_\rho^\star(\tilde{\theta}_{nT}), \tag{4.23}$$

where $C_\rho^\star(\tilde{\theta}_{nT}) = \big[C_\rho(\tilde{\theta}_{nT}) - B_{\rho\phi\cdot\beta}(\tilde{\theta}_{nT}) B_{\phi\cdot\beta}^{-1}(\tilde{\theta}_{nT}) C_\phi(\tilde{\theta}_{nT})\big]$.

Corollary 2. *Under our stated assumptions, the following results hold.*

1. *$LM_\rho(\tilde{\theta}_{nT}) \xrightarrow{d} \chi_p^2(\vartheta_8)$ under H_A^ρ and H_0^ϕ; and $LM_\rho(\tilde{\theta}_{nT}) \xrightarrow{d} \chi_p^2(\vartheta_9)$ under H_0^ρ and H_A^ϕ, where $\vartheta_8 = \delta_\rho' \mathcal{H}_{\rho\cdot\beta}\delta_\rho$ and $\vartheta_9 = \delta_\phi' \mathcal{H}_{\rho\phi\cdot\beta}' \mathcal{H}_{\rho\cdot\beta}^{-1} \mathcal{H}_{\rho\phi\cdot\beta}\delta_\phi$.*

2. *The asymptotic null distribution of $LM_\rho^\star(\tilde{\theta}_{nT})$ is χ_p^2, irrespective of whether H_0^ϕ or H_A^ϕ holds. Under H_A^ρ and irrespective of whether H_0^ϕ or H_A^ϕ holds, we have $LM_\rho^\star(\tilde{\theta}_{nT}) \xrightarrow{d} \chi_p^2(\vartheta_{10})$, where $\vartheta_{10} = \delta_\rho'\big(\mathcal{H}_{\rho\cdot\beta} - \mathcal{H}_{\rho\phi\cdot\beta}\mathcal{H}_{\phi\cdot\beta}^{-1}\mathcal{H}_{\rho\phi\cdot\beta}'\big)\delta_\rho$.*

Proof. See Section C.3. □

Finally, we consider $H_0^\gamma : \gamma_0 = 0$. Here, we have $\psi = \gamma$ and $\phi = (\lambda', \rho')'$ in terms of notation of Proposition 2. The standard one directional test can be expressed as

$$LM_\gamma(\tilde{\theta}_{nT}) = N\, C_\gamma'(\tilde{\theta}_{nT})\big[B_{\gamma\cdot\beta}(\tilde{\theta}_{nT})\big]^{-1} C_\gamma(\tilde{\theta}_{nT}), \tag{4.24}$$

where $B_{\gamma\cdot\beta}(\tilde{\theta}_{nT}) = B_\gamma(\tilde{\theta}_{nT}) - B_{\gamma\beta}(\tilde{\theta}_{nT}) B_\beta^{-1}(\tilde{\theta}_{nT}) B_{\beta\gamma}(\tilde{\theta}_{nT})$. The robust version is stated as

$$LM_\gamma^\star(\tilde{\theta}_{nT}) = N\, C_\gamma^{\star\prime}(\tilde{\theta}_{nT})\big[B_{\gamma\cdot\beta}(\tilde{\theta}_{nT}) - B_{\gamma\phi\cdot\beta}(\tilde{\theta}_{nT}) B_{\phi\cdot\beta}^{-1}(\tilde{\theta}_{nT}) B_{\gamma\phi\cdot\beta}'(\tilde{\theta}_{nT})\big]^{-1} C_\gamma^\star(\tilde{\theta}_{nT}), \tag{4.25}$$

where $C_\gamma^\star(\tilde{\theta}_{nT}) = \big[C_\gamma(\tilde{\theta}_{nT}) - B_{\gamma\phi\cdot\beta}(\tilde{\theta}_{nT}) B_{\phi\cdot\beta}^{-1}(\tilde{\theta}_{nT}) C_\phi(\tilde{\theta}_{nT})\big]$ is the adjusted score function.

Corollary 3. *Under our stated assumptions, the following results hold.*

1. *$LM_\gamma(\tilde{\theta}_{nT}) \xrightarrow{d} \chi_1^2(\vartheta_{11})$ under H_A^γ and H_0^ϕ; and $LM_\gamma(\tilde{\theta}_{nT}) \xrightarrow{d} \chi_1^2(\vartheta_{12})$ under H_0^γ and H_A^ϕ, where $\vartheta_{11} = \delta_\gamma' \mathcal{H}_{\gamma\cdot\beta}\delta_\gamma$ and $\vartheta_{12} = \delta_\phi' \mathcal{H}_{\gamma\phi\cdot\beta}' \mathcal{H}_{\gamma\cdot\beta}^{-1} \mathcal{H}_{\gamma\phi\cdot\beta}\delta_\phi$.*

2. *The asymptotic null distribution of $LM^{\star}_{\gamma}(\tilde{\theta}_{nT})$ is χ^2_1, irrespective of whether H^{ϕ}_0 or H^{ϕ}_A holds. Finally, under H^{γ}_A and irrespective of whether H^{ϕ}_0 or H^{ϕ}_A holds, we have $LM^{\star}_{\gamma}(\tilde{\theta}_{nT}) \xrightarrow{d} \chi^2_1(\vartheta_{13})$, where $\vartheta_{13} = \delta^{'}_{\gamma}\big(\mathcal{H}_{\gamma\cdot\beta} - \mathcal{H}_{\gamma\phi\cdot\beta}\mathcal{H}^{-1}_{\phi\cdot\beta}\mathcal{H}^{'}_{\gamma\phi\cdot\beta}\big)\delta_{\gamma}$.*

Proof. See Section C.3. □

Our results in Propositions 1 and 2 on the asymptotic properties of test statistics are general enough and can be used to determine the asymptotic properties test statistics for the possible extended versions of (2.1). For example, Lee and Yu (2012b) consider an SDPD model with time varying weight matrices, which has the following transformed version.

$$J_nY^{*}_{nt} = \sum_{j=1}^{p}\lambda_{j0}J_nW_{njt}Y^{*}_{nt} + \gamma_0J_nY^{(*,-1)}_{n,t-1} + \sum_{j=1}^{p}\rho_{j0}J_nW_{njt}Y^{(*,-1)}_{n,t-1} + J_nX^{*}_{nt}\beta_0 + J_nV^{*}_{nt}, \quad (4.26)$$

for $t = 1,\ldots,T-1$. In (4.26), the spatial weight matrices $W_{nj1}, W_{nj2},\ldots,W_{nj,T-1}$ are now time varying. The test statistics for hypotheses about the parameters of (4.26) can easily be obtained by adjusting the test statistics presented in Corollaries 1-3. The adjustment involves the following steps: (i) the set of moment functions in (3.1) remains the same with the change of $\mathbf{V}^{*}_{n,T-1}(\theta) = \big(V^{*'}_{n1}(\theta),\ldots,V^{*'}_{n,T-1}(\theta)\big)^{'}$, where $V^{*}_{nt}(\theta) = S_{nt}(\lambda)Y^{*}_{nt} - Z^{*}_{nt}\delta - \alpha^{*}_{t}l_n$ and $S_{nt}(\lambda) = \big(I_n - \sum_{j=1}^{p}\lambda_jW_{njt}\big)$, and (ii) the term $\mathbf{W}_{nj,T-1} = I_{T-1}\otimes W_{nj}$ in $G_a(\theta)$ is replaced with $\mathbf{W}_{nj,T-1} = \mathrm{Diag}\,(W_{nj1}, W_{nj2},\ldots,W_{nj,T-1})$ in the expressions stated in Appendix B, where $\mathrm{Diag}(\cdot)$ is an operator that creates a block diagonal matrix from a given list of matrices.

Another extended version of (2.1) for $p = 1$ is considered in Anselin et al. (2008) and Lee and Yu (2016), which can be stated as

$$\begin{aligned} Y_{nt} &= \lambda_0W_nY_{nt} + \gamma_0Y_{n,t-1} + \rho_0W_nY_{n,t-1} + X_{nt}\beta_{10} + W_nX_{nt}\beta_{20} + X_{n,t-1}\beta_{30} \\ &\quad + W_nX_{n,t-1}\beta_{40} + c_{n0} + \alpha_{0t}l_n + V_{nt} \end{aligned} \quad (4.27)$$

In (4.27), W_nX_{nt} and $W_nX_{n,t-1}$ can be called the Durbin terms (LeSage and Pace 2009). Let $\mathbb{X}_{nt} = (X_{nt}, W_nX_{nt}, X_{n,t-1}, W_nX_{n,t-1})$ and $\beta_0 = \big(\beta^{'}_{10}, \beta^{'}_{20}, \beta^{'}_{30}, \beta^{'}_{40}\big)^{'}$. Then, (5.1) can be written as

$$Y_{nt} = \lambda_0W_nY_{nt} + \gamma_0Y_{n,t-1} + \rho_0W_nY_{n,t-1} + \mathbb{X}_{nt}\beta_0 + c_{n0} + \alpha_{0t}l_n + V_{nt} \quad (4.28)$$

Thus, the test statistics stated in Corollaries 1-3 become applicable to (4.27) when X_{nt} is replaced with $\mathbb{X}_{nt}$ in the formulation of test statistics.

5 The Monte Carlo Study

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 (2015b). For the model in (2.1), we will focus on the case where $p = 1$:

$$Y_{nt} = \lambda_0W_nY_{nt} + \gamma_0Y_{n,t-1} + \rho_0W_nY_{n,t-1} + X_{nt}\beta_0 + \mathbf{c}_{n0} + \alpha_{t0}l_n + V_{nt}, \quad (5.1)$$

for $t = 1, 2,\ldots,T$. We generate the weights matrix according to (i) Rook contiguity and (ii) Queen 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 (left/right/above or below) to the square of the spatial unit i. In the Queen contiguity, $w_{ij,n} = 1$ if

Table 1: Summary of test statistics

Null hypothesis	Parameter		Test statistic
	Spatial time lag: ρ_0	Time lag: γ_0	
$\mathrm{H}_0 : \lambda_0 = 0$	Set to zero	Set to zero	LM_λ in (4.20)
$\mathrm{H}_0 : \lambda_0 = 0$	Unrestricted, not estimated	Unrestricted, not estimated	$\mathrm{LM}^\star_\lambda$ in (4.21)
$\mathrm{H}_0 : \lambda_0 = 0$	Unrestricted, estimated	Unrestricted, estimated	LM^A_λ in (4.3)
	Contemporaneous spatial lag: λ_0	Time lag: γ_0	
$\mathrm{H}_0 : \rho_0 = 0$	Set to zero	Set to zero	LM_ρ in (4.22)
$\mathrm{H}_0 : \rho_0 = 0$	Unrestricted, not estimated	Unrestricted, not estimated	$\mathrm{LM}^\star_\rho$ in (4.23)
$\mathrm{H}_0 : \rho_0 = 0$	Unrestricted, estimated	Unrestricted, estimated	LM^A_ρ in (4.3)
	Contemporaneous spatial lag: λ_0	Spatial time lag: ρ_0	
$\mathrm{H}_0 : \gamma_0 = 0$	Set to zero	Set to zero	LM_γ in (4.24)
$\mathrm{H}_0 : \gamma_0 = 0$	Unrestricted, not estimated	Unrestricted, not estimated	$\mathrm{LM}^\star_\gamma$ in (4.26)
$\mathrm{H}_0 : \gamma_0 = 0$	Unrestricted, estimated	Unrestricted, estimated	LM^A_γ in (4.3)
$\mathrm{H}_0 : \lambda_0 = 0,\ \rho_0 = 0,\ \gamma_0 = 0$	–	–	LM_J in (4.6)
$\mathrm{H}_0 : \lambda_0 = 0,\ \rho_0 = 0,\ \gamma_0 = 0$	–	–	C_J in (4.8)

the spatial unit j is in a square that is adjacent to or shares a corner with the square of the spatial unit i. In both cases, W_n is row normalized.

We allow for two exogenous regressors. The first one is generated as $X_{1,nt} = \Psi_n + 0.01\, t\, l_n + U_{nt}$, 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, $X_{nt} = \big(X_{1,nt}, W_n X_{2,nt}\big)$ 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 α_{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) $V_{i,nt} \sim \mathrm{Gamma}(1,1) - 1$. 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: $(n,\, T) = \{(100, 10),\ (20, 200)\}$.[11]

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 (DPD) with no spatial effects, i.e., when $\lambda_0 = \rho_0 = 0$ and $\gamma_0 \neq 0$ in (5.1). The second specification is a spatial static panel data (SSPD) model, 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 model 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 marginal tests, their robust counterparts and the conditional tests in the following way: (i) the DPD model for LM_ρ, $\mathrm{LM}^\star_\rho$, LM^A_ρ, LM_λ, $\mathrm{LM}^\star_\lambda$, LM^A_λ; (ii) the SSPD model for LM_ρ, $\mathrm{LM}^\star_\rho$, LM^A_ρ, LM_γ, $\mathrm{LM}^\star_\gamma$, LM^A_γ; (iii) the SDPDW model for LM_ρ, $\mathrm{LM}^\star_\rho$, LM^A_ρ. We let λ_0, γ_0 and ρ_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, 36 and 216 combinations. Re-sampling is carried out for $5,000$ times.

Table 1 shows 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 $\mathrm{H}_0 : \lambda_0 = 0$ is the presence of ρ_0 and γ_0 in the alternative model. All test statis-

[11] For the sake of brevity, we only provide estimation results for $(n,\, T) = (100, 10)$.

tics, except the conditional test statistics $\mathrm{LM}_{\lambda}^{A}$, LM_{ρ}^{A} and LM_{γ}^{A} in Table 1 are computed by the estimates obtained from the two-way model (2WE): $Y_{nt} = X_{nt}\beta_0 + \mathbf{c}_{n0} + \alpha_{t0}l_n + V_{nt}$. The computation of the conditional test statistics require the estimates obtained from the corresponding constrained optimal GMMEs. For all test statistics, we also need to specify the set of moment functions. The set of linear moments consists of $Q_{nt} = \left(Y_{n,t-1},\, W_nY_{n,t-1},\, W_n^2Y_{n,t-1},\, X_{n,t}^{*},\, W_nX_{n,t}^{*},\, W_n^2X_{n,t}^{*}\right)$ for $t = 1, 2, \ldots, T-1$. For the quadratic moments, we employ $P_{n1} = W_n - \mathrm{tr}(W_nJ_n)/(n-1)J_n$ and $P_{n2} = W_n^2 - \mathrm{tr}(W_n^2J_n)/(n-1)J_n$.

5.1 Results on Size Properties

The P value discrepancy plots (or size-discrepancy plots) are generated from the empirical distribution function (edf) of p-values. To see how, let τ denote a test statistic, and τ_j for $j = 1, \ldots, \mathcal{R}$ be the $\mathcal{R}$ realizations of τ generated in a Monte Carlo experiment. Let $F(z)$ denote the cumulative distribution function (cdf) of the asymptotic distribution of τ evaluated at the level z. Then, the p-value associated with τ_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 $\{z_i\}$ for $i = 1, \ldots, m$ from the interval (0, 1). Then, an estimate of the cdf of $p(\tau)$ is given by $\widehat{F}(z_i) = \sum_{j=1}^{\mathcal{R}} \mathbf{1}\left(p(\tau_j) \leq z_i\right)/\mathcal{R}$.[12]The P value discrepancy plot is created by plotting $\widehat{F}(z_i) - z_i$ against z_i under the assumption that the true data generating process is characterized by the null hypothesis.

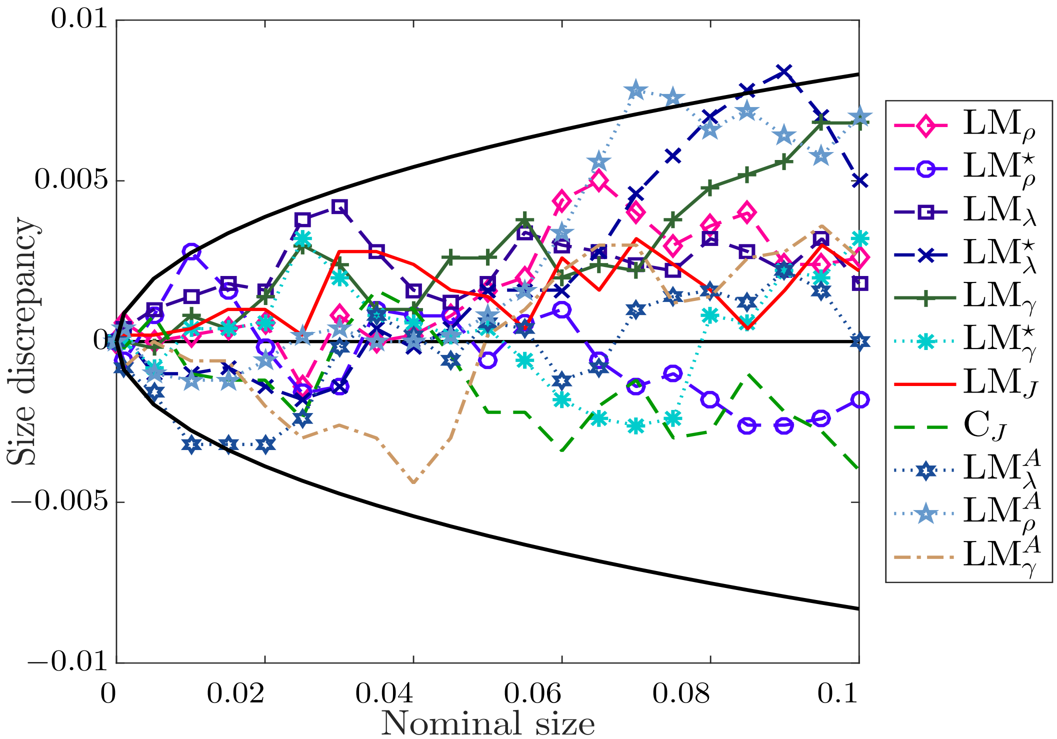
To assess the significance of discrepancies in a P value discrepancy plot, we construct a point-wise 95% confidence interval for a nominal size by using a normal approximation to the binomial distribution (Anselin et al. 1996). Let α denote the nominal size at which the test is carried out. Using a normal approximation to the binomial distribution, a point-wise 95% confidence interval centered on α 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 black solid lines.

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 λ_0, γ_0 and ρ_0, which would yield several 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 deviations at this level only in Tables C.4 through C.7 in Appendix C.4. The general observations on the size properties of tests from Figures 1 and 2, and Tables C.4 through C.7 are listed as follows.

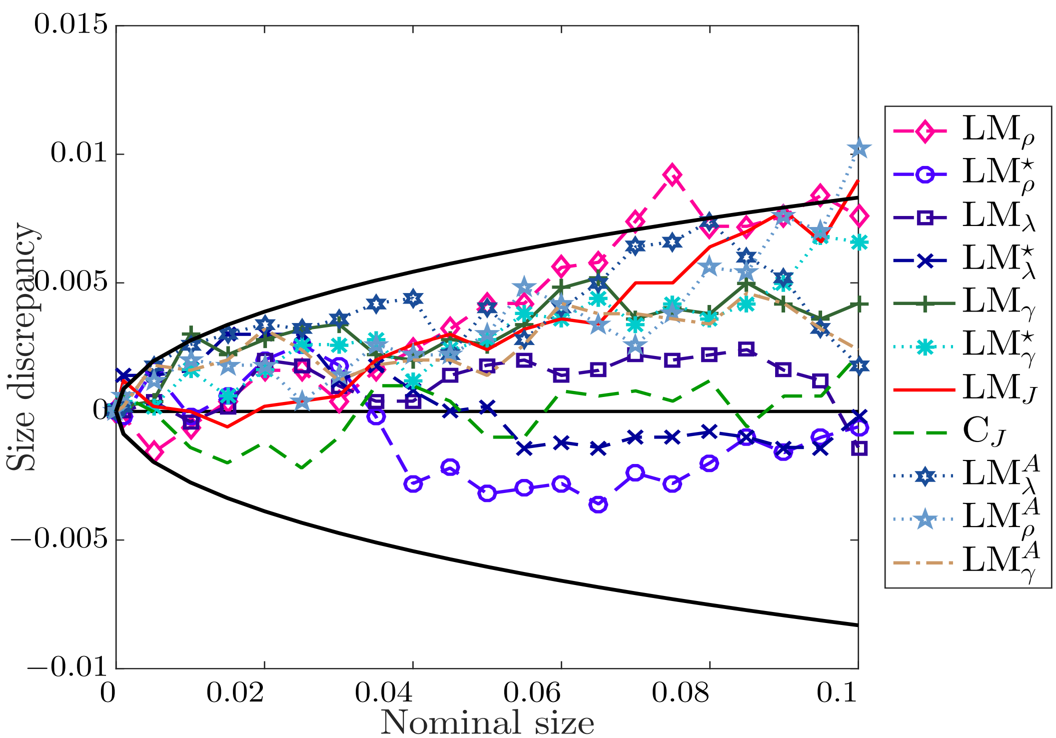
1. Figures 1 and 2 present the size discrepancy plots when the null model is 2WE. The results show that all tests have little size distortions and their size discrepancies generally lie inside the 95% confidence intervals. The size discrepancies are relatively larger in the case of queen weight matrix and non-normal errors.

2. Table C.4 provides some evidences on the magnitude of size distortions as a function of the size of local misspecification in the alternative model, the DPD model. We would expect to see robust versions of one directional tests, $\mathrm{LM}_{\rho}^{\star}$ and $\mathrm{LM}_{\lambda}^{\star}$, to perform better than LM_{ρ} and LM_{λ}, respectively, when the magnitude of misspecification is small. Overall, this seems to

[12]We choose the following sequence and focus on the levels smaller than or equal to 0.1: $\{z_i\}_{i=1}^{m} = \{0.001 : 0.001 : 0.010 \quad 0.015 : 0.005 : 0.990 \quad 0.991 : 0.001 : 0.999\}$.

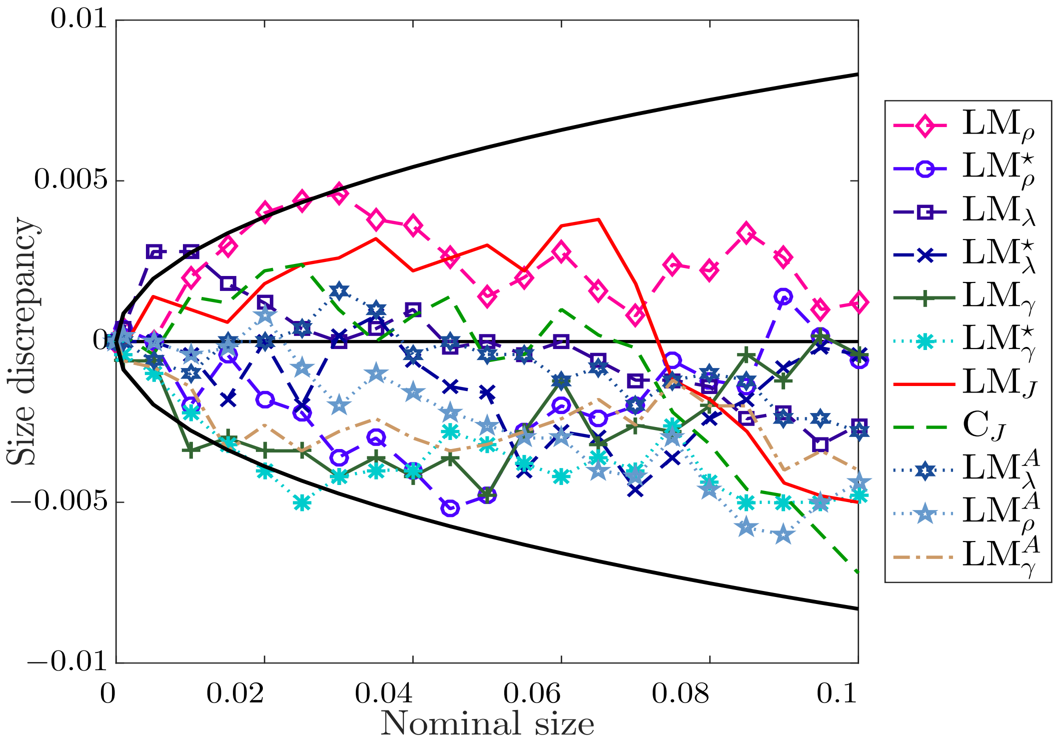


(a) Rook weight matrix and normal errors

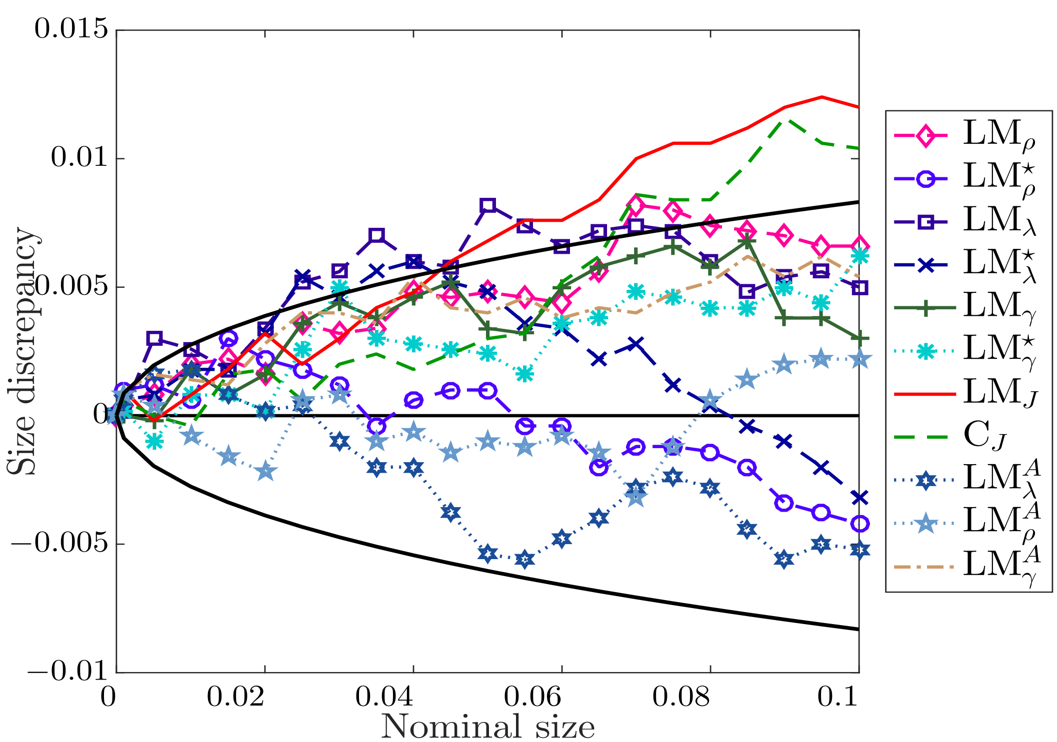


(b) Rook weight matrix and non-normal errors

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



(a) Queen weight matrix and normal errors



(b) Queen weight matrix and non-normal errors

Figure 2: Size discrepancy plots when $(n, T) = (100, 10)$.

be the case. For example, when the value of γ_0 is 0.05 in absolute value in the true model, the actual size of the robust tests are very close to the nominal size of 5%. However, as the misspecification deteriorates, the robust tests do not perform well as expected and are undersized. The conditional tests $\mathrm{LM}_{\lambda}^{A}$ and LM_{ρ}^{A}, which require the estimation of γ_0, are performing relatively better than the robust tests only for large values of γ_0.

3. Similar results hold for Table C.5 as well, the robust versions of one directional tests, $\mathrm{LM}_{\rho}^{\star}$ and $\mathrm{LM}_{\gamma}^{\star}$, perform better than LM_{ρ} and LM_{γ}, respectively, when λ_0 deviates locally from zero in the null model. The performance of conditional tests LM_{ρ}^{A} and LM_{γ}^{A}, which require the estimation of λ_0, is again comparable with the corresponding robust test for small values of λ_0.

4. Tables C.6 and C.7 confirm our previous findings: $\mathrm{LM}_{\rho}^{\star}$ performs better than LM_{ρ}, when λ_0 and γ_0 deviate locally from zero. For example, in Table C.6, when true values of λ_0 and γ_0 are 0.1, the actual size of $\mathrm{LM}_{\rho}^{\star}$ is 0.045 in the case of normal errors, whereas the actual size of LM_{ρ} is 0.985. In this case, the computation of LM_{ρ}^{A} requires the estimation of λ_0 and γ_0. When the size of misspecification is small in the alternative model, i.e., when λ_0 and γ_0 are close to zero, both LM_{ρ}^{A} and $\mathrm{LM}_{\rho}^{\star}$ perform similarly.

5. Recall that the robust tests use the residuals from the estimation of 2WE model and implements 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 of spatial dependence as well as on the sparseness 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.

6. Finally, Tables C.4, C.6 and C.7 indicate that as the temporal dependence strengthens, i.e., the misspecification in γ_0 gets larger in absolute value, the performance of robust one-directional tests deteriorates relative to their marginal counterparts. This is not surprising in the sense that the bias in the residuals from the estimation of 2WE model increases as the dependence over time strengthens.

5.2 Results on Power Properties

To investigate power properties of all tests, we use the approach described in Davidson and MacKinnon (1998) to generate the size power curves against the actual size obtained under the corresponding 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 denote the resulting edf by $\widetilde{F}(z)$. Second, the data generating process satisfies the null hypothesis, and as before $\widehat{F}(z)$ denotes the resulting edf of p-values. Then, a size-power curve is generated by plotting $\widetilde{F}(z_i)$ against $\widehat{F}(z_i)$ for $i = 1, \ldots, m$. As stated in Davidson and MacKinnon (1998), the size-power curve avoids the size adjustments made to generate the power curves.

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 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. To save space, we instead summarize the results in Tables C.8 through C.12 in Appendix C.4, where the level for all tests is 5%. As we mentioned in the Monte Carlo design, the DPD, SSPD and SDPDW models can be considered as null models for one-directional tests and their robust

Table 2: The null and alternative models for power investigations

Case 1	
Null Models	Alternative Models
2WE	DPD, SSPD, SDPDW, SDPD
Case 2	
Null Models	Alternative Models
DPD, SSPD, SDPDW	SDPDW, SDPD

counterparts. Therefore, we can 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_λ and $\mathrm{LM}^\star_\lambda$ where the null model is the DPD model and the alternative model is SDPDW model. Similarly, for LM_λ and $\mathrm{LM}^\star_\lambda$, the null of the DPD and the alternative of the SDPD would yield another size power curve. We chose to present some representative cases in Figures 3 and 6. The null and alternative models that we use to investigate the power properties are summarized in Table 2.[13]

We will start with the general observations on the power properties of our proposed tests for Case 1. The results are presented in Tables C.8 through C.12 in Appendix C.4.

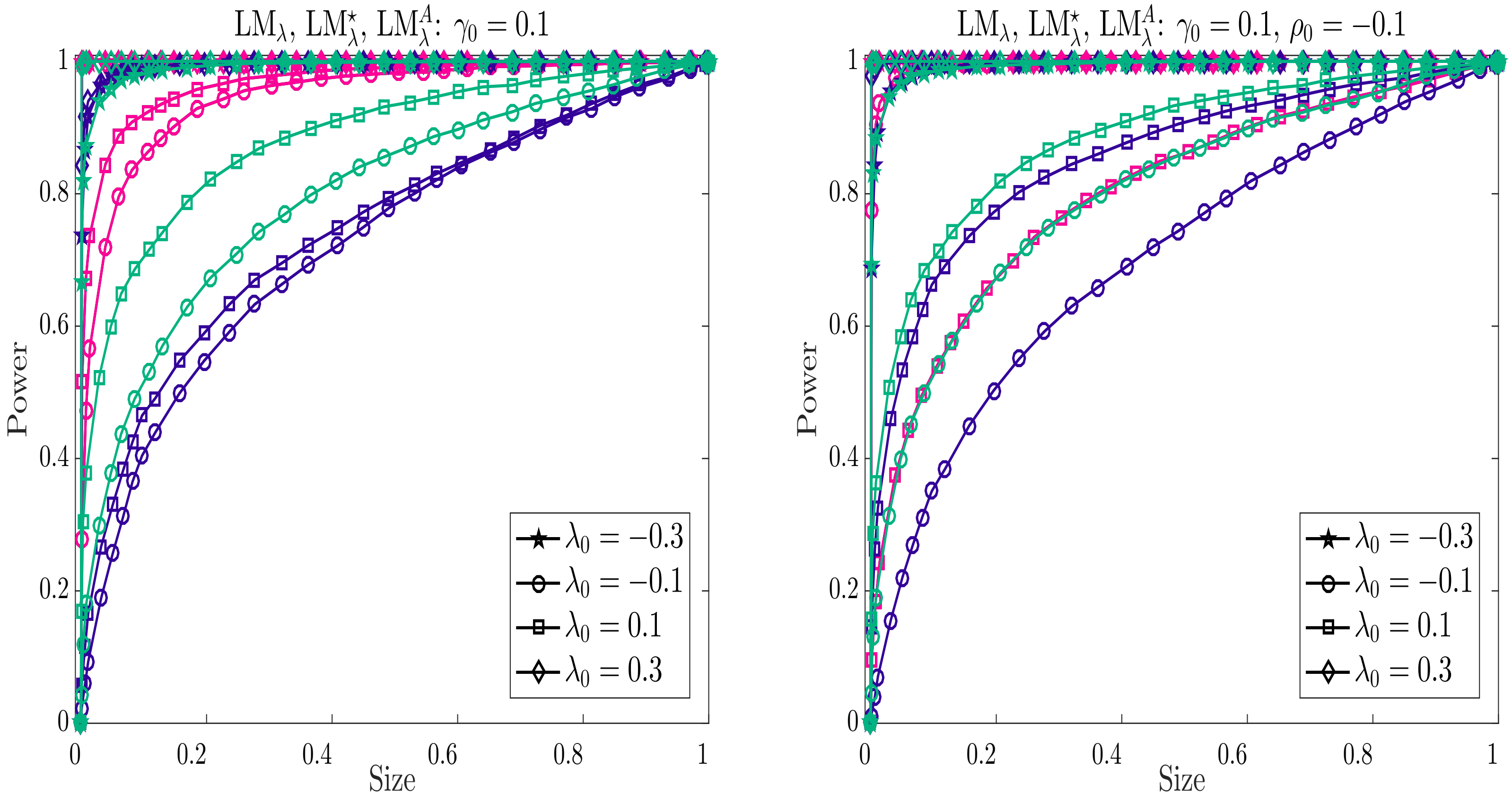
1. Table C.8 shows that (i) the joint test statistics, (ii) LM_γ, $\mathrm{LM}^\star_\gamma$ and LM^A_γ in the case of H_1 : The DPD model, and (iii) LM_λ, $\mathrm{LM}^\star_\lambda$ and LM^A_λ in the case of H_1 : The SSPD model, have desirable power. It is important to note that there is very little loss in power from using $\mathrm{LM}^\star_\gamma$ and $\mathrm{LM}^\star_\lambda$. For example, in the case where H_1: The DPD model and H_0: The 2WE model, the rejection frequencies reported by LM_γ, $\mathrm{LM}^\star_\gamma$ and LM^A_γ are 0.598, 0.586, and 0.558, respectively, when $\gamma_0 = 0.10$.

2. In Table C.8, the robust versions of one directional tests generally perform similar to their non-robust counterparts. However, as the value of γ_0 increases in the DPD model for example, we see that the rejection frequency of $\mathrm{LM}^\star_\rho$ remains low whereas LM_ρ over rejects the true null, confirming the (over) size problem reported in Table C.4. A similar finding applies to $\mathrm{LM}^\star_\lambda$. 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 C.8, $\mathrm{LM}^\star_\gamma$ and $\mathrm{LM}^\star_\rho$ report relatively smaller rejection frequencies, and hence perform better than their non-robust counterparts. Again, in case of spatial dependence in the data generating process, the robust tests are preferable.

3. Table C.9 reveals similar findings. The joint test statistics, LM_γ, $\mathrm{LM}^\star_\gamma$, LM^A_γ, LM_λ, $\mathrm{LM}^\star_\lambda$ and LM^A_λ, generally have desirable power. The rejection frequency reported by $\mathrm{LM}^\star_\rho$ remains low for smaller deviations of λ_0 and γ_0 from zero, whereas LM_ρ over rejects the true null, confirming the (over) size problem reported in Table C.6. Therefore, in case of spatial and temporal dependence in the data generating process, the robust tests are preferable. The rejection rates reported by LM^A_ρ are close to the nominal sizes in all cases.

[13] In this section, we only provide simulation results based on the following design: (i) The queen weight matrix, (ii) the normally distributed errors, (iii) (n, T)=(100,10), and (iv) the nominal size of 0.05. The results based on the rook weight matrix are available upon request. The results based on the gamma distributed error case are similar. Also, for the case of the SDPD model, we focus on some representative tables.

4. Tables C.10, C.11 and C.12 demonstrate that all tests have proper power. The non-robust tests have higher power relative to their robust counterparts in some cases but the differences are generally negligible.

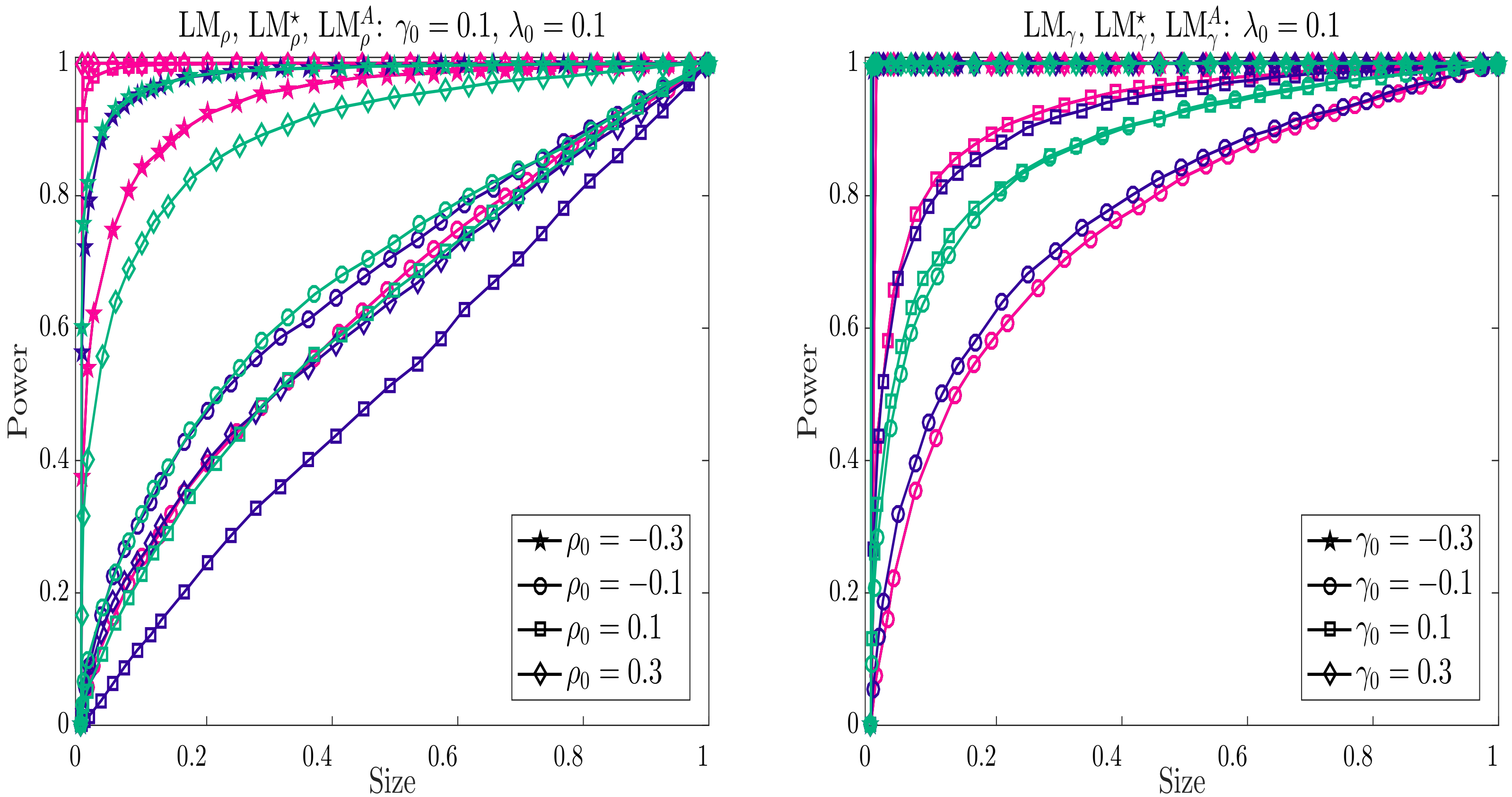


(a) H_0: The DPD model, H_1: The SDPDW model

(b) H_0: The DPD model, H_1: The SDPD model

Notes: Each color represents a test statistic. The size-power curve for the robust statistic is in blue color, for the non-robust one directional statistic is in mulberry color, and for the conditional test statistic is in green color. Markers represent different values of the parameter in the alternative model.

Figure 3: Size-power curves

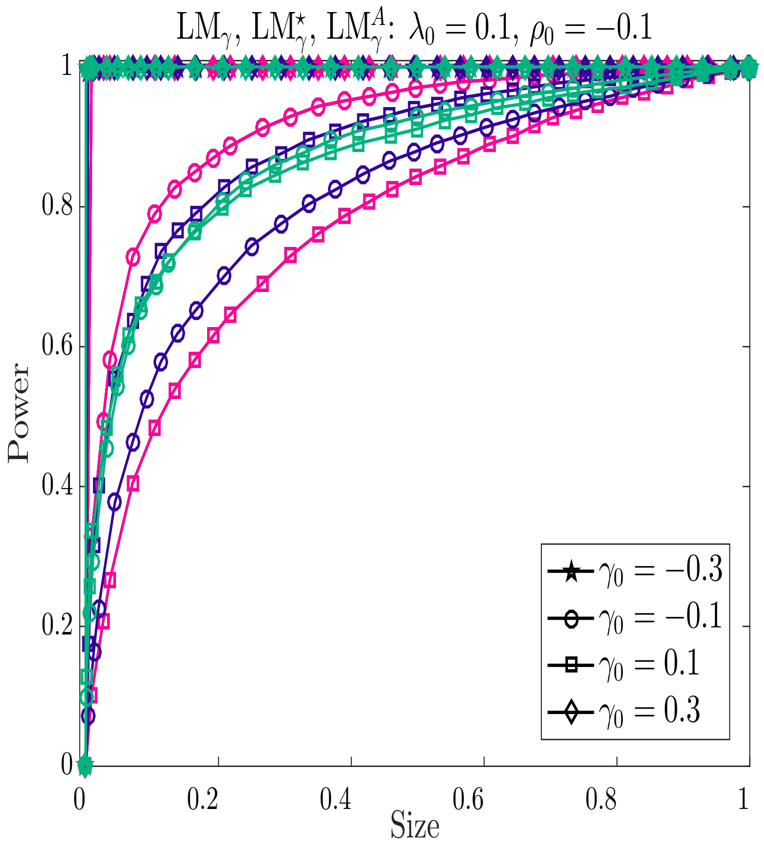


(a) H_0: The DPD model, H_1: The SDPD model

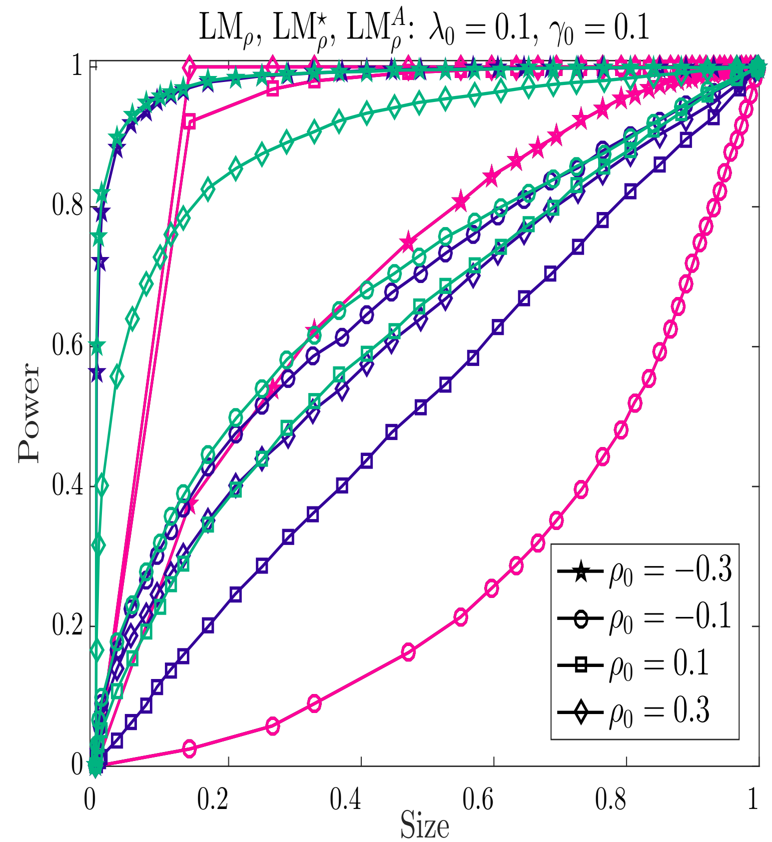
(b) H_0: The SSPD model, H_1: The SDPDW model

Notes: Each color represents a test statistic. The size-power curve for the robust statistic is in blue color, for the non-robust one directional statistic is in mulberry color, and for the conditional test statistic is in green color. Markers represent different values of the parameter in the alternative model.

Figure 4: Size-power curves



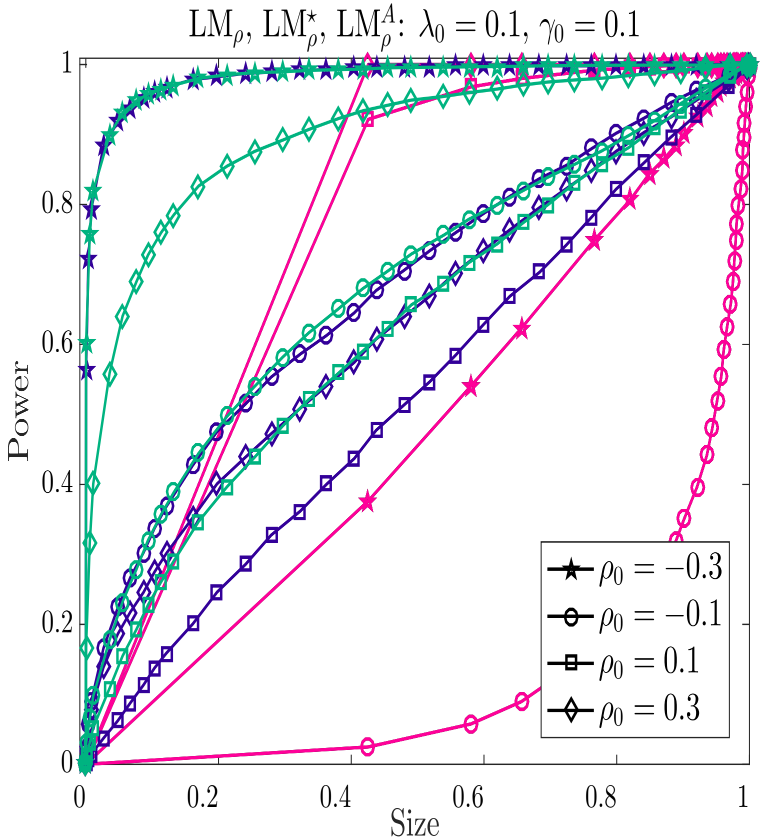
(a) H_0: The SSPD model, H_1: The SDPD model



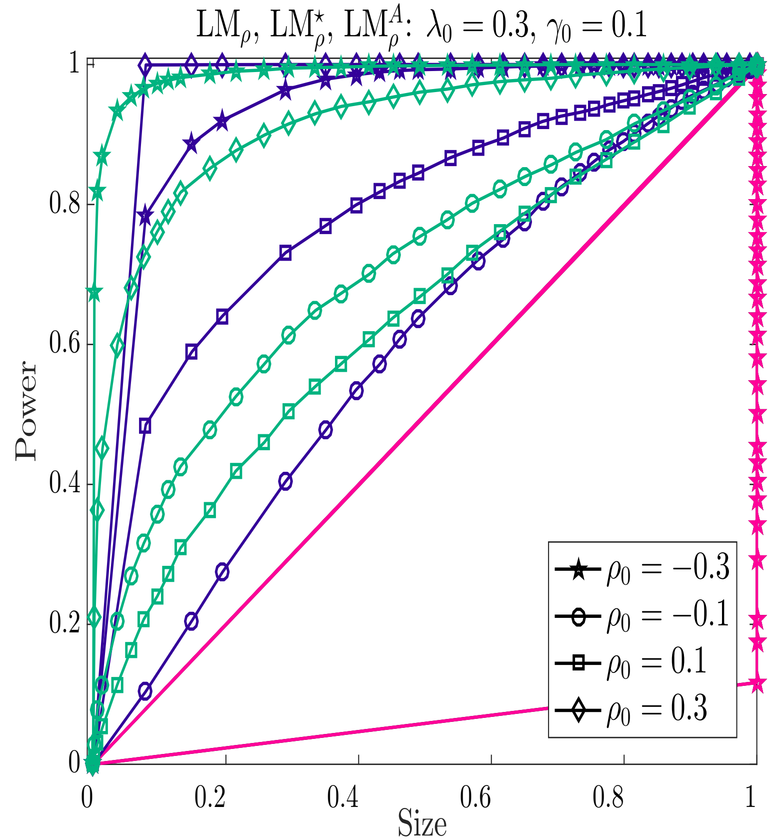
(b) H_0: The SSPD model, H_1: The SDPD model

Notes: Each color represents a test statistic. The size-power curve for the robust statistic is in blue color, for the non-robust one directional statistic is in mulberry color, and for the conditional test statistic is in green color. Markers represent different values of the parameter in the alternative model.

Figure 5: Size-power curves



(a) H_0: The SDPDW model, H_1: The SDPD model



(b) H_0: The SDPDW model, H_1: The SDPD model

Notes: Each color represents a test statistic. The size-power curve for the robust statistic is in blue color, for the non-robust one directional statistic is in mulberry color, and for the conditional test statistic is in green color. Markers represent different values of the parameter in the alternative model.

Figure 6: Size-power curves

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 DPD, SSPD, SDPDW and SDPD models. We will not generate size power curves for these cases as we already summarized the results in Tables C.8 through C.12. Furthermore, for the one directional 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 a size power curve for LM_λ and $\mathrm{LM}^\star_\lambda$ using the DPD model as the null model and the SDPDW model as the alternative. Another size power curve for LM_λ and $\mathrm{LM}^\star_\lambda$ can be obtained from the DPD model as the null model and the SDPD model as the alternative.

In Figures 3 through 6, the size power curves in mulberry color correspond to the non-robust one directional statistic, those in blue color correspond to robust test statistics, and those in green color correspond to the conditional test statistic. Different markers are used to identify varying true values of the spatial parameter in the corresponding alternative model.

The general observations on the power properties of our proposed tests are listed in the following.

1. In Figure 3(a), the null model is the DPD model and the alternative model is the SDPDW model. Generally, $\mathrm{LM}^\star_\lambda$ has relatively less power, except in the case of $\lambda_0 = -0.3$. LM_λ has more power than other two tests in the case where $\lambda_0 = 0.1$ and $\lambda_0 = -0.1$. In Figure 3(b), the null model is the DPD model and the alternative model is the SDPD model. $\mathrm{LM}^\star_\lambda$ is less powerful than LM_λ and LM^A_λ only in the case where $\lambda_0 = -0.1$.

2. In Figure 4(a), the null model is the DPD model and the alternative model is again the SDPD model. $\mathrm{LM}^\star_\rho$ is slightly less powerful than LM_ρ and LM^A_ρ except for the case where $\rho_0 = -0.1$. In Figure 4(b), the null model is the SSPD model and the alternative model is again the SDPDW model. $\mathrm{LM}^\star_\gamma$ and LM_γ behave similarly and both lack power when $\gamma_0 = -0.1$.

3. In Figure 5(a), the null model is the SSPD model and the alternative model is again the SDPD model. Generally, $\mathrm{LM}^\star_\gamma$ behaves similar to LM_γ and LM^A_γ, except in the case of $\gamma_0 = -0.1$. We see that when $\gamma_0 = 0.1$, $\mathrm{LM}^\star_\gamma$ is more powerful than LM_γ. But this picture reverses when $\gamma_0 = -0.1$.

4. In Figure 5(b), the null model is the SSPD model and the alternative model is again the SDPD model. The size power curves confirm the results on the one directional tests of ρ_0 reported in Table C.5. LM_ρ over rejects when the true model involves dependence over space and time. When $\rho_0 = -0.1$, LM_ρ suffers from positive size distortion and lacks of power. The difference in power between LM^A_ρ and $\mathrm{LM}^\star_\rho$ is negligible, except when $\rho_0 = 0.1$.

5. In Figure 6(a), the null model is the SDPDW model and the alternative model is the SDPD model. These plots confirm the results on the one directional tests of ρ_0 reported in Table C.6. Clearly, LM_ρ over rejects when the true model involves dependence over space and time. Again, we see that $\mathrm{LM}^\star_\rho$ lacks power when $\rho_0 = 0.1$. LM^A_ρ has more power than $\mathrm{LM}^\star_\rho$, except in the case of $\rho_0 = -0.3$. The over rejection problem of LM_ρ is also evident in Figure 6(b). $\mathrm{LM}^\star_\rho$ behaves similar to LM^A_ρ, except for the case of $\rho_0 = 0.1$, where it is more powerful.

6 Empirical Illustration

In this section, we provide an empirical illustration based on the spatial Durbin panel models considered in Lee and Yu (2016) for a space-time analysis of regional growth process for 26 OECD countries over the period 1970–2005. The empirical equations considered in Lee and Yu (2016) are based on a spatially augmented growth model which allows interaction among countries through a specification that addresses the international technological interdependence (Ertur and Koch 2007; Lee and Yu 2012a; Lesage and Fischer 2008). The production function of this model is characterized with decreasing returns to physical capital, which ensures a steady state equilibrium for the level of output per-worker. The transition dynamics of the economy to the steady state can be studied by exploring the evolution of physical capital. The speed of transition to the steady-state equilibrium is measured by a convergence rate that is assumed to be the same for all countries. Under this assumption, there are two empirical equations that can be derived as the analogues of the theoretical equation of the transition dynamics of the model: (i) a growth-initial level specification that implies a cross sectional regression model over the period consisting of the time between the initial point ($t_1 = 0$) and an arbitrary point in time ($t_2 = T$) (Ertur and Koch 2007; Mankiw et al. 1992), and (ii) a dynamic panel data specification that divides the whole period T into several shorter time spans (Islam 1995; Knight et al. 1993; Lee and Yu 2012a).

In this study, we consider the dynamic panel data specification, where the whole period T is divided into equal time-spans of τ.[14] There are different suggestions about the appropriate length of time spans in the growth literature. Short time-spans are generally considered inappropriate for studying growth convergence in a panel data framework, because unobservable factors may loom large and the disturbance terms over short time-spans are more likely to be influenced by business cycle fluctuations. Following Lee and Yu (2016), we choose 5-year time intervals, i.e., $\tau = 5$, so that we have 7 data points over the period 1970-2005. The relevant exogenous variables are constructed by taking averages over non-overlapping time-spans. The resulting regression specification in the form of a spatial Durbin model is

$$\begin{aligned}\ln Y_{it} = {} & \lambda_0 \sum_{j=1}^{n} w_{ij,t} \ln Y_{jt} + \rho_0 \sum_{j=1}^{n} w_{ij,t-1} \ln Y_{j,t-1} + \gamma_0 \ln Y_{i,t-1} + c_{i0} + \alpha_{t0} \\ & + \beta_{10} \ln (N_{it} + 0,05) + \beta_{20} \ln S_{it} + \beta_{30} \sum_{j=1}^{n} w_{ij,t} \ln (N_{it} + 0,05) + \beta_{20} \sum_{j=1}^{n} w_{ij,t} \ln S_{it} + V_{it},\end{aligned} \tag{6.1}$$

where Y_{it} is the GDP per worker in country i at time t, N_{it} is the average annual working-age population growth over the last 5-years, $(N_{it} + 0.05)$ is a proxy variable for the sum of working-age population growth rate, the exogenous technological growth rate and the capital depreciation rate, and S_{it} is the average investment share in GDP over the last 5 years.[15] Following Lee and Yu (2016), we use $w_{ij} = \exp(-d_{ij})$, where d_{ij} is the geographic or economic distance between countries. We use three economic distance measures: (i) the bilateral trade flow (Export+Import), (ii) the total exports, and (iii) the total imports.

We use the OLS estimates, obtained from (6.1) when the joint null hypothesis $H_0 : \lambda_0 = \rho_0 = \gamma_0 = 0$ holds, to compute the robust and non-robust test statistics. The estimates of test statistics are presented in Table 3. The test statistics for the joint null hypothesis $H_0 : \lambda_0 = \rho_0 = \gamma_0 = 0$ in the last two columns of Table 3 are significant in all cases, suggesting that at least one parameter is different from zero. The estimates of joint test statistics are relatively smaller for the specification with the geographic distance based weight matrix.

As expected, the estimates reported for non robust tests statistics are larger than the corresponding robust versions. The test statistics LM_γ and $\mathrm{LM}^\star_\gamma$ for $\mathrm{H}_0 : \gamma_0 = 0$ in all cases are in

[14] For advantages of the dynamic panel data specification over the cross-sectional regression model, see Knight et al. (1993), Islam (1995), Caselli et al. (1996), and Acemoglu (2008).

[15] This data set is available in the Journal of Applied Econometrics Data Archive at `http://qed.econ.queensu.ca/jae/2016-v31.1/lee-yu/`.

agreement and reject the null hypothesis. This result highlights the evidence reported in the literature for the presence of conditional convergence. In all cases, the test statistic $\mathrm{LM}_\rho^\star$ for $\mathrm{H}_0 : \rho_0 = 0$ fails to reject the null hypothesis whereas LM_ρ fails to reject the null hypothesis only for the specification with the geographic distance based weight matrix. The Wald statistics for $\mathrm{H}_0 : \rho_0 = 0$ reported in Lee and Yu (2016) that needed estimation of the full model are insignificant in all cases except in the case of the specification with the geographic distance based weight matrix. These results suggest that the space-time lag effect may not be significant for this application. The test statistic $\mathrm{LM}_\lambda^\star$ for $\mathrm{H}_0 : \lambda_0 = 0$ fails to reject the null hypothesis in all cases, while LM_λ fails to reject the null hypothesis only for the specification with the geographic distance based weight matrix. This result is not surprising since the Wald statistics reported in Lee and Yu (2016) for $\mathrm{H}_0 : \lambda_0 = 0$ provide mixed results across cases, suggesting that $\widehat{\lambda}$ may not be significant for this application.

Overall, this empirical illustration clearly indicates that the robust tests statistics are more informative as they account for the presence of non-tested parameters in the model. The results in Table 3 show that by taking account for the presence of time lag effects in the model, the robust tests $\mathrm{LM}_\rho^\star$ and $\mathrm{LM}_\lambda^\star$ can be more informative than the non-robust tests. As reported in Lee and Yu (2016), the Wald counterparts of these robust tests lead to similar conclusions though the former tests are based on a full estimation.

Table 3: Test statistics from the empirical illustration

Weight matrix	LM_λ	$\mathrm{LM}_\lambda^\star$	LM_ρ	$\mathrm{LM}_\rho^\star$	LM_γ	$\mathrm{LM}_\gamma^\star$	LM_J	C_J
Export+Import	11.319	1.112	17.631	1.756	51.427	30.777	53.654	52.504
	[0.001]	[0.292]	[0.000]	[0.185]	[0.000]	[0.000]	[0.000]	[0.000]
Export	7.810	0.412	12.450	0.869	42.510	27.353	44.151	43.968
	[0.005]	[0.521]	[0.000]	[0.351]	[0.000]	[0.000]	[0.000]	[0.000]
Import	12.764	1.416	20.668	2.053	59.528	36.049	61.770	59.055
	[0.000]	[0.234]	[0.000]	[0.152]	[0.000]	[0.000]	[0.000]	[0.000]
Geographic distance	1.089	0.001	1.423	0.020	35.797	34.881	37.481	37.750
	[0.297]	[0.970]	[0.233]	[0.889]	[0.000]	[0.000]	[0.000]	[0.000]

Notes: P values are in brackets.

7 Conclusion

In this paper, we introduced the robust LM tests within the GMM framework for a spatial dynamic panel data model. These tests are robust in the sense that their asymptotic distributions under the null hypothesis are still a central chi-squared distribution when the alternative model is misspecified. On the other hand, when the alternative model is misspecified, the asymptotic null distributions of the conventional LM tests deviate from the central chi-squared distributions. Hence, the robust tests asymptotically obtain the correct size. We derive the asymptotic distributions of our proposed tests 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. Since these tests are robust to the misspecification of the alternative models, they are much more suitable for the detection of the source of dependence in a spatial dynamic panel data model. 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 be made available for practical applications by using the standard statistical software.

Appendix

A A Useful Lemma

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

1. $\frac{1}{N}\mathrm{E}\left(g_{nT}(\theta_0)g_{nT}^{'}(\theta_0)\right) = \Sigma_{nT} + o(1)$ *and* $\widehat{\Sigma}_{nT} = \Sigma_{nT} + o_p(1)$*, where* $\widehat{\Sigma}_{nT}$ *and* Σ_{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* $O(\frac{1}{T})$ *and* $\widehat{\theta}_{nT}$ *is any consistent estimator of* θ_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* θ_0*.*

4. *Let* a_{nT} *be a* $k_a \times (m+q)$ *non-stochastic matrix. Then*

$$\frac{1}{\sqrt{N}}a_{nT}g_{nT}\left(\theta_0\right) \xrightarrow{d} N\left(0, \mathrm{plim}_{n,T\to\infty}\, a_{nT}\Sigma_{nT}a_{nT}^{'}\right) \tag{A.1}$$

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 $\mathbf{V}_{n,T-1}^{*} = \left(V_{n1}^{*'}, \ldots, V_{n,T-1}^{*'}\right)^{'}$ with $V_{nt}^{*}(\theta) = S_n(\lambda)Y_{nt}^{*} - Z_{nt}^{*}\delta - \alpha_t^{*}l_n$, $\mathbf{Y}_{n,T-1}^{*} = \left(Y_{n1}^{*'}, \ldots, Y_{n,T-1}^{*'}\right)^{'}$, $\mathbf{Y}_{n,T-1}^{*,-1} = \left(Y_{n0}^{*'}, \ldots, Y_{n,T-2}^{*'}\right)^{'}$, $\mathbf{X}_{n,T-1}^{*} = \left(X_{n1}^{*'}, \ldots, X_{n,T-1}^{*'}\right)^{'}$ and $\mathbf{W}_{nj,T-1} = I_{T-1} \otimes W_{nj}$. Let the jth column of $G_a\left(\theta\right)$ be denoted by $G_a\left(\theta\right)[:,j]$. We start with $G\left(\theta\right) = \left(G_\lambda\left(\theta\right), G_\gamma\left(\theta\right), G_\rho\left(\theta\right), G_\beta\left(\theta\right)\right)$, where

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

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

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

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

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

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

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

1. $C_{\lambda}(\theta) = G'_{\lambda}(\theta)\begin{pmatrix} O_{11} & O_{12} \\ O_{21} & O_{22} \end{pmatrix}\bar{g}_{nT}(\theta), \quad C_{\gamma}(\theta) = G'_{\gamma}(\theta)\begin{pmatrix} O_{11} & O_{12} \\ O_{21} & O_{22} \end{pmatrix}\bar{g}_{nT}(\theta)$ (B.6)
2. $C_{\rho}(\theta) = G'_{\rho}(\theta)\begin{pmatrix} O_{11} & O_{12} \\ O_{21} & O_{22} \end{pmatrix}\bar{g}_{nT}(\theta), \quad C_{\beta}(\theta) = G'_{\beta}(\theta)\begin{pmatrix} O_{11} & O_{12} \\ O_{21} & O_{22} \end{pmatrix}\bar{g}_{nT}(\theta)$ (B.7)

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

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

Expressions for $H_0^{\lambda}: \lambda_0 = 0$:

$$C_{\lambda}^{\star}\big(\tilde{\theta}_{nT}\big) = \big[C_{\lambda}\big(\tilde{\theta}_{nT}\big) - B_{\lambda\phi\cdot\beta}\big(\tilde{\theta}_{nT}\big)B_{\phi\cdot\beta}^{-1}\big(\tilde{\theta}_{nT}\big)C_{\phi}\big(\tilde{\theta}_{nT}\big)\big], \tag{B.8}$$

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

$$\begin{aligned} B_{\lambda\phi\cdot\beta}\big(\tilde{\theta}_{nT}\big) &= B_{\lambda\phi}\big(\tilde{\theta}_{nT}\big) - B_{\lambda\beta}\big(\tilde{\theta}_{nT}\big)B_{\beta}^{-1}\big(\tilde{\theta}_{nT}\big)B_{\beta\phi}\big(\tilde{\theta}_{nT}\big) \\ &= \big(B_{\lambda\rho}\big(\tilde{\theta}_{nT}\big), B_{\lambda\gamma}\big(\tilde{\theta}_{nT}\big)\big) - B_{\lambda\beta}\big(\tilde{\theta}_{nT}\big)B_{\beta}^{-1}\big(\tilde{\theta}_{nT}\big)\big(B_{\beta\rho}\big(\tilde{\theta}_{nT}\big), B_{\beta\gamma}\big(\tilde{\theta}_{nT}\big)\big) \end{aligned} \tag{B.9}$$

$$\begin{aligned} B_{\phi\cdot\beta}\big(\tilde{\theta}_{nT}\big) &= B_{\phi}\big(\tilde{\theta}_{nT}\big) - B_{\phi\beta}\big(\tilde{\theta}_{nT}\big)B_{\beta}^{-1}\big(\tilde{\theta}_{nT}\big)B_{\beta\phi}\big(\tilde{\theta}_{nT}\big) \\ &= \begin{bmatrix} B_{\rho}\big(\tilde{\theta}_{nT}\big) & B_{\rho\gamma}\big(\tilde{\theta}_{nT}\big) \\ B_{\gamma\rho}\big(\tilde{\theta}_{nT}\big) & B_{\rho}\big(\tilde{\theta}_{nT}\big) \end{bmatrix} - \begin{bmatrix} B_{\rho\beta}\big(\tilde{\theta}_{nT}\big) \\ B_{\gamma\beta}\big(\tilde{\theta}_{nT}\big) \end{bmatrix} B_{\beta}^{-1}\big(\tilde{\theta}_{nT}\big) \Big[B_{\beta\rho}\big(\tilde{\theta}_{nT}\big), B_{\beta\gamma}\big(\tilde{\theta}_{nT}\big)\Big]. \end{aligned} \tag{B.10}$$

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

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

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

$$\begin{aligned} B_{\rho\phi\cdot\beta}\big(\tilde{\theta}_{nT}\big) &= B_{\rho\phi}\big(\tilde{\theta}_{nT}\big) - B_{\rho\beta}\big(\tilde{\theta}_{nT}\big)B_{\beta}^{-1}\big(\tilde{\theta}_{nT}\big)B_{\beta\phi}\big(\tilde{\theta}_{nT}\big) \\ &= \big(B_{\rho\lambda}\big(\tilde{\theta}_{nT}\big), B_{\rho\gamma}\big(\tilde{\theta}_{nT}\big)\big) - B_{\rho\beta}\big(\tilde{\theta}_{nT}\big)B_{\beta}^{-1}\big(\tilde{\theta}_{nT}\big)\big(B_{\beta\lambda}\big(\tilde{\theta}_{nT}\big), B_{\beta\gamma}\big(\tilde{\theta}_{nT}\big)\big), \end{aligned} \tag{B.12}$$

$$\begin{aligned} B_{\phi\cdot\beta}\big(\tilde{\theta}_{nT}\big) &= B_{\phi}\big(\tilde{\theta}_{nT}\big) - B_{\phi\beta}\big(\tilde{\theta}_{nT}\big)B_{\beta}^{-1}\big(\tilde{\theta}_{nT}\big)B_{\beta\phi}\big(\tilde{\theta}_{nT}\big) \\ &= \begin{bmatrix} B_{\lambda}\big(\tilde{\theta}_{nT}\big) & B_{\lambda\gamma}\big(\tilde{\theta}_{nT}\big) \\ B_{\gamma\lambda}\big(\tilde{\theta}_{nT}\big) & B_{\gamma}\big(\tilde{\theta}_{nT}\big) \end{bmatrix} - \begin{bmatrix} B_{\lambda\beta}\big(\tilde{\theta}_{nT}\big) \\ B_{\gamma\beta}\big(\tilde{\theta}_{nT}\big) \end{bmatrix} B_{\beta}^{-1}\big(\tilde{\theta}_{nT}\big) \Big[B_{\beta\lambda}\big(\tilde{\theta}_{nT}\big), B_{\beta\gamma}\big(\tilde{\theta}_{nT}\big)\Big]. \end{aligned} \tag{B.13}$$

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

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

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

$$\begin{aligned} B_{\gamma\phi\cdot\beta}\big(\tilde{\theta}_{nT}\big) &= B_{\gamma\phi}\big(\tilde{\theta}_{nT}\big) - B_{\gamma\beta}\big(\tilde{\theta}_{nT}\big)B_{\beta}^{-1}\big(\tilde{\theta}_{nT}\big)B_{\beta\phi}\big(\tilde{\theta}_{nT}\big) \\ &= \big(B_{\gamma\lambda}\big(\tilde{\theta}_{nT}\big), B_{\gamma\rho}\big(\tilde{\theta}_{nT}\big)\big) - B_{\gamma\beta}\big(\tilde{\theta}_{nT}\big)B_{\beta}^{-1}\big(\tilde{\theta}_{nT}\big)\big(B_{\beta\lambda}\big(\tilde{\theta}_{nT}\big), B_{\beta\rho}\big(\tilde{\theta}_{nT}\big)\big), \end{aligned} \tag{B.15}$$

$$\begin{aligned} B_{\phi\cdot\beta}\big(\tilde{\theta}_{nT}\big) &= B_{\phi}\big(\tilde{\theta}_{nT}\big) - B_{\phi\beta}\big(\tilde{\theta}_{nT}\big)B_{\beta}^{-1}\big(\tilde{\theta}_{nT}\big)B_{\beta\phi}\big(\tilde{\theta}_{nT}\big) \\ &= \begin{bmatrix} B_{\lambda}\big(\tilde{\theta}_{nT}\big) & B_{\lambda\rho}\big(\tilde{\theta}_{nT}\big) \\ B_{\rho\lambda}\big(\tilde{\theta}_{nT}\big) & B_{\rho}\big(\tilde{\theta}_{nT}\big) \end{bmatrix} - \begin{bmatrix} B_{\lambda\beta}\big(\tilde{\theta}_{nT}\big) \\ B_{\rho\beta}\big(\tilde{\theta}_{nT}\big) \end{bmatrix} B_{\beta}^{-1}\big(\tilde{\theta}_{nT}\big) \Big[B_{\beta\lambda}\big(\tilde{\theta}_{nT}\big), B_{\beta\rho}\big(\tilde{\theta}_{nT}\big)\Big]. \end{aligned} \tag{B.16}$$

C Proofs of Propositions

C.1 Proof of Proposition 1

Let $g_{nT}(\theta)$ denote the $m+q$ dimensional vector of empirical moments such that $m+q \geq 2p+k_x+1$. Define the OGMME $\widehat{\theta}_{nT} = \text{argmin}_\theta\, g'_{nT}(\theta)\widetilde{\Sigma}_{nT}^{-1}g_{nT}(\theta)$, where $\widetilde{\Sigma}_{nT}$ is a consistent estimate of Σ_{nT} by Lemma 1. By the implicit function theorem, the set of k_r restrictions on θ_0 can also be stated as $h(\xi_0) = \theta_0$, where $h : \mathbb{R}^{\bar{q}} \to \mathbb{R}^{2p+k_x+1}$ is continuously differentiable, ξ_0 contains the free parameters, and $\bar{q} = 2p + k_x + 1 - k_r$. Define $\widehat{\xi}_{nT} = \text{argmin}_\xi\, g'_{nT}(h(\xi))\,\widehat{\Sigma}_{nT}^{-1}g_{nT}(h(\xi))$. Then, we have $\widehat{\theta}_{c,nT} = h\left(\widehat{\xi}_{nT}\right)$ as the constrained OGMME of θ_0. Let $\tilde{\xi}_{nT}$ denote a $\sqrt{N}$-consistent estimate of ξ_0.

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

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 (2012c). Let

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

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*

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

Then,

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

Proof. Note that

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

By Lemma 1, we have

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

Now, write down

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

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

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

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

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

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

Claim 3. *For any $\sqrt{N}$-consistent estimate of $\tilde{\xi}_{nT}$ of ξ_0, we have $\mathcal{T}_{nT}\big(\mathcal{A}_{nT}^{*}, \tilde{\xi}_{nT}\big) = \mathcal{T}_{nT}\big(\mathcal{A}_{nT}^{*}, \xi_0\big) + o_p(1)$.*

Proof. By assumption $\tilde{\xi}_{nT}$ is a $\sqrt{N}$-consistent estimator. Hence $\sqrt{N}\big(\tilde{\xi}_{nT} - \phi_0\big) = O_p(1)$. By the mean value theorem, we obtain

$$\mathcal{T}_{nT}\big(\mathcal{A}_{nT}, \tilde{\xi}_{nt}\big) = \mathcal{T}_{nT}\big(\mathcal{A}_{nT}, \xi_0\big) + \frac{1}{\sqrt{N}}\frac{\partial\mathcal{T}_{nT}\big(\mathcal{A}_{nT}, \bar{\xi}_{nT}\big)}{\partial\xi'}\sqrt{N}\big(\tilde{\xi}_{nt} - \tilde{\xi}_0\big)$$

where $\bar{\xi}_{nT}$ lies between $\tilde{\xi}_{nt}$ and $\tilde{\xi}_0$. By $\bar{\xi}_{nT} \xrightarrow{p} \xi_0$ and Lemma 1, we obtain

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

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

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

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

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

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

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

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, $\mathcal{T}_{nT}\big(\mathcal{A}^*_{nT},\xi_0\big)$ can be written as

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

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

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

where Ω_{nT} is singular with rank k_r. By Lemma 1, $\frac{1}{\sqrt{N}}g_{nT}(\theta_0) \xrightarrow{d} N\big(0,\text{plim}_{n,T\to\infty}\Sigma_{nT}\big)$. Hence, $\mathcal{T}_{nT}\big(\mathcal{A}^*_{nT},\xi_0\big) \xrightarrow{d} N\big(0,\Omega\big)$. □

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

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

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

Note that

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

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

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

□

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

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

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

$$LM = N\,C'\big(\widehat{\theta}_{nT,r}\big)B^{-1}\big(\widehat{\theta}_{nT,r}\big)C\big(\widehat{\theta}_{nT,r}\big). \tag{C.7}$$

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

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

To evaluate (C.8), we need to consider the limiting behavior of $\sqrt{N}\big(\widehat{\theta}_{nT,r} - \theta_0\big)$. 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

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

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

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

Substituting (C.10) into $\widetilde{LM}$ yields $\widetilde{LM} = \sqrt{N}\,C'(\theta_0)\mathcal{H}^{-1}R'\big(R\mathcal{H}^{-1}R'\big)^{-1}R\mathcal{H}^{-1}\sqrt{N}\,C\big(\theta_0\big) + o_p(1)$. By Lemma 1, we have $R\mathcal{H}^{-1}\sqrt{N}\,C\big(\theta_0\big) \xrightarrow{d} N\big(0, R\mathcal{H}^{-1}R'\big)$, which implies that $\widetilde{LM} \xrightarrow{d} \chi^2_{k_r}$. Then, the desired results follows from the asymptotic equivalence of $\widetilde{LM}$ and LM.

C.2 Proof of Proposition 2

The first three results follows directly from $LM_{\psi}\left(\tilde{\theta}_{nT}\right) \xrightarrow{d} \chi^2_{k_{\psi}}\left(\vartheta_1\right)$ under H^{ψ}_A and H^{ϕ}_A, where $\vartheta_1 = \delta'_{\psi}\mathcal{H}_{\psi\cdot\beta}\delta_{\psi} + \delta'_{\psi}\mathcal{H}_{\psi\phi\cdot\beta}\delta_{\phi} + \delta'_{\phi}\mathcal{H}'_{\psi\phi\cdot\beta}\delta_{\psi} + \delta'_{\phi}\mathcal{H}'_{\psi\phi\cdot\beta}\mathcal{H}^{-1}_{\psi\cdot\beta}\mathcal{H}_{\psi\phi\cdot\beta}\delta_{\phi}$ is the non-centrality parameter. Here, we will prove the last two results. For this purpose, we consider the distribution of $\mathbf{C}_{\psi\phi}\big(\tilde{\theta}_{nT}\big) = \big(C'_{\psi}\big(\tilde{\theta}_{nT}\big),\, C'_{\phi}\big(\tilde{\theta}_{nT}\big)\big)'$. The first order Taylor expansions of $\mathbf{C}_{\psi\phi}\big(\tilde{\theta}_{nT}\big)$ and $C_{\beta}\big(\tilde{\theta}_{nT}\big)$ around $\theta_0 = (\beta'_0,\, \psi'_0,\, \phi'_0)'$ under H^{ψ}_A and H^{ϕ}_A are given by

$$\begin{aligned}\sqrt{N}\,\mathbf{C}_{\psi\phi}(\tilde{\theta}_{nT}) &= \sqrt{N}\,\mathbf{C}_{\psi\phi}(\theta_0) - \mathbf{G}'_{\psi\phi}(\theta_0)\widehat{\Sigma}^{-1}_{nT}\mathbf{G}_{\psi\phi}(\bar{\theta}) \times (\delta'_{\psi},\, \delta'_{\phi})'\\ &\quad + \sqrt{N}\,\mathbf{G}'_{\psi\phi}(\theta_0)\widehat{\Sigma}^{-1}_{nT}G_{\beta}(\bar{\theta}) \times \big(\tilde{\beta}_{nT} - \beta_0\big) + o_p(1),\end{aligned} \tag{C.11}$$

$$\begin{aligned}\sqrt{N}\,C_{\beta}\big(\tilde{\theta}_{nT}\big) &= \sqrt{N}\,C_{\beta}(\theta_0) - G'_{\beta}(\theta_0)\widehat{\Sigma}^{-1}_{nT}\mathbf{G}_{\psi\phi}(\bar{\theta}) \times (\delta'_{\psi},\, \delta'_{\phi})'\\ &\quad + \sqrt{N}\,G'_{\beta}(\theta_0)\widehat{\Sigma}^{-1}_{nT}G_{\beta}(\bar{\theta}) \times \big(\tilde{\beta}_{nT} - \beta_0\big) + o_p(1),\end{aligned} \tag{C.12}$$

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

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

where $\mathcal{G}_{\psi\phi} = \big(\mathcal{G}_{\psi},\ \mathcal{G}_{\phi}\big)$, $\mathbf{H}_{\psi\phi\beta} = \big(\mathcal{H}_{\psi\beta}^{'},\ \mathcal{H}_{\phi\beta}^{'}\big)^{'}$, 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 $\sqrt{N}\,\mathbf{C}_{\psi\phi}\big(\tilde{\theta}_{nT}\big)$ under H_0^{ψ} and H_A^{ϕ} from (C.13). It can be shown that

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

The result in (C.14) can be used to determine the distribution of the adjusted score given in Proposition 2. Note that under our stated assumptions, we have

$$\sqrt{N}\,C_{\psi}^{\star}\big(\tilde{\theta}_{nT}\big) = \big(I,\ -\mathcal{H}_{\psi\phi\cdot\beta}\mathcal{H}_{\phi\cdot\beta}^{-1}\big)\sqrt{N}\,\mathbf{C}_{\psi\phi}\big(\tilde{\theta}_{nT}\big) + o_p(1). \tag{C.15}$$

Using (C.14) in (C.15) yields

$$\sqrt{N}\left[C_{\psi}\big(\tilde{\theta}_{nT}\big) - \mathcal{H}_{\psi\phi\cdot\beta}\mathcal{H}_{\phi\cdot\beta}^{-1}C_{\phi}\big(\tilde{\theta}_{nT}\big)\right] \xrightarrow{d} N\big(0, \mathcal{H}_{\psi\cdot\beta} - \mathcal{H}_{\psi\phi\cdot\beta}\mathcal{H}_{\phi\cdot\beta}^{-1}\mathcal{H}_{\phi\psi\cdot\beta}\big). \tag{C.16}$$

under H_0^{ψ} and H_A^{ϕ}. Then, this last result and Lemma 1 yield the fourth result of Proposition 2.

Using (C.13), (C.15) and Lemma 1, we can also determine the distribution of $\sqrt{N}\,C_{\psi}^{\star}\big(\tilde{\theta}_{nT}\big)$ under H_A^{ψ} and H_0^{ϕ} for the asymptotic power analysis. It can easily be discerned that

$$\sqrt{N}\,C_{\phi}^{\star}\big(\tilde{\theta}_{nT}\big) \xrightarrow{d} N\big(-\big(\mathcal{H}_{\psi\cdot\beta} - \mathcal{H}_{\psi\phi\cdot\beta}\mathcal{H}_{\phi\cdot\beta}^{-1}\mathcal{H}_{\psi\phi\cdot\beta}^{'}\big)\delta_{\psi},\ \mathcal{H}_{\psi\cdot\beta} - \mathcal{H}_{\psi\phi\cdot\beta}\mathcal{H}_{\phi\cdot\beta}^{-1}\mathcal{H}_{\psi\phi\cdot\beta}^{'}\big). \tag{C.17}$$

Therefore, $LM_{\psi}^{\star}\big(\tilde{\theta}_{nT}\big) \xrightarrow{d} \chi_{k_{\psi}}^{2}(\vartheta_4)$, where $\vartheta_4 = \delta_{\psi}^{'}\big(\mathcal{H}_{\psi\cdot\beta} - \mathcal{H}_{\psi\phi\cdot\beta}\mathcal{H}_{\phi\cdot\beta}^{-1}\mathcal{H}_{\psi\phi\cdot\beta}^{'}\big)\delta_{\psi}$. It follows that $\vartheta_2 - \vartheta_4 \geq 0$. This result indicates that $LM_{\psi}^{\star}\big(\tilde{\theta}_{nT}\big)$ has less asymptotic power than $LM_{\psi}\big(\tilde{\theta}_{nT}\big)$ when there is no local misspecification, i.e., when $H_0^{\phi}: \phi_0 = \phi_{\star}$ holds.

Finally, using (C.13), (C.15) and Lemma 1, we can also determine the distribution of $\sqrt{N}\,C_{\psi}^{\star}\big(\tilde{\theta}_{nT}\big)$ under H_A^{ψ} and H_A^{ϕ}. Lemma 1 and (C.13) implies that

$$\sqrt{N}\,\mathbf{C}_{\psi\phi}\big(\tilde{\theta}_{nT}\big) \xrightarrow{d} N\left(-\begin{pmatrix} \mathcal{H}_{\psi\cdot\beta} & \mathcal{H}_{\psi\phi\cdot\beta} \\ \mathcal{H}_{\phi\psi\cdot\beta} & \mathcal{H}_{\phi\cdot\beta} \end{pmatrix} \begin{pmatrix} \delta_{\psi} \\ \delta_{\phi} \end{pmatrix}, \begin{pmatrix} \mathcal{H}_{\psi\cdot\beta} & \mathcal{H}_{\psi\phi\cdot\beta} \\ \mathcal{H}_{\phi\psi\cdot\beta} & \mathcal{H}_{\phi\cdot\beta} \end{pmatrix} \right). \tag{C.18}$$

By using (C.18) in (C.15), we will get (C.17) under H_A^{ψ} and H_A^{ϕ}, namely

$$\sqrt{N}\,C_{\phi}^{\star}\big(\tilde{\theta}_{nT}\big) \xrightarrow{d} N\big(-\big(\mathcal{H}_{\psi\cdot\beta} - \mathcal{H}_{\psi\phi\cdot\beta}\mathcal{H}_{\phi\cdot\beta}^{-1}\mathcal{H}_{\psi\phi\cdot\beta}^{'}\big)\delta_{\psi},\ \mathcal{H}_{\psi\cdot\beta} - \mathcal{H}_{\psi\phi\cdot\beta}\mathcal{H}_{\phi\cdot\beta}^{-1}\mathcal{H}_{\psi\phi\cdot\beta}^{'}\big). \tag{C.19}$$

Hence, $LM_{\psi}^{\star}\big(\tilde{\theta}_{nT}\big) \xrightarrow{d} \chi_{k_{\psi}}^{2}(\vartheta_4)$, where $\vartheta_4 = \delta_{\psi}^{'}\big(\mathcal{H}_{\psi\cdot\beta} - \mathcal{H}_{\psi\phi\cdot\beta}\mathcal{H}_{\phi\cdot\beta}^{-1}\mathcal{H}_{\psi\phi\cdot\beta}^{'}\big)\delta_{\psi}$. This result is not surprising since the asymptotic distribution of $LM_{\psi}^{\star}\big(\tilde{\theta}_{nT}\big)$ does not depend on the presence of ϕ_0.

C.3 Proof of Corollaries

The results in Corollaries 1-3 directly follow from Proposition 2. Therefore, their proofs are omitted.

C.4 Simulation Results

Table C.4: Empirical sizes when H_0: The DPD model and $(n, T) = (100, 10)$

	Normal Distribution						*Gamma Distribution*					
γ_0	LM_ρ	$LM_\rho^\star$	LM_ρ^A	LM_λ	$LM_\lambda^\star$	LM_λ^A	LM_ρ	$LM_\rho^\star$	LM_ρ^A	LM_λ	$LM_\lambda^\star$	LM_λ^A
						Rook						
-0.30	0.046	0.015	0.050	0.042	0.005	0.051	0.047	0.016	0.048	0.042	0.008	0.048
-0.10	0.044	0.038	0.051	0.042	0.039	0.054	0.040	0.042	0.049	0.041	0.037	0.050
-0.05	0.040	0.049	0.053	0.048	0.051	0.057	0.043	0.045	0.049	0.047	0.046	0.054
0.05	0.061	0.046	0.048	0.061	0.056	0.051	0.057	0.051	0.051	0.056	0.052	0.051
0.10	0.074	0.042	0.050	0.064	0.039	0.044	0.070	0.041	0.052	0.061	0.043	0.048
0.30	0.135	0.028	0.050	0.100	0.024	0.046	0.128	0.035	0.053	0.099	0.028	0.051
						Queen						
-0.30	0.063	0.020	0.047	0.053	0.012	0.049	0.062	0.018	0.055	0.049	0.011	0.051
-0.10	0.044	0.047	0.056	0.046	0.043	0.057	0.039	0.038	0.053	0.044	0.038	0.048
-0.05	0.049	0.053	0.056	0.051	0.048	0.053	0.044	0.048	0.053	0.042	0.044	0.052
0.05	0.055	0.046	0.044	0.058	0.051	0.049	0.062	0.049	0.056	0.055	0.050	0.052
0.10	0.075	0.050	0.055	0.060	0.050	0.048	0.070	0.045	0.053	0.061	0.043	0.051
0.30	0.099	0.012	0.056	0.062	0.017	0.051	0.083	0.015	0.051	0.051	0.020	0.045

Table C.5: Empirical sizes when H_0: The SSPD model and $(n, T) = (100, 10)$

	Normal Distribution						*Gamma Distribution*					
λ_0	LM_ρ	$LM_\rho^\star$	LM_ρ^A	LM_γ	$LM_\gamma^\star$	LM_γ^A	LM_ρ	$LM_\rho^\star$	LM_ρ^A	LM_γ	$LM_\gamma^\star$	LM_γ^A
						Rook						
-0.30	1.000	0.791	0.056	1.000	0.999	0.050	1.000	0.793	0.054	1.000	1.000	0.051
-0.10	0.913	0.051	0.052	0.334	0.184	0.055	0.913	0.048	0.052	0.335	0.168	0.049
-0.05	0.394	0.053	0.053	0.087	0.071	0.054	0.379	0.052	0.047	0.085	0.065	0.054
0.05	0.326	0.048	0.049	0.077	0.068	0.057	0.336	0.051	0.052	0.073	0.064	0.056
0.10	0.853	0.051	0.046	0.204	0.136	0.050	0.863	0.053	0.054	0.215	0.143	0.050
0.30	1.000	0.730	0.052	0.998	0.997	0.049	1.000	0.708	0.051	0.999	0.998	0.049
						Queen						
-0.30	0.994	0.134	0.050	0.604	0.431	0.048	0.997	0.144	0.054	0.614	0.451	0.054
-0.10	0.393	0.058	0.055	0.070	0.068	0.056	0.374	0.055	0.047	0.072	0.064	0.047
-0.05	0.134	0.052	0.053	0.056	0.057	0.054	0.132	0.047	0.046	0.049	0.049	0.056
0.05	0.171	0.046	0.047	0.073	0.063	0.061	0.187	0.045	0.050	0.060	0.054	0.054
0.10	0.550	0.053	0.054	0.103	0.071	0.049	0.539	0.055	0.044	0.116	0.073	0.046
0.30	0.999	0.202	0.054	0.972	0.970	0.053	0.999	0.195	0.045	0.972	0.969	0.057

Table C.6: Empirical sizes when H_0: The SDPDW model and $(n,T) = (100,10)$: Rook Weight Matrix

		Normal Distribution			*Gamma Distribution*		
λ_0	γ_0	LM_ρ	$LM_\rho^\star$	LM_ρ^A	LM_ρ	$LM_\rho^\star$	LM_ρ^A
-0.30	-0.30	0.580	0.299	0.059	0.578	0.310	0.052
-0.30	-0.10	0.999	0.833	0.054	1.000	0.831	0.051
-0.30	-0.05	1.000	0.831	0.051	1.000	0.830	0.056
-0.30	0.05	1.000	0.772	0.050	1.000	0.778	0.048
-0.30	0.10	1.000	0.885	0.048	1.000	0.892	0.053
-0.30	0.30	1.000	1.000	0.050	1.000	1.000	0.053
-0.10	-0.30	0.120	0.045	0.049	0.118	0.043	0.050
-0.10	-0.10	0.466	0.039	0.049	0.466	0.039	0.049
-0.10	-0.05	0.734	0.046	0.049	0.751	0.048	0.052
-0.10	0.05	0.971	0.048	0.055	0.974	0.042	0.050
-0.10	0.10	0.990	0.047	0.054	0.987	0.049	0.048
-0.10	0.30	1.000	0.739	0.051	0.999	0.739	0.049
-0.05	-0.30	0.073	0.025	0.050	0.067	0.024	0.051
-0.05	-0.10	0.128	0.044	0.054	0.137	0.045	0.051
-0.05	-0.05	0.242	0.050	0.051	0.255	0.047	0.056
-0.05	0.05	0.515	0.051	0.053	0.502	0.048	0.052
-0.05	0.10	0.612	0.049	0.055	0.610	0.046	0.049
-0.05	0.30	0.819	0.219	0.052	0.835	0.215	0.051
0.05	-0.30	0.068	0.018	0.049	0.062	0.015	0.054
0.05	-0.10	0.121	0.046	0.052	0.115	0.036	0.047
0.05	-0.05	0.207	0.049	0.053	0.208	0.053	0.052
0.05	0.05	0.474	0.051	0.050	0.469	0.053	0.052
0.05	0.10	0.557	0.042	0.046	0.585	0.045	0.056
0.05	0.30	0.598	0.022	0.052	0.597	0.017	0.054
0.10	-0.30	0.133	0.031	0.050	0.134	0.035	0.052
0.10	-0.10	0.360	0.042	0.051	0.347	0.051	0.052
0.10	-0.05	0.639	0.053	0.049	0.639	0.056	0.059
0.10	0.05	0.956	0.054	0.050	0.957	0.055	0.044
0.10	0.10	0.985	0.045	0.048	0.985	0.046	0.046
0.10	0.30	0.990	0.151	0.050	0.991	0.157	0.046
0.30	-0.30	0.763	0.296	0.050	0.764	0.302	0.049
0.30	-0.10	0.976	0.746	0.050	0.970	0.768	0.052
0.30	-0.05	1.000	0.757	0.051	1.000	0.754	0.053
0.30	0.05	1.000	0.643	0.046	1.000	0.652	0.050
0.30	0.10	1.000	0.637	0.051	1.000	0.627	0.051
0.30	0.30	1.000	1.000	0.050	1.000	1.000	0.051

Table C.7: Empirical sizes when H_0: The SDPDW model and $(n,T) = (100, 10)$: Queen Weight Matrix

		Normal Distribution			*Gamma Distribution*		
λ_0	γ_0	LM_ρ	$LM_\rho^\star$	LM_ρ^A	LM_ρ	$LM_\rho^\star$	LM_ρ^A
-0.30	-0.30	0.223	0.021	0.048	0.227	0.017	0.046
-0.30	-0.10	0.670	0.125	0.055	0.662	0.118	0.050
-0.30	-0.05	0.935	0.153	0.054	0.934	0.153	0.054
-0.30	0.05	1.000	0.105	0.048	1.000	0.106	0.055
-0.30	0.10	1.000	0.067	0.049	1.000	0.065	0.048
-0.30	0.30	1.000	0.418	0.052	1.000	0.432	0.050
-0.10	-0.30	0.046	0.021	0.058	0.041	0.021	0.049
-0.10	-0.10	0.126	0.042	0.053	0.120	0.043	0.044
-0.10	-0.05	0.230	0.049	0.048	0.234	0.048	0.051
-0.10	0.05	0.541	0.045	0.048	0.533	0.050	0.054
-0.10	0.10	0.638	0.048	0.051	0.636	0.043	0.051
-0.10	0.30	0.675	0.021	0.053	0.670	0.019	0.052
-0.05	-0.30	0.043	0.020	0.050	0.045	0.020	0.051
-0.05	-0.10	0.058	0.039	0.053	0.062	0.042	0.045
-0.05	-0.05	0.092	0.048	0.057	0.094	0.047	0.052
-0.05	0.05	0.179	0.050	0.053	0.175	0.051	0.055
-0.05	0.10	0.221	0.053	0.048	0.210	0.049	0.049
-0.05	0.30	0.209	0.009	0.047	0.208	0.010	0.052
0.05	-0.30	0.121	0.024	0.049	0.117	0.021	0.050
0.05	-0.10	0.065	0.042	0.054	0.061	0.041	0.058
0.05	-0.05	0.105	0.045	0.053	0.114	0.043	0.054
0.05	0.05	0.264	0.050	0.052	0.274	0.050	0.048
0.05	0.10	0.344	0.049	0.050	0.364	0.042	0.058
0.05	0.30	0.477	0.049	0.050	0.484	0.047	0.046
0.10	-0.30	0.230	0.032	0.052	0.220	0.035	0.054
0.10	-0.10	0.157	0.044	0.050	0.153	0.042	0.050
0.10	-0.05	0.328	0.047	0.054	0.326	0.043	0.055
0.10	0.05	0.713	0.056	0.047	0.732	0.050	0.053
0.10	0.10	0.821	0.048	0.055	0.821	0.049	0.053
0.10	0.30	0.912	0.178	0.054	0.918	0.187	0.051
0.30	-0.30	0.866	0.028	0.051	0.858	0.030	0.053
0.30	-0.10	0.783	0.170	0.047	0.789	0.170	0.049
0.30	-0.05	0.977	0.194	0.048	0.974	0.204	0.054
0.30	0.05	1.000	0.231	0.051	1.000	0.241	0.058
0.30	0.10	1.000	0.350	0.057	1.000	0.351	0.047
0.30	0.30	1.000	1.000	0.048	1.000	1.000	0.052

Table C.8: Power of tests when H_1: The DPD/SSPD model and H_0: The 2WE model

γ_0/λ_0	LM_ρ	$LM_\rho^\star$	LM_ρ^A	LM_λ	$LM_\lambda^\star$	LM_λ^A	LM_γ	$LM_\gamma^\star$	LM_γ^A	LM_J	C_J
H_1: The DPD model											
-0.30	0.063	0.020	0.047	0.053	0.012	0.049	1.000	1.000	1.000	1.000	1.000
-0.10	0.044	0.047	0.056	0.046	0.043	0.057	0.533	0.522	0.528	0.364	0.362
-0.05	0.049	0.053	0.056	0.051	0.048	0.053	0.174	0.167	0.174	0.114	0.111
0.05	0.055	0.046	0.044	0.058	0.051	0.049	0.224	0.220	0.203	0.145	0.143
0.10	0.075	0.050	0.055	0.060	0.050	0.048	0.598	0.586	0.558	0.441	0.439
0.30	0.099	0.012	0.056	0.062	0.017	0.051	1.000	1.000	1.000	1.000	1.000
H_1: The SSPD model											
-0.30	0.994	0.134	0.050	1.000	1.000	0.966	0.604	0.431	0.048	1.000	1.000
-0.10	0.393	0.058	0.055	0.768	0.532	0.392	0.070	0.068	0.056	0.596	0.587
-0.05	0.134	0.052	0.053	0.246	0.165	0.137	0.056	0.057	0.054	0.165	0.160
0.05	0.171	0.046	0.047	0.323	0.203	0.203	0.073	0.063	0.061	0.230	0.225
0.10	0.550	0.053	0.054	0.840	0.564	0.584	0.103	0.071	0.049	0.711	0.706
0.30	0.999	0.202	0.054	1.000	0.999	1.000	0.972	0.970	0.053	1.000	1.000

Notes: The simulation results are based on the following design: (i) The queen weight matrix, (ii) the normally distributed errors, (iii) the nominal size of 0.05, and (iii) $(n, T) = (100, 10)$.

Table C.9: Power of tests when H_1 : The SDPDW model and H_0: The 2WE model

λ_0	γ_0	LM_ρ	$\mathrm{LM}^\star_\rho$	LM^A_ρ	LM_λ	$\mathrm{LM}^\star_\lambda$	LM^A_λ	LM_γ	$\mathrm{LM}^\star_\gamma$	LM^A_γ	LM_J	C_J
-0.30	-0.30	0.223	0.021	0.048	1.000	1.000	0.988	1.000	0.988	1.000	1.000	1.000
-0.30	-0.10	0.670	0.125	0.055	1.000	1.000	0.974	0.093	0.069	0.530	1.000	1.000
-0.30	-0.05	0.935	0.153	0.054	1.000	1.000	0.965	0.185	0.140	0.172	1.000	1.000
-0.30	0.05	1.000	0.105	0.048	1.000	0.998	0.968	0.919	0.794	0.209	1.000	1.000
-0.30	0.10	1.000	0.067	0.049	1.000	0.978	0.959	0.994	0.963	0.546	1.000	1.000
-0.30	0.30	1.000	0.418	0.052	1.000	0.999	0.969	1.000	1.000	1.000	1.000	1.000
-0.10	-0.30	0.046	0.021	0.058	0.598	0.428	0.445	1.000	1.000	1.000	1.000	1.000
-0.10	-0.10	0.126	0.042	0.053	0.709	0.622	0.412	0.458	0.429	0.536	0.760	0.778
-0.10	-0.05	0.230	0.049	0.048	0.731	0.598	0.399	0.136	0.127	0.175	0.605	0.623
-0.10	0.05	0.541	0.045	0.048	0.770	0.398	0.375	0.323	0.291	0.204	0.711	0.692
-0.10	0.10	0.638	0.048	0.051	0.795	0.256	0.378	0.710	0.682	0.583	0.867	0.851
-0.10	0.30	0.675	0.021	0.053	0.789	0.278	0.354	1.000	1.000	1.000	1.000	1.000
-0.05	-0.30	0.043	0.020	0.050	0.181	0.082	0.149	1.000	1.000	1.000	1.000	1.000
-0.05	-0.10	0.058	0.039	0.053	0.226	0.188	0.139	0.532	0.511	0.546	0.471	0.485
-0.05	-0.05	0.092	0.048	0.057	0.234	0.187	0.137	0.162	0.155	0.177	0.221	0.232
-0.05	0.05	0.179	0.050	0.053	0.272	0.128	0.136	0.236	0.229	0.215	0.297	0.287
-0.05	0.10	0.221	0.053	0.048	0.279	0.095	0.132	0.643	0.629	0.568	0.601	0.589
-0.05	0.30	0.209	0.009	0.047	0.242	0.065	0.137	1.000	1.000	1.000	1.000	1.000
0.05	-0.30	0.121	0.024	0.049	0.278	0.107	0.212	1.000	1.000	1.000	1.000	1.000
0.05	-0.10	0.065	0.042	0.054	0.278	0.221	0.199	0.514	0.501	0.540	0.513	0.522
0.05	-0.05	0.105	0.045	0.053	0.294	0.219	0.205	0.155	0.150	0.172	0.269	0.277
0.05	0.05	0.264	0.050	0.052	0.354	0.161	0.210	0.276	0.237	0.200	0.362	0.344
0.05	0.10	0.344	0.049	0.050	0.380	0.125	0.212	0.673	0.624	0.573	0.631	0.620
0.05	0.30	0.477	0.049	0.050	0.463	0.141	0.212	1.000	1.000	1.000	1.000	1.000
0.10	-0.30	0.230	0.032	0.052	0.725	0.505	0.612	1.000	1.000	1.000	1.000	1.000
0.10	-0.10	0.157	0.044	0.050	0.779	0.676	0.602	0.433	0.396	0.532	0.815	0.828
0.10	-0.05	0.328	0.047	0.054	0.812	0.650	0.587	0.119	0.112	0.183	0.715	0.723
0.10	0.05	0.713	0.056	0.047	0.865	0.453	0.587	0.439	0.340	0.190	0.808	0.795
0.10	0.10	0.821	0.048	0.055	0.888	0.330	0.600	0.825	0.742	0.573	0.925	0.919
0.10	0.30	0.912	0.178	0.054	0.938	0.524	0.621	1.000	1.000	1.000	1.000	1.000
0.30	-0.30	0.866	0.028	0.051	1.000	1.000	1.000	1.000	0.691	1.000	1.000	1.000
0.30	-0.10	0.783	0.170	0.047	1.000	1.000	1.000	0.339	0.517	0.543	1.000	1.000
0.30	-0.05	0.977	0.194	0.048	1.000	1.000	1.000	0.754	0.826	0.172	1.000	1.000
0.30	0.05	1.000	0.231	0.051	1.000	0.992	1.000	0.999	0.999	0.214	1.000	1.000
0.30	0.10	1.000	0.350	0.057	1.000	0.977	1.000	1.000	1.000	0.586	1.000	1.000
0.30	0.30	1.000	1.000	0.048	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Notes: The simulation results are based on the following design: (i) The queen weight matrix, (ii) the normally distributed errors, (iii) the nominal size of 0.05, and (iii) $(n, T) = (100, 10)$.

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

λ_0	γ_0	ρ_0	LM_ρ	$LM^\star_\rho$	LM^A_ρ	LM_λ	$LM^\star_\lambda$	LM^A_λ	LM_γ	$LM^\star_\gamma$	LM^A_γ	LM_J	C_J
0.05	-0.30	-0.30	1.000	0.862	0.976	0.797	0.239	0.205	1.000	1.000	1.000	1.000	1.000
0.05	-0.30	-0.10	0.688	0.088	0.263	0.398	0.017	0.208	1.000	1.000	1.000	1.000	1.000
0.05	-0.30	-0.05	0.366	0.022	0.107	0.318	0.045	0.221	1.000	1.000	1.000	1.000	1.000
0.05	-0.30	0.05	0.064	0.080	0.089	0.241	0.195	0.208	1.000	1.000	1.000	1.000	1.000
0.05	-0.30	0.10	0.213	0.226	0.216	0.226	0.324	0.213	1.000	1.000	1.000	1.000	1.000
0.05	-0.30	0.30	0.997	0.930	0.814	0.332	0.807	0.223	0.999	1.000	1.000	1.000	1.000
0.05	-0.10	-0.30	0.999	0.928	0.948	0.179	0.076	0.209	0.981	0.794	0.543	0.998	0.997
0.05	-0.10	-0.10	0.358	0.210	0.254	0.160	0.115	0.214	0.737	0.548	0.536	0.660	0.651
0.05	-0.10	-0.05	0.106	0.076	0.100	0.213	0.173	0.208	0.622	0.514	0.537	0.520	0.524
0.05	-0.10	0.05	0.227	0.085	0.079	0.394	0.262	0.206	0.427	0.494	0.536	0.599	0.611
0.05	-0.10	0.10	0.564	0.222	0.185	0.541	0.275	0.220	0.360	0.490	0.543	0.772	0.773
0.05	-0.10	0.30	1.000	0.832	0.716	0.975	0.188	0.210	0.204	0.589	0.554	1.000	0.999
0.05	-0.05	-0.30	0.997	0.927	0.941	0.143	0.044	0.202	0.861	0.413	0.177	0.987	0.985
0.05	-0.05	-0.10	0.247	0.225	0.245	0.127	0.154	0.206	0.361	0.182	0.170	0.365	0.358
0.05	-0.05	-0.05	0.069	0.086	0.105	0.186	0.184	0.203	0.242	0.174	0.162	0.252	0.251
0.05	-0.05	0.05	0.347	0.093	0.074	0.459	0.222	0.204	0.117	0.151	0.181	0.417	0.424
0.05	-0.05	0.10	0.689	0.205	0.170	0.643	0.193	0.213	0.099	0.153	0.173	0.653	0.652
0.05	-0.05	0.30	0.999	0.735	0.689	0.995	0.111	0.211	0.357	0.222	0.171	0.999	0.998
0.05	0.05	-0.30	0.972	0.919	0.930	0.377	0.061	0.200	0.191	0.132	0.211	0.932	0.931
0.05	0.05	-0.10	0.095	0.238	0.236	0.079	0.223	0.201	0.131	0.195	0.206	0.252	0.249
0.05	0.05	-0.05	0.089	0.096	0.096	0.160	0.190	0.193	0.194	0.222	0.208	0.240	0.232
0.05	0.05	0.05	0.581	0.074	0.076	0.607	0.118	0.198	0.383	0.257	0.195	0.562	0.548
0.05	0.05	0.10	0.833	0.146	0.167	0.825	0.086	0.204	0.530	0.262	0.198	0.783	0.771
0.05	0.05	0.30	1.000	0.264	0.642	1.000	0.234	0.209	0.978	0.367	0.196	1.000	1.000
0.05	0.10	-0.30	0.953	0.907	0.928	0.599	0.110	0.205	0.275	0.462	0.576	0.945	0.945
0.05	0.10	-0.10	0.072	0.222	0.234	0.063	0.213	0.209	0.486	0.587	0.565	0.470	0.469
0.05	0.10	-0.05	0.127	0.092	0.092	0.167	0.176	0.194	0.574	0.604	0.565	0.511	0.507
0.05	0.10	0.05	0.660	0.056	0.071	0.671	0.093	0.210	0.778	0.649	0.554	0.798	0.785
0.05	0.10	0.10	0.882	0.083	0.169	0.880	0.093	0.209	0.878	0.662	0.574	0.919	0.915
0.05	0.10	0.30	1.000	0.100	0.643	1.000	0.446	0.221	1.000	0.830	0.573	1.000	1.000
0.05	0.30	-0.30	0.985	0.187	0.905	0.961	0.030	0.211	1.000	1.000	1.000	1.000	1.000
0.05	0.30	-0.10	0.093	0.019	0.232	0.043	0.035	0.214	1.000	1.000	1.000	0.999	0.999
0.05	0.30	-0.05	0.130	0.022	0.097	0.125	0.059	0.207	1.000	1.000	1.000	1.000	1.000
0.05	0.30	0.05	0.860	0.109	0.078	0.837	0.288	0.219	1.000	1.000	1.000	1.000	1.000
0.05	0.30	0.10	0.990	0.215	0.144	0.983	0.475	0.236	1.000	1.000	1.000	1.000	1.000
0.05	0.30	0.30	1.000	0.528	0.618	1.000	0.866	0.241	1.000	1.000	1.000	1.000	1.000

Notes: The simulation results are based on the following design: (i) The queen weight matrix, (ii) the normally distributed errors, (iii) the nominal size of 0.05, and (iii) $(n, T) = (100, 10)$.

Table C.11: Power of tests when H_1: The SDPD model and H_0: The 2WE model

λ_0	γ_0	ρ_0	LM_ρ	$LM_\rho^\star$	LM_ρ^A	LM_λ	$LM_\lambda^\star$	LM_λ^A	LM_γ	$LM_\gamma^\star$	LM_γ^A	LM_J	C_J
0.10	-0.30	-0.30	1.000	0.778	0.980	0.974	0.048	0.630	1.000	1.000	1.000	1.000	1.000
0.10	-0.30	-0.10	0.833	0.049	0.286	0.823	0.193	0.612	1.000	1.000	1.000	1.000	1.000
0.10	-0.30	-0.05	0.549	0.015	0.114	0.771	0.335	0.611	1.000	1.000	1.000	1.000	1.000
0.10	-0.30	0.05	0.087	0.130	0.091	0.700	0.659	0.609	1.000	1.000	1.000	1.000	1.000
0.10	-0.30	0.10	0.177	0.325	0.204	0.679	0.771	0.621	1.000	1.000	1.000	1.000	1.000
0.10	-0.30	0.30	0.998	0.961	0.816	0.808	0.978	0.620	0.988	1.000	1.000	1.000	1.000
0.10	-0.10	-0.30	0.998	0.926	0.956	0.556	0.071	0.604	0.995	0.762	0.556	0.999	0.999
0.10	-0.10	-0.10	0.303	0.190	0.261	0.629	0.525	0.603	0.789	0.464	0.544	0.828	0.834
0.10	-0.10	-0.05	0.101	0.074	0.103	0.708	0.627	0.584	0.617	0.415	0.542	0.793	0.806
0.10	-0.10	0.05	0.459	0.099	0.081	0.864	0.704	0.587	0.285	0.376	0.531	0.883	0.892
0.10	-0.10	0.10	0.812	0.231	0.178	0.930	0.694	0.595	0.193	0.383	0.548	0.959	0.960
0.10	-0.10	0.30	1.000	0.813	0.705	0.999	0.415	0.620	0.366	0.393	0.542	1.000	1.000
0.10	-0.05	-0.30	0.994	0.937	0.951	0.365	0.143	0.594	0.954	0.382	0.180	0.991	0.990
0.10	-0.05	-0.10	0.161	0.226	0.259	0.563	0.572	0.603	0.395	0.132	0.169	0.626	0.627
0.10	-0.05	-0.05	0.102	0.084	0.102	0.696	0.625	0.595	0.221	0.115	0.175	0.620	0.629
0.10	-0.05	0.05	0.686	0.090	0.087	0.906	0.625	0.589	0.082	0.094	0.168	0.838	0.841
0.10	-0.05	0.10	0.916	0.203	0.180	0.961	0.548	0.593	0.127	0.095	0.172	0.944	0.944
0.10	-0.05	0.30	1.000	0.637	0.680	1.000	0.315	0.610	0.759	0.134	0.171	1.000	1.000
0.10	0.05	-0.30	0.913	0.936	0.943	0.112	0.333	0.605	0.319	0.120	0.212	0.889	0.889
0.10	0.05	-0.10	0.127	0.243	0.240	0.493	0.601	0.605	0.133	0.251	0.201	0.531	0.510
0.10	0.05	-0.05	0.377	0.101	0.098	0.685	0.545	0.594	0.263	0.297	0.194	0.651	0.626
0.10	0.05	0.05	0.918	0.068	0.076	0.959	0.341	0.594	0.653	0.386	0.204	0.931	0.925
0.10	0.05	0.10	0.986	0.130	0.168	0.992	0.252	0.598	0.808	0.433	0.204	0.984	0.983
0.10	0.05	0.30	1.000	0.126	0.672	1.000	0.596	0.626	0.999	0.691	0.202	1.000	1.000
0.10	0.10	-0.30	0.807	0.918	0.931	0.123	0.439	0.587	0.157	0.438	0.574	0.861	0.867
0.10	0.10	-0.10	0.213	0.224	0.232	0.444	0.533	0.585	0.483	0.637	0.561	0.697	0.683
0.10	0.10	-0.05	0.521	0.098	0.106	0.693	0.442	0.579	0.669	0.693	0.575	0.823	0.810
0.10	0.10	0.05	0.958	0.049	0.079	0.974	0.269	0.599	0.919	0.776	0.564	0.976	0.971
0.10	0.10	0.10	0.995	0.062	0.155	0.997	0.269	0.606	0.973	0.811	0.565	0.996	0.996
0.10	0.10	0.30	1.000	0.186	0.640	1.000	0.827	0.614	1.000	0.961	0.576	1.000	1.000
0.10	0.30	-0.30	0.828	0.294	0.916	0.604	0.121	0.602	1.000	1.000	1.000	1.000	1.000
0.10	0.30	-0.10	0.199	0.042	0.238	0.273	0.187	0.604	1.000	1.000	1.000	1.000	1.000
0.10	0.30	-0.05	0.577	0.086	0.094	0.676	0.331	0.616	1.000	1.000	1.000	1.000	1.000
0.10	0.30	0.05	0.996	0.369	0.074	0.996	0.741	0.635	1.000	1.000	1.000	1.000	1.000
0.10	0.30	0.10	1.000	0.564	0.139	1.000	0.869	0.627	1.000	1.000	1.000	1.000	1.000
0.10	0.30	0.30	1.000	0.855	0.635	1.000	0.987	0.672	1.000	1.000	1.000	1.000	1.000

Notes: The simulation results are based on the following design: (i) The queen weight matrix, (ii) the normally distributed errors, (iii) the nominal size of 0.05, and (iii) $(n, T) = (100, 10)$.

Table C.12: Power of tests when H_1: The SDPD model and H_0: The 2WE model

λ_0	γ_0	ρ_0	LM_ρ	$LM^\star_\rho$	LM^A_ρ	LM_λ	$LM^\star_\lambda$	LM^A_λ	LM_γ	$LM^\star_\gamma$	LM^A_γ	LM_J	C_J
0.30	-0.30	-0.30	1.000	0.626	0.990	1.000	0.874	1.000	1.000	0.936	1.000	1.000	1.000
0.30	-0.30	-0.10	0.997	0.103	0.332	1.000	1.000	1.000	1.000	0.774	1.000	1.000	1.000
0.30	-0.30	-0.05	0.977	0.045	0.125	1.000	1.000	1.000	1.000	0.730	1.000	1.000	1.000
0.30	-0.30	0.05	0.579	0.086	0.088	1.000	1.000	1.000	0.987	0.683	1.000	1.000	1.000
0.30	-0.30	0.10	0.284	0.238	0.238	1.000	1.000	1.000	0.895	0.661	1.000	1.000	1.000
0.30	-0.30	0.30	0.999	0.949	0.844	1.000	1.000	1.000	0.434	0.565	1.000	1.000	1.000
0.30	-0.10	-0.30	1.000	0.983	0.973	1.000	1.000	1.000	1.000	0.151	0.563	1.000	1.000
0.30	-0.10	-0.10	0.429	0.542	0.295	1.000	1.000	1.000	0.705	0.337	0.538	1.000	1.000
0.30	-0.10	-0.05	0.446	0.334	0.118	1.000	1.000	1.000	0.345	0.430	0.535	1.000	1.000
0.30	-0.10	0.05	0.970	0.090	0.083	1.000	1.000	1.000	0.714	0.616	0.537	1.000	1.000
0.30	-0.10	0.10	0.999	0.088	0.185	1.000	1.000	1.000	0.944	0.706	0.536	1.000	1.000
0.30	-0.10	0.30	1.000	0.235	0.734	1.000	0.936	1.000	1.000	0.926	0.533	1.000	1.000
0.30	-0.05	-0.30	0.994	0.989	0.965	1.000	1.000	1.000	0.999	0.179	0.178	1.000	1.000
0.30	-0.05	-0.10	0.468	0.587	0.294	1.000	1.000	1.000	0.337	0.651	0.175	1.000	1.000
0.30	-0.05	-0.05	0.792	0.367	0.114	1.000	1.000	1.000	0.380	0.744	0.173	1.000	1.000
0.30	-0.05	0.05	0.999	0.099	0.077	1.000	0.999	1.000	0.963	0.881	0.176	1.000	1.000
0.30	-0.05	0.10	1.000	0.095	0.188	1.000	0.997	1.000	0.998	0.930	0.177	1.000	1.000
0.30	-0.05	0.30	1.000	0.438	0.712	1.000	0.954	1.000	1.000	0.994	0.170	1.000	1.000
0.30	0.05	-0.30	0.668	0.987	0.961	1.000	1.000	1.000	0.871	0.696	0.201	1.000	1.000
0.30	0.05	-0.10	0.980	0.535	0.281	1.000	1.000	1.000	0.767	0.980	0.208	1.000	1.000
0.30	0.05	-0.05	0.999	0.343	0.108	1.000	0.998	1.000	0.975	0.991	0.205	1.000	1.000
0.30	0.05	0.05	1.000	0.226	0.083	1.000	0.977	1.000	1.000	0.999	0.191	1.000	1.000
0.30	0.05	0.10	1.000	0.358	0.173	1.000	0.956	1.000	1.000	1.000	0.199	1.000	1.000
0.30	0.05	0.30	1.000	0.994	0.674	1.000	1.000	1.000	1.000	1.000	0.202	1.000	1.000
0.30	0.10	-0.30	0.342	0.979	0.953	1.000	1.000	1.000	0.492	0.906	0.574	1.000	1.000
0.30	0.10	-0.10	0.999	0.477	0.270	1.000	0.999	1.000	0.975	0.998	0.574	1.000	1.000
0.30	0.10	-0.05	1.000	0.337	0.109	1.000	0.992	1.000	0.999	1.000	0.573	1.000	1.000
0.30	0.10	0.05	1.000	0.519	0.079	1.000	0.968	1.000	1.000	1.000	0.563	1.000	1.000
0.30	0.10	0.10	1.000	0.770	0.164	1.000	0.986	1.000	1.000	1.000	0.567	1.000	1.000
0.30	0.10	0.30	1.000	1.000	0.680	1.000	1.000	1.000	1.000	1.000	0.583	1.000	1.000
0.30	0.30	-0.30	0.848	0.608	0.945	0.986	0.993	1.000	0.996	1.000	1.000	1.000	1.000
0.30	0.30	-0.10	1.000	0.947	0.261	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.30	0.30	-0.05	1.000	0.995	0.116	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.30	0.30	0.05	1.000	1.000	0.063	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	0.135	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	0.750	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Notes: The simulation results are based on the following design: (i) The queen weight matrix, (ii) the normally distributed errors, (iii) the nominal size of 0.05, and (iii) $(n, T) = (100, 10)$.

References

Acemoglu, Daron (2008). *Introduction to Modern Economic Growth*. NJ: Princeton: Princeton University Press.

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 (1988). *Spatial econometrics: Methods and Models*. New York: Springer.

Anselin, Luc (2001). "Rao's score test in spatial econometrics". In: *Journal of Statistical Planning and Inference* 97.1, pp. 113 –139.

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 Dong Li (2001). "LM Tests for Functional Form and Spatial Error Correlation". In: *International Regional Science Review* 24.2, pp. 194–225.

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, pp. 103–134.

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 Bilias (2001). "Rao's score, Neyman's C(α) 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, pp. 649–658.

Bera, Anil K., Gabriel Montes-Rojas, and Walter Sosa-Escudero (2010). "General Specification Testing with Locally Misspecified Models". In: *Econometric Theory* 26 (06), pp. 1838–1845.

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.

Burridge, P. (1980). "On the Cliff-Ord test for spatial correlation". In: *Journal of the Royal Statistical Society, Series B: Methodological* 42, pp. 107–108.

Burridge, Peter and Bernard Fingleton (2010). "Bootstrap Inference in Spatial Econometrics: the J-test". In: *Spatial Economic Analysis* 5.1, pp. 93–119.

Caselli, Francesco, Gerardo Esquivel, and Fernando Lefort (1996). "Reopening the convergence debate: A new look at cross-country growth empirics". In: *Journal of Economic Growth* 1.3, pp. 363–389.

Cliff, Andrew and Keith Ord (1972). "Testing for Spatial Autocorrelation Among Regression Residuals". In: *Geographical Analysis* 4.3, pp. 267–284.

Davidson, Russell and James G. MacKinnon (1987). "Implicit Alternatives and the Local Power of Test Statistics". English. In: *Econometrica* 55.6, pp. 1305–1329.

— (1998). "Graphical Methods for Investigating the Size and Power of Hypothesis Tests". In: *The Manchester School* 66.1, pp. 1–26.

Debarsy, Nicolas and Cem Ertur (2010). "Testing for spatial autocorrelation in a fixed effects panel data model". In: *Regional Science and Urban Economics* 40.6, pp. 453 –470.

Elhorst, J. Paul (2005). “Unconditional Maximum Likelihood Estimation of Linear and Log-Linear Dynamic Models for Spatial Panels”. In: *Geographical Analysis* 37.1, pp. 85–106.
— (2010a). “Dynamic panels with endogenous interaction effects when T is small”. In: *Regional Science and Urban Economics* 40.5. Advances In Spatial Econometrics, pp. 272 –282.
— (2010b). “Spatial Panel Data Models”. In: *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications*. Ed. by Manfred M. Fischer and Arthur Getis. Berlin, Heidelberg: Springer Berlin Heidelberg.
— (2014). *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*. SpringerBriefs in Regional Science. Springer Berlin Heidelberg.
Ertur, Cem and Wilfried Koch (2007). “Growth, Technological Interdependence and Spatial Externalities: Theory and Evidence”. In: *Journal of Applied Econometrics* 22.6, pp. 1033–1062.
Fingleton, Bernard and Julie Le Gallo (2008). “Estimating spatial models with endogenous variables, a spatial lag and spatially dependent disturbances: Finite sample properties”. In: *Papers in Regional Science* 87.3, pp. 319–339.
Hall, Alastair R. (2004). *Generalized Method of Moments*. Advanced Texts in Econometrics. Oxford University Press.
Han, Xiaoyi, Chih-Sheng Hsieh, and Lung-fei Lee (2016). “Estimation and Model Selection of Higher-order Spatial Autoregressive Model: An Efficient Bayesian Approach”. In: *Regional Science and Urban Economics*, pp. –. DOI: http://dx.doi.org/10.1016/j.regsciurbeco.2016.12.003.
Hsiao, Cgeng. (2014). *Analysis of Panel Data*. Analysis of Panel Data. Cambridge University Press.
Islam, Nazrul (1995). “Growth Empirics: A Panel Data Approach”. In: *The Quarterly Journal of Economics* 110.4, pp. 1127–1170.
Jin, Fei and Lung-fei Lee (2015). “On the bootstrap for MoranâĂŹs test for spatial dependence”. In: *Journal of Econometrics* 184.2, pp. 295 –314.
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 (2001). “On the Asymptotic Distribution of the Moran I Test Statistic with Applications”. In: *Journal of Econometrics* 104.2, pp. 219–257.
— (2010). “Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances”. In: *Journal of Econometrics* 157, pp. 53–67.
Kelejian, Harry H. and Dennis P. Robinson (2004). “The Influence of Spatially Correlated Heteroskedasticity on Tests for Spatial Correlation”. In: *Advances in Spatial Econometrics: Methodology, Tools and Applications*. Ed. by Luc Anselin, Raymond J. G. M. Florax, and Sergio J. Rey. Berlin, Heidelberg: Springer Berlin Heidelberg.
Knight, Malcolm, Norman Loayza, and Delano Villanueva (1993). “Testing the Neoclassical Theory of Economic Growth: A Panel Data Approach”. In: *Staff Papers - International Monetary Fund* 40.3, pp. 512–541.
Korniotis, George M. (2010). “Estimating Panel Models With Internal and External Habit Formation”. In: *Journal of Business & Economic Statistics* 28.1, pp. 145–158.
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, pp. 255–271.
— (2011). “A Unified Transformation Approach for the Estimation of Spatial Dynamic Panel Data Models: Stability, Spatial Cointegration and Explosive Roots”. In: *Handbook of Empirical Economics 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 & Francis Group.
Lee, Lung-fei and Jihai Yu (2012a). "Convergence: A Spatial Dynamic Panel Data Approach". In: *Global Journal of Economics* 01.01, p. 36.
— (2012b). "QML Estimation of Spatial Dynamic Panel Data Models with Time Varying Spatial Weights Matrices". In: *Spatial Economic Analysis* 7.1, pp. 31–74.
— (2012c). "The C(α)-type gradient test for spatial dependence in spatial autoregressive models". In: *Letters in Spatial and Resource Sciences* 5.3, pp. 119–135.
— (2014). "Efficient GMM estimation of spatial dynamic panel data models with fixed effects". In: *Journal of Econometrics* 180.2, pp. 174 –197.
— (2015a). "Estimation of fixed effects panel regression models with separable and non-separable space-time filters". In: *Journal of Econometrics* 184.1, pp. 174 –192.
— (2015b). "Spatial Panel Data Models". In: *The Oxford Handbook of Panel Data*. Ed. by Badi H. Baltagi. Advanced Studies in Theoretical and Applied Econometrics. Oxford University Press.
— (2016). "Identification of Spatial Durbin Panel Models". In: *Journal of Applied Econometrics* 31.1, pp. 133–162.
LeSage, James and Robert K. Pace (2009). *Introduction to Spatial Econometrics (Statistics: A Series of Textbooks and Monographs*. London: Chapman and Hall/CRC.
Lesage, James P. and Manfred M. Fischer (2008). "Spatial Growth Regressions: Model Specification, Estimation and Interpretation". In: *Spatial Economic Analysis* 3.3, pp. 275–304.
Mankiw, N Gregory, David Romer, and David N Weil (1992). "A Contribution to the Empirics of Economic Growth". In: *The Quarterly Journal of Economics* 107.2, pp. 407–37.
Moran, P. A. P. (1950). "A Test for the Serial Independence of Residuals". In: *Biometrika* 37.1/2, pp. 178–181.
Newey, Whitney K. and Kenneth D. West (1987). "Hypothesis Testing with Efficient Method of Moments Estimation". In: *International Economic Review* 28.3, pp. 777–787.
Parent, Olivier and P. James LeSage (2010). "A spatial dynamic panel model with random effects applied to commuting times". In: *Transportation Research Part B: Methodological* 44.5, pp. 633 –645.
— (2011). "A space-time filter for panel data models containing random effects". In: *Computational Statistics & Data Analysis* 55.1, pp. 475 –490.
— (2012). "Spatial dynamic panel data models with random effects". In: *Regional Science and Urban Economics* 42.4, pp. 727 –738.
Robinson, Peter M. and Francesca Rossi (2014). "Improved Lagrange multiplier tests in spatial autoregressions". In: *The Econometrics Journal* 17.1, pp. 139–164.
— (2015a). "Refined Tests for Spatial Correlation". In: *Econometric Theory* 31.6, pp. 1249–1280.
— (2015b). "Refinements in maximum likelihood inference on spatial autocorrelation in panel data". In: *Journal of Econometrics* 189.2. Frontiers in Time Series and Financial Econometrics, pp. 447 –456.
Robinson, P.M. (2008). "Correlation testing in time series, spatial and cross-sectional data". In: *Journal of Econometrics* 147.1, pp. 5 –16.
Saikkonen, Pentti (1989). "Asymptotic relative efficiency of the classical test statistics under misspecification". In: *Journal of Econometrics* 42.3, pp. 351 –369.
Su, Liangjun and Zhenlin Yang (2015). "QML estimation of dynamic panel data models with spatial errors". In: *Journal of Econometrics* 185.1, pp. 230 –258.
Taspinar, Suleyman, Osman Dogan, and Wim P. M Vijverberg (2016). "GMM inference in spatial autoregressive models". In: *Econometric Reviews* 0.0, pp. 1–24.

White, Halbert (2001). *Asymptotic Theory for Econometricians*. Economic theory, econometrics, and mathematical economics. Academic Press.

Yang, Zhenlin (2010). “A robust LM test for spatial error components”. In: *Regional Science and Urban Economics* 40.5. Advances In Spatial Econometrics, pp. 299 –310.

— (2015a). “LM tests of spatial dependence based on bootstrap critical values”. In: *Journal of Econometrics* 185.1, pp. 33 –59.

— (2015b). *Unified M-Estimation of Fixed-Effects Spatial Dynamic Models with Short 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”. In: *Econometric Theory* 26 (05), pp. 1332–1362.

Yu, Jihai, Robert de Jong, and Lung fei Lee (2008). “Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large”. In: *Journal of Econometrics* 146.1, pp. 118 –134.

— (2012). “Estimation for spatial dynamic panel data with fixed effects: The case of spatial cointegration”. In: *Journal of Econometrics* 167.1, pp. 16 –37.