Simple Tests for Social Interaction Models with Network Structures

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

17 August 2017

Simple Tests for Social Interaction Models with Network Structures

Osman Doğan[a,*], Süleyman Taşpınar[b], Anil K. Bera[c]

[a] *University of Illinois at Urbana-Champaign, Illinois, United States.*
[b] *Economics Program, Queens College, The City University of New York, New York, United States.*
[c] *University of Illinois at Urbana-Champaign, Illinois, United States.*

Abstract

We consider an extended spatial autoregressive model that can incorporate possible endogenous interactions, exogenous interactions, unobserved group fixed effects and correlation of unobservables. In the generalized method of moments (GMM) and the maximum likelihood (ML) frameworks, we introduce simple gradient based tests that can be used to test the presence of endogenous effects, the correlation of unobservables and the contextual effects. We show the asymptotic distributions of tests, and formulate robust tests that have central chi-square distributions under both the null and local misspecification. The proposed tests are easy to compute and only require the estimates from a transformed linear regression model. We carry out an extensive Monte Carlo study to investigate the size and power properties of the proposed tests. Our results show that the proposed tests have good finite sample properties and are useful for testing the presence of endogenous effects, correlation of unobservables and contextual effects in a social interaction model.

Keywords: Social interactions, Endogenous effects, Spatial dependence, GMM inference, LM tests, Robust LM test, Local misspecification.

1. Introduction

In a social interaction model, an individual's outcome is affected by the outcomes and characteristics of her reference group's members, i.e., her peers. The effects channeled through the outcomes of the reference group is known as the endogenous effects. The effects arising from the characteristics of the group is called the contextual effects. Identification of these effects within an estimation framework is important because their policy implications greatly differ. Manski (1993) shows that endogenous and contextual effects cannot be separately identified in a linear-in-means model. This identification problem, known as the "reflection problem," has led to various adjustments to the linear-in-means specification to allow for partial or full identification of these effects (Brock and Durlauf, 2001; Lee, 2007; Calvo-Armengol et al., 2009; Bramoullé et al., 2009; Lin, 2010; Liu and Lee, 2010; Goldsmith-Pinkham and Imbens, 2013; Hsieh and Lee, 2014; Burridge et al., 2016).

Tools from spatial econometrics can be useful to reformulate social interaction models thereby identification of various effects become possible (for spatial econometrics, see Anselin (1988), LeSage and Pace (2009), Elhorst (2010, 2014)). The group relation can be represented by means of a so-called spatial weights (or connectivity) matrix. The outcomes of a group members are included into a model through a so-called spatial lag operator which constructs a new variable consisting of a weighted average of the group members' outcomes. Similarly, the contextual effect variables are formulated through a spatial lag of the group members' characteristics. This class of models is referred to as the social interaction models with network structures. Lee (2007), Lee et al. (2010) and Liu et al. (2014) consider this type of social interaction models in which

*Corresponding author
Email address: odogan@illinois.edu (Osman Doğan)

We are grateful to the editor-in-chief and two anonymous referees for many pertinent comments and constructive suggestions. 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.

the endogenous effects, the contextual effects and the correlation of unobservables are formulated through the spatial lag operators.

In the literature, diagnostic testing for social interaction models with network structures have received scant attention. The gradient or score based tests within the GMM or ML frameworks can be formulated for testing the presence of various effects by following White (1982), Newey (1985a,b,c), Tauchen (1985), Newey and West (1987) and Smith (1987). However, these gradient based tests, i.e., the Lagrange multiplier (LM) tests, are not robust to the local parametric misspecification in the alternative models. Within the ML framework, Davidson and MacKinnon (1987), Saikkonen (1989) and Bera and Yoon (1993) show that the conventional LM test statistic has a non-central chi-square distribution when the alternative hypothesis deviates (locally) from the true data generating process (DGP). Bera et al. (2010) extend this result to a GMM framework and show that the asymptotic distribution of the LM test is a non-central chi-square distribution when the alternative model deviates locally from the true DGP. Thus, the conventional LM tests will over reject the true null hypothesis and lead to incorrect inference under parametric misspecification. Bera and Yoon (1993) and Bera et al. (2010) formulate robust (or adjusted) versions that have, asymptotically, central chi-square distributions irrespective of the local deviation of the alternative model from the true data generating process.

In this paper, we formulate robust LM tests in the GMM and ML frameworks for a social interaction model that has a network structure. We show the asymptotic distributions of these tests under the null and the local alternatives within the context of our social interaction model. These tests can be used to detect the presence of endogenous effects, the correlation of unobservables and the contextual effects. Besides being robust to local parametric misspecification in the alternative models, these tests are computationally very simple and only require estimates from a transformed linear regression model. We design an extensive Monte Carlo study to investigate the size and power properties of our proposed tests. Our results show that the proposed tests have good finite sample properties and can be useful for the identification of the source of dependence in a social interaction model.

The rest of this paper is organized as follows. In Section 2, we introduce the social interaction model. In Section 3, we review the GMM estimation approach and introduce the GMM gradient tests for testing linear and nonlinear restrictions on the spatial autoregressive parameters. We adjust these procedures for our social interaction model and formulate the robust LM test statistics. In Section 4, we consider the ML estimation approach for the model, and formulate various versions of the LM tests. In Section 5, we introduce test statistics for testing the presence of contextual effects in both GMM and ML frameworks. In Section 6, we show the relationships among the test statistics. In Sections 7, 8 and 9, we compare the size and power properties of tests through a Monte Carlo study. Section 10 closes the paper with concluding remarks. Some technical details are relegated to appendices.

2. The Model Specification

We consider a group interaction set up that consists of R groups. Let m_r be the number of individuals in the rth group, and $n = \sum_{r=1}^{R} m_r$ be the total number of individuals. Let $Y_r = (Y_{1r}, Y_{2r}, \ldots, Y_{m_r r})'$ be the $m_r \times 1$ vector of observed outcomes in the rth group. Then, the DGP stated for the rth group is given by

$$Y_r = \lambda_0 W_r Y_r + X_{1r}\beta_{01} + W_r X_{2r}\beta_{02} + l_{m_r}\alpha_{0r} + u_r, \tag{2.1}$$

$$u_r = \rho_0 M_r u_r + \varepsilon_r \quad \text{for} \quad r = 1, \ldots, R. \tag{2.2}$$

In (2.1) and (2.2), the network weights matrices W_r and M_r are $m_r \times m_r$ matrices with known constant terms and zero diagonal elements. The matrices of exogenous variables are denoted with X_{1r} and X_{2r}, which have dimensions of $m_r \times k_1$ and $m_r \times k_2$, respectively.[2] The matching parameters for the exogenous variables are denoted by β_{01} and β_{02}. The endogenous social interaction effects in (2.1) is captured by $W_r Y_r$ with the scalar coefficient λ_0. The contextual effects are captured by $W_r X_{2r}$ with the matching parameter vector of β_{02}. The model differs from the cross-sectional spatial econometric models by including the unobserved group fixed effect, denoted by $l_{m_r}\alpha_{0r}$, where l_{m_r} is an $m_r \times 1$ vector of ones and α_{0r} represents the unobserved group fixed effect. The regression disturbance term $u_r = (u_{1r}, \ldots, u_{m_r r})'$ and the innovation term $\varepsilon_r = (\varepsilon_{1r}, \ldots, \varepsilon_{m_r r})'$

[2] Note that X_{1r} and X_{2r} may or may not be the same.

are m_r-dimensional vectors. The distributional assumption is imposed on the elements of ε_r by assuming that ε_{ir}s are i.i.d with mean zero and variance σ_0^2. Finally, through the spatial autoregressive process given in (2.2), the unobserved correlation effects within the rth group is captured by $M_r u_r$ with the scalar coefficient ρ_0. In the spatial econometric literature, λ_0 and ρ_0 are called the spatial autoregressive parameters.

The network structure specified through weight matrices W_r and M_r has implications for the estimation approaches adopted for the model. In Lee (2007), $W_r = \frac{1}{m_r-1}\left(l_{m_r} l_{m_r}^{'} - I_{m_r}\right)$ is the $m_r \times m_r$ network matrix, which indicates that each individual in the group is equally affected by the other members of the group. Hence, the spatial lag term $W_r Y_r$ denotes the average outcome of the group r. The zero diagonal property of W_r indicates that Y_{ir} is not included in the calculation of the group mean outcome for the ith individual, which is not the case in Manski (1993). The network matrices considered in Lee et al. (2010) may differ from above W_r, but their rows still sum to a constant. In the case where this property is violated, the likelihood function of the model can not be derived, and therefore Liu and Lee (2010) propose 2SLS and GMM methods for estimation.

In certain interaction scenarios, the elements of weight matrices might be a function of sample size n. For cross-sectional spatial autoregressive models without group fixed effects, Lee (2004) assumes a large group interaction setting and specifies the elements of weight matrix by $w_{ij} = O(1/h_n)$, where w_{ij} is the (i, j)th element of weight matrix W and $\{h_n\}$ is a sequence of real numbers that can be bounded or divergent with the property that $\lim_{n\to\infty} h_n/n = 0$. For the case where $W_r = \frac{1}{m_r-1}\left(l_{m_r} l_{m_r}^{'} - I_{m_r}\right)$, we have $h_n = m_r - 1$ and $h_n/n = (m_r - 1)/n$, where $n = \sum_{r=1}^{R} m_r$. If there is no variation in group sizes and the increase in n is generated by the increase in m_r and R, then clearly $\lim_{n\to\infty} h_n/n = 0$. However, as shown in Lee (2007), the endogenous effect cannot be identified in this case. In addition, Lee (2007) shows that both the endogenous and exogenous interaction effects would be weakly identified and their rates of convergence can be quite low when all group sizes are large, even if there is group size variation. Therefore, following Lee et al. (2010) and Liu and Lee (2010), we assume interaction scenarios in which $\{h_n\}$ is bounded in this study.

In order to write the model for the entire sample, define $Y = (Y_1^{'}, \ldots, Y_R^{'})^{'}$, $X = (X_1^{'}, \ldots, X_R^{'})^{'}$ with $X_r = (X_{1r}, W_r X_{2r})$, $u = (u_1^{'}, \ldots, u_R^{'})^{'}$, $\alpha_0 = (\alpha_{01}, \ldots, \alpha_{0R})^{'}$, and $\varepsilon = (\varepsilon_1^{'}, \ldots, \varepsilon_R^{'})^{'}$. Let $\mathrm{D}\left(\{C_r\}_{r=1}^{R}\right)$ be the operator that creates a block diagonal matrix in which the diagonal blocks are m_r by n_r matrices C_r. Let $W = \mathrm{D}\,(W_1, \ldots, W_R)$, $M = \mathrm{D}\,(M_1, \ldots, M_R)$ and $l_n = \mathrm{D}\,(l_{m_1}, \ldots, l_{m_R})$. Then, the model for the entire sample is given by

$$Y = \lambda_0 WY + X\beta_0 + l_n\alpha_0 + u, \quad u = \rho_0 Mu + \varepsilon, \tag{2.3}$$

where $\beta_0 = (\beta_{01}^{'}, \beta_{02}^{'})^{'}$. To obtain the reduced form of (2.3), define $R(\rho) = (I_n - \rho M)$ and $S(\lambda) = (I_n - \lambda W)$. At the true parameter values, let $R(\rho_0) = R$ and $S(\lambda_0) = S$. Then, if R and S are not singular, the reduced form of the model becomes

$$Y = S^{-1}X\beta_0 + S^{-1}l_n\alpha_0 + S^{-1}R^{-1}\varepsilon. \tag{2.4}$$

3. The GMM Estimation Approach

The model can be stated in terms of innovations in the following way

$$RY = RZ\delta_0 + Rl_n\alpha_0 + \varepsilon, \tag{3.1}$$

where $Z = (WY, X)$ and $\delta_0 = (\lambda_0, \beta_0^{'})^{'}$. To wipe out fixed effects from (3.1), an orthogonal projector that projects a vector to the column space of Rl_n can be used. For this purpose, the rth diagonal block of Rl_n, which is given by $R_r l_{m_r} = A \times (1, \rho_0)^{'}$ where $A = (l_{m_r}, M_r l_{m_r})$, can be used to construct a projector. Define $J_r = I_{m_r} - A(A^{'}A)^{-}A^{'}$, where A^{-} is the generalized inverse of A. In the case where M_r has rows all sum to a constant c such that $R_r l_{m_r} = (1 - c\rho_0) l_{m_r}$, the projector reduces to the usual deviation from group mean maker $J_r = I_{m_r} - \frac{1}{m_r} l_{m_r} l_{m_r}^{'}$. In any case, since $J_r R_r l_{m_r} = 0$, the fixed effects can be eliminated from (3.1). Let $J = \mathrm{D}\,(J_1, \ldots, J_R)$. Then, the pre-multiplication of (3.1) by J yields

$$JRY = JRZ\delta_0 + J\varepsilon. \tag{3.2}$$

The GMM estimation approach requires the following assumptions.

Assumption 1. *The innovation term ε_{ir}s are i.i.d with zero mean and variance σ_0^2, and* $\mathrm{E}\left(|\varepsilon_{ir}|^{4+\tau}\right) < \infty$ *for some $\tau > 0$, for all $i = 1, \ldots, m_r$ and $r = 1, \ldots, R$.*

Assumption 2. *(i) The matrix X has full column rank of $k = k_1 + k_2$, and it has uniformly bounded elements, and $\lim_{n\to\infty} \frac{1}{n} X^{'}X$ is a finite nonsingular matrix, (ii) $\mathcal{X}(\rho) = \lim_{n\to\infty} \frac{1}{n} f^{'}(\rho) f(\rho)$, where $f(\rho) = JR(\rho)\mathrm{E}(Z)$, exist and is non-singular for all values of ρ such that $R(\rho)$ is non-singular.*

Assumption 3. *The row and column sums of matrices W, M, S^{-1}, and R^{-1} are bounded uniformly in absolute value.*[3]

Assumption 4. *The parameter vector $\theta_0 = (\rho_0, \delta_0^{'})^{'}$ is in the interior of bounded parameter space Θ.*

3.1. The Moment Conditions

The internal instrumental variables (IVs) for the endogenous variable JRZ can be determined from the reduced form of the model in (2.4). By definition, the best set of instruments is $f = JR\mathrm{E}(Z) = (JRGX\beta_0 + JRGl_n\alpha_0,\, JRX)$, where $G = WS^{-1}$. Since $R = I_n - \rho_0 M$, the best IV set is a linear combination of IVs in $Q_\infty = J\left(Q^0,\, MQ^0\right)$, where $Q^0 = (GX,\, Gl_n,\, X)$. Furthermore, since $G = \sum_{j=0}^{\infty} \lambda^j W^{j+1}$, Q^0 is a linear combination of elements of $Q_\infty^0 = \left(WX, W^2X, \ldots, Wl_n, W^2l_n, \ldots,\, X\right)$. Since l_n has R columns, the number of IVs increases as the number of groups increases. Let Q_K^0 be a sub-matrix of Q_∞^0 and define $Q_K = J\left(Q_K^0,\, MQ_K^0\right)$ as the $n \times K$ IV matrix, where $K \geq k+1$. Then, the linear moment function is defined by $g_1(\delta_0) = Q_K^{'} J\varepsilon$, which satisfies the orthogonality condition under Assumption 1:

$$\mathrm{E}\big(g_1(\delta_0)\big) = \mathrm{E}\big(Q_K^{'} J\varepsilon\big) = Q_K^{'}\mathrm{E}\big(\varepsilon\big) = 0_{K\times 1}, \tag{3.3}$$

where $J\varepsilon(\theta_0) = JR(Y - Z\delta_0)$. The result in (2.4) indicates that the endogenous term JRZ is also a function of a stochastic term. Liu and Lee (2010) formulate additional quadratic moment functions to exploit the information in the stochastic part. Both types of moment functions can be used in the GMM framework to estimate all parameters jointly. Let $U_1, \ldots, U_q$ be $n \times n$ non-stochastic matrices satisfying $\mathrm{tr}(JU_j) = 0$ for $j = 1, \ldots, q$.[4] Using these non-stochastic matrices, additional quadratic moment functions can be formulated as $\mathrm{E}\big(\varepsilon^{'}(\theta_0)JU_jJ\varepsilon(\theta_0)\big)$ for $j = 1, \ldots, q$, where $\varepsilon(\theta_0) = JR\big(Y - Z\delta_0\big)$. Let $g_2(\theta) = \big(\varepsilon^{'}(\theta)JU_1J\varepsilon(\theta), \ldots, \varepsilon^{'}(\theta)JU_qJ\varepsilon(\theta)\big)^{'}$ be the set of quadratic moment functions. The combined set of moment functions for the GMM estimation is then given by

$$g(\theta) = \Big[g_1^{'}(\theta),\, g_2^{'}(\theta)\Big]^{'}, \tag{3.4}$$

where $\theta = (\rho, \delta^{'})^{'}$. The population moment condition for each quadratic moment function in (3.4) is satisfied since $\mathrm{E}\left(\varepsilon^{'}(\theta_0)JU_jJ\varepsilon(\theta_0)\right) = \sigma_0^2 \mathrm{tr}\left(JU_jJ\right) = 0$ for all j by assumption.[5]

For the notational simplicity, let $T_j = JU_jJ$ for $j = 1, \ldots, q$, $H = MR^{-1}$, $\bar{G} = RGR^{-1}$ and $A^s = A + A^{'}$ for any square matrix A. Also, let $\mathrm{vec}(\cdot)$ be the operator that creates a column vector from the elements of an input matrix, $\mathrm{vec}_D(\cdot)$ be the operator that creates a column vector from the diagonal elements of an input matrix, and e_i be the ith unit column vector of dimension $k+1$. Define $\Omega = \mathrm{E}\big[g(\theta_0)g^{'}(\theta_0)\big]$ and $D_2 = \mathrm{E}\big[\frac{\partial g_2(\theta)}{\partial \theta^{'}}\big|_{\theta_0}\big]$. For our generic set of moment functions in (3.4), these matrices are given by

$$\Omega = \begin{bmatrix} \sigma_0^2 Q_K^{'} Q_K & \mu_3 Q_K^{'}\omega \\ \mu_3 \omega^{'} Q_K & (\mu_4 - 3\sigma_0^4)\omega^{'}\omega + \sigma_0^4 \Upsilon \end{bmatrix}, \tag{3.5}$$

[3] For properties of matrices that have row and column sums bounded uniformly in absolute value, see Kelejian and Prucha (2010).

[4] The row and column sums of these matrices are assumed to be uniformly bounded in absolute value. That is, *Assumption 3* holds for these matrices.

[5] The conditions for the identification of parameters can be investigated from moment functions. The identification requires that $\mathrm{E}\left(g(\theta)\right) = 0$ if and only if $\theta = \theta_0$ (Newey and McFadden, 1994, Lemma 2.3). Liu and Lee (2010) state the identification conditions . Here, we simply assume that θ_0 is identified.

$$
\mathrm{D}_2 = -\sigma_0^2 \begin{bmatrix} \mathrm{tr}(T_1^s H) & \mathrm{tr}(T_1^s \bar{G}) & 0_{1\times k} \\ \mathrm{tr}(T_2^s H) & \mathrm{tr}(T_2^s \bar{G}) & 0_{1\times k} \\ \vdots & \vdots & \vdots \\ \mathrm{tr}(T_q^s H) & \mathrm{tr}(T_q^s \bar{G}) & 0_{1\times k} \end{bmatrix}, \tag{3.6}
$$

where μ_3 and μ_4 are, respectively, the third and the fourth moments of ε_{ir}, $\omega = [\mathrm{vec}_D(T_1), \ldots, \mathrm{vec}_D(T_q)]$ and $\Upsilon = \frac{1}{2}\left[\mathrm{vec}(T_1^s), \ldots, \mathrm{vec}(T_q^s)\right]^{'}\left[\mathrm{vec}(T_1^s), \ldots, \mathrm{vec}(T_q^s)\right]$.

The optimal GMM estimation requires an initial estimate of Ω. The result in (3.5) indicates that a consistent estimate of Ω can be recovered from consistent estimates of σ_0^2, μ_3 and μ_4 under the stated assumptions. Let $\widehat{\Omega}$ be an initial consistent estimate of Ω. Then, the optimal GMM estimator (GMME) is defined by

$$
\hat{\theta} = \arg\min_{\theta\in\Theta} g^{'}(\theta)\widehat{\Omega}^{-1}g(\theta), \tag{3.7}
$$

The GMME defined in (3.7) is consistent but may not be centered properly around the true parameter vector. The asymptotic bias arises since the dimension of $g_1(\theta)$ increases as the number of groups increases, i.e., there is too many IV problem for the GMM estimation. Under the condition that $K^{3/2}/n \to 0$, Liu and Lee (2010) establish the following fundamental result:

$$
\sqrt{n}\left(\hat{\theta} - \theta_0 - Bias\right) \xrightarrow{d} N\left[0_{(k+2)\times 1}, \mathcal{H}^{-1}\right], \tag{3.8}
$$

where $\mathcal{H} = \sigma_0^{-2}\,\mathrm{D}\left(0, \mathcal{X}\left(\rho_0\right)\right) + \lim_{n\to\infty}\frac{1}{n}\bar{D}_2^{'}V_{22}\bar{D}_2$, $V_{22} = \left[\left(\mu_4 - 3\sigma_0^4\right)\omega^{'}\omega + \sigma_0^4\Upsilon - \frac{\mu_3^2}{\sigma_0^2}\omega^{'}P_K\omega\right]^{-1}$, $Bias = \left[\sigma_0^{-2}\,\mathrm{D}\left(0, Z^{'}R^{'}P_K RZ\right) + \check{D}_2^{'}V_{22}\check{D}_2\right]^{-1}\left[\mathrm{tr}\left(P_K M R^{-1}\right), \mathrm{tr}\left(P_K\bar{G}\right)e_1^{'}\right]^{'}$, $\check{D}_2 = D_2 - \frac{\mu_3}{\sigma_0^2}\left[0, \omega^{'}P_K RZ\right]$, $\bar{D}_2 = D_2 - \frac{\mu_3}{\sigma_0^2}\left[0, \omega^{'}f\right]$ and $P_K = Q_K(Q_K^{'}Q_K)^{-}Q_K^{'}$.[6]

3.2. The GMM Gradients Tests for Spatial Autoregressive Parameters

In this section, we formulate the GMM gradient tests when the number of linear IVs is fixed, i.e., when K is fixed. The standard LM test statistic requires computation of the restricted model implied by the null hypotheses. Consider the set of restrictions given by $\pi(\theta_0) = 0$, where $\pi : \Theta \to \mathbb{R}^p$ is a continuously differentiable function such that its Jacobian $\partial\pi(\theta_0)/\partial\theta^{'}$ is finite and has full row rank p. Then, the restricted GMME is defined by $\hat{\theta}_r = \arg\min_{\{\theta:\pi(\theta)=0\}} g^{'}(\theta)\widehat{\Omega}^{-1}g(\theta)$. The restricted estimator can also be defined in an alternative way by using the implicit function theorem to state the set of restrictions in an explicit way. By the implicit function theorem, there exists a continuously differentiable function $\kappa : \mathbf{R}^{k+2-p} \to \mathbf{R}^{k+2}$ such that $\partial\kappa\left(\varrho\right)/\varrho^{'}$ has full row rank $k+2-p$, where ϱ is the vector of free parameters. Define $\hat{\varrho} = \arg\min_{\varrho} g^{'}\left(\kappa\left(\varrho\right)\right)\widehat{\Omega}^{-1}g\left(\kappa\left(\varrho\right)\right)$. Then, the restricted GMME is, alternatively, defined by $\hat{\theta}_r = \kappa(\hat{\varrho})$. Let $G_a\left(\theta\right) = \frac{\partial g(\theta)}{\partial a^{'}}$ and $C_a(\theta) = \frac{1}{n}G_a^{'}(\theta)\hat{\Omega}^{-1}g(\theta)$ where $a = \rho, \lambda, \beta$. Define $G\left(\theta\right) = \left[G_\rho\left(\theta\right), G_\lambda\left(\theta\right), G_\beta\left(\theta\right)\right]$, $C(\theta) = [C_\rho(\theta), C_\lambda(\theta), C_\beta(\theta)]$ and $B(\theta) = \frac{1}{n}G^{'}(\theta)\hat{\Omega}^{-1}G(\theta)$.[7] The standard gradient test, i.e. the LM test, is based on the idea that the sample gradients evaluated at $\hat{\theta}_r$ should be close to zero when the restrictions are valid. The test statistic is given by

$$
\mathrm{LM}_0^{\mathrm{g}}(\hat{\theta}_r) = n\,C^{'}(\hat{\theta}_r)\left[B(\hat{\theta}_r)\right]^{-1}C(\hat{\theta}_r). \tag{3.9}
$$

In the literature, the asymptotic properties of the LM test are investigated under local parametric misspecification in the alternative model (Davidson and MacKinnon, 1987; Saikkonen, 1989; Bera and Yoon, 1993; Bera and Bilias, 2001; Bera et al., 2010). Bera and Yoon (1993) and Bera et al. (2010) suggest robust LM tests when there is a local parametric misspecification in the alternative model that used to construct the test statistics. We consider similar robust LM tests for the following null hypothesis:

[6] The bias term is $O\left(\frac{K}{n}\right)$, and the result in (3.8) indicates that it will vanish only when $\frac{K^2}{n} \to 0$.

[7] The test statistics suggested in this section are formulated with $G\left(\theta\right)$ and $B(\theta)$. In Appendix B, we give explicit expressions for these terms.

1. On the correlations of error terms:

$$\mathrm{H}_0^{\rho} : \rho_0 = \rho_{\star}. \tag{3.10}$$

2. On the endogenous effects:

$$\mathrm{H}_0^{\lambda} : \lambda = \lambda_{\star}. \tag{3.11}$$

In (3.10) and (3.11), $\rho_{\star}$ and $\lambda_{\star}$ are hypothesized known quantities. For these hypotheses, we construct LM tests that are robust to local parametric misspecification. For this purpose, we consider the sequence of local alternatives formulated for hypotheses in 3.10 and 3.11. The sequence of local alternatives, also known as Pitman drifts, takes the following forms: $\mathrm{H}_A^{\lambda} : \lambda_0 = \lambda_{\star} + \delta_{\lambda}/\sqrt{n}$, and $\mathrm{H}_A^{\rho} : \rho_0 = \rho_{\star} + \delta_{\rho}/\sqrt{n}$, where δ_{λ} and δ_{ρ} are bounded scalars. As will be illustrated, this device of sequence of local alternatives is not only the basis of the ensuing discussion of power properties of test statistics, it is also instrumental in the formulation of our robust test statistics. Let $\mathcal{H} = \sigma_0^{-2}\,\mathrm{D}\left(0,\, \mathcal{X}\left(\rho_0\right)\right) + \lim_{n\to\infty} \frac{1}{n}\bar{D}_2^{'}V_{22}\bar{D}_2$. To formulate the test statistic, consider the following partition of $B\left(\theta\right)$ and $\mathcal{H}$:

$$B\left(\theta\right) = \begin{bmatrix} \underbrace{B_{\rho\rho}\left(\theta\right)}_{1\times 1} & \underbrace{B_{\rho\lambda}\left(\theta\right)}_{1\times 1} & \underbrace{B_{\rho\beta}\left(\theta\right)}_{1\times k} \\ \underbrace{B_{\lambda\rho}\left(\theta\right)}_{1\times 1} & \underbrace{B_{\lambda\lambda}\left(\theta\right)}_{1\times 1} & \underbrace{B_{\lambda\beta}\left(\theta\right)}_{1\times k} \\ \underbrace{B_{\beta\rho}\left(\theta\right)}_{k\times 1} & \underbrace{B_{\beta\lambda}\left(\theta\right)}_{k\times 1} & \underbrace{B_{\beta\beta}\left(\theta\right)}_{k\times k} \end{bmatrix}, \quad \mathcal{H} = \begin{bmatrix} \underbrace{\mathcal{H}_{\rho\rho}}_{1\times 1} & \underbrace{\mathcal{H}_{\rho\lambda}}_{1\times 1} & \underbrace{\mathcal{H}_{\rho\beta}}_{1\times k} \\ \underbrace{\mathcal{H}_{\lambda\rho}}_{1\times 1} & \underbrace{\mathcal{H}_{\lambda\lambda}}_{1\times 1} & \underbrace{\mathcal{H}_{\lambda\beta}}_{1\times k} \\ \underbrace{\mathcal{H}_{\beta\rho}}_{k\times 1} & \underbrace{\mathcal{H}_{\beta\lambda}}_{k\times 1} & \underbrace{\mathcal{H}_{\beta\beta}}_{k\times k} \end{bmatrix}. \tag{3.12}$$

Let $\tilde{\theta} = \left(\rho_{\star}, \lambda_{\star}, \tilde{\beta}^{'}\right)^{'}$ be a restricted GMME under the joint null hypothesis $\mathrm{H}_0 : \rho_0 = \rho_{\star}$ and $\lambda_0 = \lambda_{\star}$. The LM test statistic for this joint null hypothesis can be expressed as

$$\mathrm{LM}^{\mathrm{g}}_{\rho\lambda}(\tilde{\theta}) = n\,\mathbf{C}^{'}_{\rho\lambda}(\tilde{\theta})\left[\mathbf{B}_{1\cdot 3}(\tilde{\theta})\right]^{-1}\mathbf{C}_{\rho\lambda}(\tilde{\theta}), \tag{3.13}$$

where $\mathbf{C}_{\rho\lambda}(\tilde{\theta}) = \left[C^{'}_{\rho}(\tilde{\theta}),\, C^{'}_{\lambda}(\tilde{\theta})\right]^{'}$, $\mathbf{B}_{1\cdot 3}(\tilde{\theta}) = \mathbf{B}_{11}(\tilde{\theta}) - \mathbf{B}_{13}(\tilde{\theta})B^{-1}_{\beta\beta}(\tilde{\theta})\mathbf{B}_{31}(\tilde{\theta})$, $\mathbf{B}_{11}(\tilde{\theta}) = \begin{bmatrix} B_{\rho\rho}(\tilde{\theta}) & B_{\rho\lambda}(\tilde{\theta}) \\ B_{\lambda\rho}(\tilde{\theta}) & B_{\lambda\lambda}(\tilde{\theta}) \end{bmatrix}$,
112 and $\mathbf{B}_{31}(\tilde{\theta}) = \mathbf{B}^{'}_{13}(\tilde{\theta}) = \left[B_{\beta\rho}(\tilde{\theta}),\, B_{\beta\lambda}(\tilde{\theta})\right]$.

Now, we consider the problem of testing H_0^{ρ} when H_0^{λ} holds. Then, the standard LM test can be stated as

$$\mathrm{LM}^{\mathrm{g}}_{\rho}(\tilde{\theta}) = n\,C^{'}_{\rho}(\tilde{\theta})\left[B_{\rho\cdot\beta}(\tilde{\theta})\right]^{-1}C_{\rho}(\tilde{\theta}), \tag{3.14}$$

where $B_{\rho\cdot\beta}(\tilde{\theta}) = B_{\rho\rho}(\tilde{\theta}) - B_{\rho\beta}(\tilde{\theta})B^{-1}_{\beta\beta}(\tilde{\theta})B_{\beta\rho}(\tilde{\theta})$. The distribution of (3.14) under H_A^{ρ} and H_A^{λ} can be investigated from the first order Taylor expansion of pseudo-gradients $C_{\rho}(\tilde{\theta})$ and $C_{\beta}(\tilde{\theta})$ around θ_0. These expansions can be stated as

$$\begin{aligned} \sqrt{n}\,C_{\rho}(\tilde{\theta}) = \sqrt{n}\,C_{\rho}(\theta_0) &- \frac{1}{n}G^{'}_{\rho}(\theta_0)\widehat{\Omega}^{-1}G_{\rho}(\bar{\theta})\delta_{\rho} - \frac{1}{n}G^{'}_{\rho}(\theta_0)\widehat{\Omega}^{-1}G_{\lambda}(\bar{\theta})\delta_{\lambda} \\ &+ \frac{1}{n}\,G^{'}_{\rho}(\theta_0)\widehat{\Omega}^{-1}G_{\beta}(\bar{\theta})\sqrt{n}(\tilde{\beta} - \beta_0) + o_p(1), \end{aligned} \tag{3.15}$$

$$\begin{aligned} \sqrt{n}\,C_{\beta}(\tilde{\theta}) = \sqrt{n}\,C_{\beta}(\theta_0) &- \frac{1}{n}G^{'}_{\beta}(\theta_0)\widehat{\Omega}^{-1}G_{\rho}(\bar{\theta})\delta_{\rho} - \frac{1}{n}G^{'}_{\beta}(\theta_0)\widehat{\Omega}^{-1}G_{\lambda}(\bar{\theta})\delta_{\lambda} \\ &+ \frac{1}{n}\,G^{'}_{\beta}(\theta_0)\widehat{\Omega}^{-1}G_{\beta}(\bar{\theta})\sqrt{n}(\tilde{\beta} - \beta_0) + o_p(1), \end{aligned} \tag{3.16}$$

where $\bar{\theta}$ lies between $\tilde{\theta}$ and θ_0. Using the asymptotic results in Lemma 1, we obtain the following result from (3.15) and (3.16).

$$\begin{aligned} \sqrt{n}\,C_{\rho}(\tilde{\theta}) = \left[-1,\, \mathcal{H}_{\rho\beta}\mathcal{H}^{-1}_{\beta\beta}\right] \times \begin{bmatrix} -\sqrt{n}C_{\rho}(\theta_0) \\ -\sqrt{n}C_{\beta}(\theta_0) \end{bmatrix} &- \left[\mathcal{H}_{\rho\rho} - \mathcal{H}_{\rho\beta}\mathcal{H}^{-1}_{\beta\beta}\mathcal{H}_{\beta\rho}\right]\delta_{\rho} \\ &- \left[\mathcal{H}_{\rho\lambda} - \mathcal{H}_{\rho\beta}\mathcal{H}^{-1}_{\beta\beta}\mathcal{H}_{\beta\lambda}\right]\delta_{\lambda} + o_p(1). \end{aligned} \tag{3.17}$$

Under our stated assumptions, the pseudo-gradients have an asymptotic normal distribution as shown in Lemma 1. Thus, the result in (3.17) implies that $\sqrt{n}\, C_\rho(\tilde{\theta}) \xrightarrow{d} N\left[-\mathcal{H}_{\rho\cdot\beta}\delta_\rho - \mathcal{H}_{\rho\lambda\cdot\beta}\delta_\lambda, \mathcal{H}_{\rho\cdot\beta}\right]$, where $\mathcal{H}_{\rho\cdot\beta} = \left[\mathcal{H}_{\rho\rho} - \mathcal{H}_{\rho\beta}\mathcal{H}_{\beta\beta}^{-1}\mathcal{H}_{\beta\rho}\right]$, and $\mathcal{H}_{\rho\lambda\cdot\beta} = \left[\mathcal{H}_{\rho\lambda} - \mathcal{H}_{\rho\beta}\mathcal{H}_{\beta\beta}^{-1}\mathcal{H}_{\beta\lambda}\right]$.[8] Hence, $\mathrm{LM}_\rho^{\mathrm{g}}(\tilde{\theta}) \xrightarrow{d} \chi_1^2(\vartheta_1)$ under H_A^ρ and H_A^λ, where $\vartheta_1 = \delta_\rho^2\mathcal{H}_{\rho\cdot\beta} + \delta_\rho^{'}\mathcal{H}_{\rho\lambda\cdot\beta}\delta_\lambda + \delta_\lambda^{'}\mathcal{H}_{\rho\lambda\cdot\beta}^{'}\delta_\rho + \delta_\lambda^2\mathcal{H}_{\rho\lambda\cdot\beta}^{'}\mathcal{H}_{\rho\cdot\beta}^{-1}\mathcal{H}_{\rho\lambda\cdot\beta}$ is the non-centrality parameter.[9]

In the case where H_A^ρ and H_0^λ hold, the result in (3.17) implies that $\sqrt{n}\, C_\rho(\tilde{\theta}) \xrightarrow{d} N\left[-\mathcal{H}_{\rho\cdot\beta}\delta_\rho, \mathcal{H}_{\rho\cdot\beta}\right]$. Hence, $\mathrm{LM}_\rho^{\mathrm{g}}(\tilde{\theta}) \xrightarrow{d} \chi_1^2(\vartheta_2)$ under H_A^ρ and H_0^λ, where $\vartheta_2 = \delta_\rho^2\mathcal{H}_{\rho\cdot\beta}$. Therefore , under H_0^ρ and H_0^λ, $\mathrm{LM}_1^{\mathrm{g}}(\tilde{\theta})$ has a central chi-squared distribution and hence asymptotically correct size. In case where H_0^ρ and H_A^λ hold, the result in (3.17) indicates that $\sqrt{n}\, C_\rho(\tilde{\theta}) \xrightarrow{d} N\left[-\mathcal{H}_{\rho\lambda\cdot\beta}\delta_\lambda, \mathcal{H}_{\rho\cdot\beta}\right]$. Hence, $\mathrm{LM}_\rho^{\mathrm{g}}(\tilde{\theta}) \xrightarrow{d} \chi_1^2(\vartheta_3)$ under H_0^ρ and H_A^λ, where $\vartheta_3 = \delta_\lambda^2\mathcal{H}_{\rho\lambda\cdot\beta}^{'}\mathcal{H}_{\rho\cdot\beta}^{-1}\mathcal{H}_{\rho\lambda\cdot\beta}$. This result is simply the extension of Bera et al. (2010) to our GMM framework. It indicates that $\mathrm{LM}_1^{\mathrm{g}}(\tilde{\theta})$ will over reject $\mathrm{H}_0^\rho : \rho_0 = \rho_\star$ when there is local parametric misspecification in the alternative model.

Bera et al. (2010) suggest a robust version in a general context such that the test statistic has a central chi-square distribution irrespective of whether H_0^λ or H_A^λ holds. Using this approach, we can adjust the asymptotic mean and variance of $\sqrt{n}\, C_\rho(\tilde{\theta})$ in such a way that the resulting score statistic $\mathrm{LM}_\rho^{\mathrm{g}}(\tilde{\theta})$ has an asymptotic centered chi-square distribution. Let $\sqrt{n}\left[C_\rho(\tilde{\theta}) - \mathcal{H}_{\rho\lambda\cdot\beta}\mathcal{H}_{\lambda\cdot\beta}^{-1}C_\lambda(\tilde{\theta})\right]$ be the adjusted unfeasible pseudo-gradient, which has a zero asymptotic mean. Under our assumptions, a feasible version of the adjusted pseudo-gradient is given by $\sqrt{n}\, C_\rho^\star(\tilde{\theta}) = \sqrt{n}\left[C_\rho(\tilde{\theta}) - B_{\rho\lambda\cdot\beta}(\tilde{\theta})B_{\lambda\cdot\beta}^{-1}(\tilde{\theta})C_\lambda(\tilde{\theta})\right]$, where $B_{\lambda\cdot\beta}(\tilde{\theta}) = \left[B_{\lambda\lambda}(\tilde{\theta}) - B_{\lambda\beta}(\tilde{\theta})B_{\beta\beta}^{-1}(\tilde{\theta})B_{\beta\lambda}(\tilde{\theta})\right]$, and $B_{\rho\lambda\cdot\beta}(\tilde{\theta}) = \left[B_{\rho\lambda}(\tilde{\theta}) - B_{\rho\beta}(\tilde{\theta})B_{\beta\beta}^{-1}(\tilde{\theta})B_{\beta\lambda}(\tilde{\theta})\right]$. Then, we can use this adjusted pseudo-gradient to formulate a robust test statistics, denoted by $\mathrm{LM}_\rho^{\mathrm{g}\star}(\tilde{\theta})$. In the following proposition, we provide this test along with the results summarized so far.

Proposition 1. — Under Assumptions 1–4, the following results hold.

1. Under H_A^ρ and H_A^λ, we have

$$\mathrm{LM}_\rho^{\mathrm{g}}(\tilde{\theta}) \xrightarrow{d} \chi_1^2(\vartheta_1), \tag{3.18}$$

   where $\vartheta_1 = \delta_\rho^2\mathcal{H}_{\rho\cdot\beta} + \delta_\rho\mathcal{H}_{\rho\lambda\cdot\beta}\delta_\lambda + \delta_\lambda\mathcal{H}_{\rho\lambda\cdot\beta}^{'}\delta_\rho + \delta_\lambda^2\mathcal{H}_{\rho\lambda\cdot\beta}^{'}\mathcal{H}_{\rho\cdot\beta}^{-1}\mathcal{H}_{\rho\lambda\cdot\beta}$.

2. Under H_0^ρ and irrespective of whether H_0^λ or H_A^λ holds, we have

$$\mathrm{LM}_\rho^{\mathrm{g}\star}(\tilde{\theta}) = n\, C_\rho^{\star'}(\tilde{\theta})\left[B_{\rho\cdot\beta}(\tilde{\theta}) - B_{\rho\lambda\cdot\beta}(\tilde{\theta})B_{\lambda\cdot\beta}^{-1}(\tilde{\theta})B_{\rho\lambda\cdot\beta}^{'}(\tilde{\theta})\right]^{-1} C_\rho^\star(\tilde{\theta}) \xrightarrow{d} \chi_1^2, \tag{3.19}$$

   where $B_{\rho\cdot\beta}(\tilde{\theta}) = \left[B_{\rho\rho}(\tilde{\theta}) - B_{\rho\beta}(\tilde{\theta})B_{\beta\beta}^{-1}(\tilde{\theta})B_{\beta\rho}(\tilde{\theta})\right]$.

3. Under H_A^ρ and irrespective of whether H_0^λ or H_A^λ holds, we have

$$\mathrm{LM}_\rho^{\mathrm{g}\star}(\tilde{\theta}) \xrightarrow{d} \chi_1^2(\vartheta_4), \tag{3.20}$$

   where $\vartheta_4 = \delta_\rho^2\big(\mathcal{H}_{\rho\cdot\beta} - \mathcal{H}_{\rho\lambda\cdot\beta}\mathcal{H}_{\lambda\cdot\beta}^{-1}\mathcal{H}_{\rho\lambda\cdot\beta}^{'}\big)$.

Proof. See Appendix D. □

The noncentrality parameters reported in Proposition 1 can be used for asymptotic local power comparisons. Note that the tail probability of a noncentral chi-squared distribution decreases with the degrees of freedom and increases with the noncentrality parameter. Also, the noncentrality parameter is related to the

[8]Note that the distribution of $\sqrt{n}\, C_\rho(\tilde{\theta})$ has an asymptotic mean of $-\left[\mathcal{H}_{\rho\cdot\beta}\delta_\rho + \mathcal{H}_{\rho\lambda\cdot\beta}\delta_\lambda\right]$. The negative sign arises since we define the objective function differently. In Bera et al. (2010), the objective function is defined as $\mathcal{Q} = -g^{'}(\theta)\widehat{\Omega}^{-1}g(\theta)$ and $\hat{\theta} = \arg\max_{\theta\in\Theta}\mathcal{Q}$.

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

approximate slope of a test. If the asymptotic distribution of a test has a relatively larger noncentrality parameter, then the test has a relatively larger approximate slope (Newey, 1985a). Under H_A^{ρ} and H_0^{λ}, we have $\mathrm{LM}_{\rho}^{\mathrm{g}\star}(\tilde{\theta}) \xrightarrow{d} \chi_1^2(\vartheta_4)$ and $\mathrm{LM}_{\rho}^{\mathrm{g}}(\tilde{\theta}) \xrightarrow{d} \chi_1^2(\vartheta_2)$ from Proposition 1. It follows that $\vartheta_2 - \vartheta_4 \geq 0$, which indicates that $\mathrm{LM}_{\rho}^{\mathrm{g}\star}(\tilde{\theta})$ has less asymptotic power than $\mathrm{LM}_{\rho}^{\mathrm{g}}(\tilde{\theta})$ when there is no local parametric misspecification, i.e., when $\lambda_0 = 0$.

The results in Proposition 1 can also be replicated for the hypothesis in 3.11. For this purpose, we consider the null hypothesis $\mathrm{H}_0^{\lambda} : \lambda_0 = \lambda_\star$ when $\mathrm{H}_0^{\rho} : \rho_0 = \rho_\star$ holds. Then, the LM test can be formulated as

$$\mathrm{LM}_{\lambda}^{\mathrm{g}}(\tilde{\theta}) = n\, C_{\lambda}^{'}(\tilde{\theta}) \left[B_{\lambda\cdot\beta}(\tilde{\theta})\right]^{-1} C_{\lambda}(\tilde{\theta}), \tag{3.21}$$

where $B_{\lambda\cdot\beta}(\tilde{\theta}) = B_{\lambda\lambda}(\tilde{\theta}) - B_{\lambda\beta}(\tilde{\theta})B_{\beta\beta}^{-1}(\tilde{\theta})B_{\beta\lambda}(\tilde{\theta})$. The asymptotic distribution of (3.21) under H_A^{λ} and H_A^{ρ} can be investigated from the first order Taylor expansions of the pseudo-gradients $C_{\lambda}(\tilde{\theta})$ and $C_{\beta}(\tilde{\theta})$ around θ_0. These expansions yield

$$\begin{aligned}\sqrt{n}\, C_{\lambda}(\tilde{\theta}) = \left[-1,\, \mathcal{H}_{\lambda\beta}\mathcal{H}_{\beta\beta}^{-1}\right] \times \begin{bmatrix}-\sqrt{n}C_{\lambda}(\theta_0)\\ -\sqrt{n}C_{\beta}(\theta_0)\end{bmatrix} - \left[\mathcal{H}_{\lambda\rho} - \mathcal{H}_{\lambda\beta}\mathcal{H}_{\beta\beta}^{-1}\mathcal{H}_{\beta\rho}\right]\delta_{\rho} \\ - \left[\mathcal{H}_{\lambda\lambda} - \mathcal{H}_{\lambda\beta}\mathcal{H}_{\beta\beta}^{-1}\mathcal{H}_{\beta\lambda}\right]\delta_{\lambda} + o_p(1).\end{aligned} \tag{3.22}$$

Using the asymptotic normality of pseudo-gradients from Lemma 1 in (3.22), we obtain $\sqrt{n}\,\mathcal{C}_{\lambda}(\tilde{\theta}) \xrightarrow{d} N\big[-\mathcal{H}_{\lambda\cdot\beta}\delta_{\lambda} - \mathcal{H}_{\lambda\rho\cdot\beta}\delta_{\rho}, \mathcal{H}_{\lambda\cdot\beta}\big]$, where $\mathcal{H}_{\lambda\cdot\beta} = \left[\mathcal{H}_{\lambda\lambda} - \mathcal{H}_{\lambda\beta}\mathcal{H}_{\beta\beta}^{-1}\mathcal{H}_{\beta\lambda}\right]$, and $\mathcal{H}_{\lambda\rho\cdot\beta} = \left[\mathcal{H}_{\lambda\rho} - \mathcal{H}_{\lambda\beta}\mathcal{H}_{\beta\beta}^{-1}\mathcal{H}_{\beta\rho}\right]$. Hence, $\mathrm{LM}_{\lambda}^{\mathrm{g}}(\tilde{\theta}) \xrightarrow{d} \chi_1^2(\zeta_1)$ under H_A^{ρ} and H_A^{λ}, where $\zeta_1 = \delta_{\lambda}^2\mathcal{H}_{\lambda\cdot\beta} + \delta_{\rho}\mathcal{H}_{\lambda\rho\cdot\beta}\delta_{\lambda} + \delta_{\lambda}\mathcal{H}_{\lambda\rho\cdot\beta}^{'}\delta_{\rho} + \delta_{\rho}^2\mathcal{H}_{\lambda\rho\cdot\beta}^{'}\mathcal{H}_{\lambda\cdot\beta}^{-1}\mathcal{H}_{\lambda\rho\cdot\beta}$ is the non-centrality parameter. Let $\mathrm{LM}_{\lambda}^{\mathrm{g}\star}(\tilde{\theta})$ be the robust version of $\mathrm{LM}_{\lambda}^{\mathrm{g}}(\tilde{\theta})$, which can be obtained by adjusting the asymptotic mean and variance of $\sqrt{n}\, C_{\lambda}(\tilde{\theta})$. To this end, let $C_{\lambda}^{\star}(\tilde{\theta}) = \left[C_{\lambda}(\tilde{\theta}) - B_{\lambda\rho\cdot\beta}(\tilde{\theta})B_{\rho\cdot\beta}^{-1}(\tilde{\theta})C_{\rho}(\tilde{\theta})\right]$ be the adjusted pseudo-gradient, where $B_{\lambda\rho\cdot\beta}(\tilde{\theta}) = \left[B_{\lambda\rho}(\tilde{\theta}) - B_{\lambda\beta}(\tilde{\theta})B_{\beta\beta}^{-1}(\tilde{\theta})B_{\beta\lambda}(\tilde{\theta})\right]$. In the following proposition, we summarize the asymptotic properties of $\mathrm{LM}_{\lambda}^{\mathrm{g}}(\tilde{\theta})$ and $\mathrm{LM}_{\lambda}^{\mathrm{g}\star}(\tilde{\theta})$.

Proposition 2. — Assumptions 1–4 ensure the following results.

1. Under H_A^{λ} and H_A^{ρ}, we have
$$\mathrm{LM}_{\lambda}^{\mathrm{g}}(\tilde{\theta}) \xrightarrow{d} \chi_1^2(\zeta_1), \tag{3.23}$$
where $\zeta_1 = \delta_{\lambda}^2\mathcal{H}_{\lambda\cdot\beta} + \delta_{\rho}\mathcal{H}_{\lambda\rho\cdot\beta}\delta_{\lambda} + \delta_{\lambda}\mathcal{H}_{\lambda\rho\cdot\beta}^{'}\delta_{\rho} + \delta_{\rho}^2\mathcal{H}_{\lambda\rho\cdot\beta}^{'}\mathcal{H}_{\lambda\cdot\beta}^{-1}\mathcal{H}_{\lambda\rho\cdot\beta}$.
2. Under H_0^{λ} and irrespective of whether H_0^{ρ} or H_A^{ρ} holds,
$$\mathrm{LM}_{\lambda}^{\mathrm{g}\star}(\tilde{\theta}) = n\, C_{\lambda}^{\star'}(\tilde{\theta}) \left[B_{\lambda\cdot\beta}(\tilde{\theta}) - B_{\lambda\rho\cdot\beta}(\tilde{\theta})B_{\rho\cdot\beta}^{-1}(\tilde{\theta})B_{\lambda\rho\cdot\beta}^{'}(\tilde{\theta})\right]^{-1} C_{\lambda}^{\star}(\tilde{\theta}) \xrightarrow{d} \chi_1^2. \tag{3.24}$$
3. Under H_A^{λ} and irrespective of whether H_0^{ρ} or H_A^{ρ} holds, we have
$$\mathrm{LM}_{\lambda}^{\mathrm{g}\star}(\tilde{\theta}) \xrightarrow{d} \chi_1^2(\zeta_2), \tag{3.25}$$
where $\zeta_2 = \delta_{\lambda}^2\big(\mathcal{H}_{\lambda\cdot\beta} - \mathcal{H}_{\lambda\rho\cdot\beta}\mathcal{H}_{\rho\cdot\beta}^{-1}\mathcal{H}_{\lambda\rho\cdot\beta}^{'}\big)$.

Proof. See Appendix D. □

4. The ML Estimation Approach

As mentioned before, if the spatial weights matrices do not have rows that sum to a unique constant, i.e., $W_r l_r \neq c l_r$, where c is a constant, then the log-likelihood function of the model cannot be derived (Liu and Lee, 2010). Therefore, in this section, we consider the ML estimation of our model when $W_r l_{m_r} = M_r l_{m_r} = l_{m_r}$ holds.[10]

[10] Note that the LM test statistics suggested in this section are only valid for models that have row normalized weight matrices.

4.1. The Log-likelihood Function

In Section 3.1 , we state that if M_r has rows all sum to a constant c such that $R_r l_{m_r} = (1 - c\rho_0) l_{m_r}$, the projector reduces to the usual deviation from group mean maker $J_r = I_{m_r} - \frac{1}{m_r} l_{m_r} l_{m_r}^{'}$. Lee et al. (2010) use the orthonormal matrix, $\left[F_r,\, l_{m_r}/\sqrt{m_r}\right]$ consisting of the eigenvectors of J_r, to wipe out group fixed effects from the model.[11] Denote $Y_r^* = F_r^{'} Y_r$, $X_r^* = F_r^{'} X_r$, $\varepsilon_r^* = F_r^{'} \varepsilon_r$, $W_r^* = F_r^{'} W_r F_r$, $M_r^* = F_r^{'} M_r F_r$, $S_r^*(\lambda) = F_r^{'} S_r(\lambda) F_r = I_{m_r^*} - \lambda W_r^*$ and $R_r^*(\rho) = F_r^{'} R_r(\rho) F_r = I_{m_r^*} - \rho W_r^*$. Using Lemma 2, the transformation of the dependent variable $R_r Y_r$ to $F_r^{'} R_r Y_r$ yields

$$R_r^* Y_r^* = \lambda_0 R_r^* W_r^* Y_r^* + R_r^* X_r^* \beta_0 + \varepsilon_r^* \tag{4.1}$$

Let $\theta = \left(\rho,\, \lambda,\, \beta^{'}, \sigma^2\right)^{'}$ be the parameter vector. The log-likelihood function for the entire sample for (4.1) can be written as

$$\ln L(\theta) = -\frac{n^*}{2} \ln\left(2\pi\sigma^2\right) + \sum_{r=1}^{R} \ln\left|S_r^*(\lambda)\right| + \sum_{r=1}^{R} \ln\left|R_r^*(\rho)\right| - \frac{1}{2\sigma^2} \sum_{r=1}^{R} \varepsilon_r^{*'}(\theta)\, \varepsilon_r^*(\theta), \tag{4.2}$$

where $n^* = n - R$, and $\varepsilon_r^*(\theta) = R_r^*(\rho) S_r^*(\lambda) Y_r^* - R_r(\rho) X_r^* \beta$. Using Lemma 2, it can be shown that $\varepsilon_r^{*'}(\theta)\, \varepsilon_r^*(\theta) = \varepsilon_r^{'}(\theta) J_r \varepsilon_r(\theta)$, where $\varepsilon_r(\theta) = R_r(\rho) S_r(\lambda) Y_r - R_r(\rho) X_r \beta$. Then, again using Lemma 2, the log-likelihood function in (4.2) can be written as

$$\ln L(\theta) \quad = -\frac{n^*}{2} \ln\left(2\pi\sigma^2\right) + \ln|S(\lambda)| + \ln|R(\rho)| - R\ln((1-\lambda)(1-\rho)) - \frac{1}{2\sigma^2} \varepsilon^{'}(\theta)\, J \varepsilon(\theta), \tag{4.3}$$

where $\varepsilon(\theta) = R(\rho) S(\lambda) Y - R(\rho) X\beta$. Thus, the log-likelihood can be evaluated without the calculation of F_r. For a given value of λ and ρ, the MLE of β_0 and σ_0^2 can computed from the first order conditions of the log likelihood function. These estimators are

$$\hat{\beta}(\lambda,\, \rho) = \left(X^{'} R^{'}(\rho) J R(\rho) X\right)^{-1} X^{'} R^{'}(\rho) J R(\rho) S(\lambda) Y, \tag{4.4}$$

$$\hat{\sigma}^2(\lambda,\, \rho) = \frac{1}{n^*} Y^{'} S^{'}(\lambda) R^{'}(\rho) P(\rho) R(\rho) S(\lambda) Y, \tag{4.5}$$

where $P(\rho) = J - J R(\rho) X \left(X^{'} R^{'}(\rho) J R(\rho) X\right)^{-1} X^{'} R^{'}(\rho) J$. Then, the concentrated log-likelihood function is given by

$$\ln L(\lambda,\, \rho) = -\frac{n^*}{2}\big(\ln(2\pi) + 1\big) - \frac{n^*}{2} \ln \hat{\sigma}^2(\lambda, \rho) + \ln|S(\lambda)| + \ln|R(\rho)| - R\ln\big((1-\lambda)(1-\rho)\big). \tag{4.6}$$

The MLE of λ_0 and ρ_0 is obtained by the maximization of (4.6). We assume the following regularity conditions for the consistency and the asymptotic distribution of the MLE.

Assumption 5. *The innovation terms ε_{ir}s are i.i.d normal with zero mean and variance σ_0^2, and* $\mathrm{E}\left(|\varepsilon_{ir}|^{2+\tau}\right) < \infty$ *for some $\tau > 0$, for all $i = 1, \ldots, m_r$ and $r = 1, \ldots, R$.*[12]

Assumption 6. *(i) The elements X are uniformly bounded constants for all n, (ii) X has the full rank of $k = k_1 + k_2$, and (iii) $\lim_{n\to\infty} \frac{1}{n} X^{'} R^{'} J R X$ exists and is nonsingular.*

Assumption 7. *(i) The row and column sums of W and M are bounded uniformly in absolute value, (ii) λ_0 and ρ_0 are in the interior of a compact parameter space Γ, (iii) the row and column sums of $S^{-1}(\lambda)$ and $R^{-1}(\rho)$ are bounded uniformly in absolute value for all $(\lambda, \rho) \in \Gamma$.*

[11] Note that F_r has the following properties: $F_r^{'} l_{m_r} = 0$, $F_r^{'} F_r = I_{m_r^*}$, where $m_r^* = m_r - 1$, and $F_r F^{'} = J_r$. For some other properties, see Lemma 2. Burridge et al. (2016) provide an explicit expression for F_r.

[12] Note that the existence of $(4+\tau)$th moments of ε_{ir} are required when ε_{ir}s are simply i.i.d. (Kelejian and Prucha, 2001).

Under Assumptions 5– 7, the following result for the MLE $\hat{\theta}$ can be established (Lee et al., 2010).[13]

$$\sqrt{n^*}\left(\hat{\theta}-\theta_0\right) \xrightarrow{d} N\left[0, \left(\lim_{n\to\infty} \Sigma\right)^{-1}\right], \tag{4.7}$$

where $\Sigma = \mathrm{E}\left[-\frac{1}{n^*}\frac{\partial \ln L(\theta_0)}{\partial\theta\partial\theta'}\right]$.[14]

4.2. The LM Tests for Spatial Autoregressive Parameters

In this section, we consider the LM statistics for testing H_0^{ρ} and H_0^{λ}. Our test statistics are similar to those suggested in Anselin et al. (1996). Note that the test statistics suggested in Anselin et al. (1996) cannot be directly used for our model, since the log-likelihood function of our model is so different and complex from the one used in Anselin et al. (1996) to formulate the test statistics. When there are no group fixed effects, i.e., $\alpha_0 = 0$, our model reduces to the cross-sectional model studied in Anselin et al. (1996). Thus, our results can be considered as an extension of results in Anselin et al. (1996).

Denote $\gamma = (\beta', \sigma^2)'$ and $\gamma_0 = (\beta_0', \sigma_0^2)'$. Let $L_a(\theta) = \frac{1}{n^*}\frac{\partial \ln L(\theta)}{\partial a}$, $L_{aa}(\theta) = \frac{1}{n^*}\frac{\partial^2 L(\theta)}{\partial a \partial a'}$, where $a = \rho, \lambda, \gamma$, $I(\theta) = \Sigma(\theta)$, and $I = \lim_{n\to\infty}\Sigma$.[15] With these new notations, the standard LM test statistic for the restrictions of the form $\pi(\theta_0) = 0$ is given by

$$\mathrm{LM}_0^{\mathrm{m}}(\hat{\theta}_r) = n^* L'(\hat{\theta}_r)\left[I(\hat{\theta}_r)\right]^{-1} L(\hat{\theta}_r), \tag{4.8}$$

where $\hat{\theta}_r = \arg\max_{\{\theta:\pi(\theta)=0\}} \ln L(\theta)$ is the restricted MLE and $I(\hat{\theta}_r)$ is the plug in estimator of I.

In order to formulate similar test statistics, consider the following partition of $I(\theta)$ and $I(\theta_0)$:

$$I(\theta) = \begin{bmatrix} \underbrace{I_{\rho\rho}(\theta)}_{1\times 1} & \underbrace{I_{\rho\lambda}(\theta)}_{1\times 1} & \underbrace{I_{\rho\gamma}(\theta)}_{1\times(k+1)} \\ \underbrace{I_{\lambda\rho}(\theta)}_{1\times 1} & \underbrace{I_{\lambda\lambda}(\theta)}_{1\times 1} & \underbrace{I_{\lambda\gamma}(\theta)}_{1\times(k+1)} \\ \underbrace{I_{\gamma\rho}(\theta)}_{(k+1)\times 1} & \underbrace{I_{\gamma\lambda}(\theta)}_{(k+1)\times 1} & \underbrace{I_{\gamma\gamma}(\theta)}_{(k+1)\times(k+1)} \end{bmatrix}, \; I = \begin{bmatrix} \underbrace{I_{\rho\rho}}_{1\times 1} & \underbrace{I_{\rho\lambda}}_{1\times 1} & \underbrace{I_{\rho\gamma}}_{1\times(k+1)} \\ \underbrace{I_{\lambda\rho}}_{1\times 1} & \underbrace{I_{\lambda\lambda}}_{1\times 1} & \underbrace{I_{\lambda\gamma}}_{1\times(k+1)} \\ \underbrace{I_{\gamma\rho}}_{(k+1)\times 1} & \underbrace{I_{\gamma\lambda}}_{(k+1)\times 1} & \underbrace{I_{\gamma\gamma}}_{(k+1)\times(k+1)} \end{bmatrix}. \tag{4.9}$$

Let $\tilde{\theta} = (\rho_\star, \lambda_\star, \tilde{\gamma})'$ be the restricted MLE when $\mathrm{H}_0 : \rho_0 = \rho_\star, \lambda_0 = \lambda_\star$ holds. First, we consider the LM test for the joint null hypothesis $\mathrm{H}_0 : \rho_0 = \rho_\star, \lambda_0 = \lambda_\star$. The test statistic is given by

$$\mathrm{LM}_{\rho\lambda}^{\mathrm{m}}(\tilde{\theta}) = n^* \mathbf{L}_{\rho\lambda}'(\tilde{\theta})\left[\mathbf{I}_{1\cdot 3}(\tilde{\theta})\right]^{-1} \mathbf{L}_{\rho\lambda}(\tilde{\theta}), \tag{4.10}$$

where $\mathbf{L}_{\rho\lambda}(\tilde{\theta}) = \left[L_\rho(\tilde{\theta}), L_\lambda(\tilde{\theta})\right]'$, $\mathbf{I}_{1\cdot 3}(\tilde{\theta}) = \mathbf{I}_{11}(\tilde{\theta}) - \mathbf{I}_{13}(\tilde{\theta}) I_{\gamma\gamma}^{-1}(\tilde{\theta}) \mathbf{I}_{31}(\tilde{\theta})$, $\mathbf{I}_{11}(\tilde{\theta}) = \begin{bmatrix} I_{\rho\rho}(\tilde{\theta}) & I_{\rho\lambda}(\tilde{\theta}) \\ I_{\lambda\rho}(\tilde{\theta}) & I_{\lambda\lambda}(\tilde{\theta}) \end{bmatrix}$, and $\mathbf{I}_{31}(\tilde{\theta}) = \mathbf{I}_{13}'(\tilde{\theta}) = \left[I_{\gamma\rho}(\tilde{\theta}), I_{\gamma\lambda}(\tilde{\theta})\right]$.

Next, following Bera and Yoon (1993), we formulate test statistics that are similar to those stated in Propositions 1 and 2 for the null hypotheses given in (3.10) and (3.11). Again, we first consider the problem of testing $\mathrm{H}_0^{\rho} : \rho_0 = \rho_\star$ when $\mathrm{H}_0^{\lambda} : \lambda_0 = \lambda_\star$ holds. Then, the one directional test statistic can be formulated as

$$\mathrm{LM}_{\rho}^{\mathrm{m}}(\tilde{\theta}) = n^* L_\rho'(\tilde{\theta})\left[I_{\rho\cdot\gamma}(\tilde{\theta})\right]^{-1} L_\rho(\tilde{\theta}), \tag{4.11}$$

[13] Lee et al. (2010) investigate the identification conditions in the ML framework and they state these conditions. Here, we simply assume that the parameters are identified.

[14] The explicit forms of Σ is given in Appendix C.

[15] The test statistics suggested in this section are formulated with $L(\theta)$ and $I(\theta)$. In Appendix C, we give explicit expressions for these terms.

where $I_{\rho\cdot\gamma}(\tilde{\theta}) = I_{\rho\rho}(\tilde{\theta}) - I_{\rho\gamma}(\tilde{\theta})I_{\gamma\gamma}^{-1}(\tilde{\theta})I_{\gamma\rho}(\tilde{\theta})$. The distribution of (4.11) under H_A^{ρ} and H_A^{λ} can be investigated from the first order Taylor expansion of $L_{\rho}(\tilde{\theta})$ and $L_{\gamma}(\tilde{\theta})$ around θ_0 (Saikkonen, 1989). The Taylor expansions can be derived as[16]

$$\sqrt{n^*}L_{\rho}(\tilde{\theta}) = \sqrt{n^*}L_{\rho}(\theta_0) - L_{\rho\rho}(\theta_0)\,\delta_{\rho} - L_{\rho\lambda}(\theta_0)\,\delta_{\lambda} + \sqrt{n^*}L_{\rho\gamma}(\theta_0)\,(\tilde{\gamma}-\gamma_0) + o_p(1), \tag{4.12}$$

$$\sqrt{n^*}L_{\gamma}(\tilde{\theta}) = \sqrt{n^*}L_{\gamma}(\theta_0) - L_{\gamma\rho}(\theta_0)\,\delta_{\rho} - L_{\gamma\lambda}(\theta_0)\,\delta_{\lambda} + \sqrt{n^*}L_{\gamma\gamma}(\theta_0)\,(\tilde{\gamma}-\gamma_0) + o_p(1). \tag{4.13}$$

Using (4.12), (4.13) and Lemma 3, we can obtain the following result.

$$\sqrt{n^*}L_{\rho}(\tilde{\theta}) = \left[1,\,-I_{\rho\gamma}I_{\gamma\gamma}^{-1}\right] \times \begin{bmatrix}\sqrt{n^*}L_{\rho}(\theta_0)\\ \sqrt{n^*}L_{\gamma}(\theta_0)\end{bmatrix} + \left[I_{\rho\rho} - I_{\rho\gamma}I_{\gamma\gamma}^{-1}I_{\gamma\rho}\right]\delta_{\rho} + \left[I_{\rho\lambda} - I_{\rho\gamma}I_{\gamma\gamma}^{-1}I_{\gamma\lambda}\right]\delta_{\lambda} + o_p(1). \tag{4.14}$$

The asymptotic distribution of $\sqrt{n^*}L_{\rho}(\tilde{\theta})$ can be determined from (4.14) by using the asymptotic normality of score functions (see Lemma 3). Hence, we can obtain $\sqrt{n^*}L_{\rho}(\tilde{\theta}) \xrightarrow{d} N\left[I_{\rho\cdot\gamma}\delta_{\rho} + I_{\rho\lambda\cdot\gamma}\delta_{\lambda}, I_{\rho\cdot\gamma}\right]$, where $I_{\rho\cdot\gamma} = \left[I_{\rho\rho} - I_{\rho\gamma}I_{\gamma\gamma}^{-1}I_{\gamma\rho}\right]$ and $I_{\rho\lambda\cdot\gamma} = \left[I_{\rho\lambda} - I_{\rho\gamma}I_{\gamma\gamma}^{-1}I_{\gamma\lambda}\right]$. This last result along with (4.14) can be used to determine the asymptotic distributions of $\mathrm{LM}_1^{\mathrm{m}}$ and its robust version $\mathrm{LM}_1^{\mathrm{m}\star}$ under the null and the local alternatives. We summarize these asymptotic results in the following proposition.

Proposition 3. — Under Assumptions 5-7, the following results hold.

1. Under H_A^{ρ} and H_A^{λ}, we have

$$\mathrm{LM}_{\rho}^{\mathrm{m}}(\tilde{\theta}) \xrightarrow{d} \chi_1^2(\varphi_1), \tag{4.15}$$

   where $\varphi_1 = \delta_{\rho}^2 I_{\rho\cdot\gamma} + \delta_{\rho}I_{\rho\lambda\cdot\gamma}\delta_{\lambda} + \delta_{\lambda}I_{\rho\lambda\cdot\gamma}^{'}\delta_{\rho} + \delta_{\lambda}^2 I_{\rho\lambda\cdot\gamma}^{'}I_{\rho\cdot\gamma}^{-1}I_{\rho\lambda\cdot\gamma}$.

2. Under $\mathrm{H}_0^{\rho}: \rho_0 = \rho_{\star}$ and irrespective of whether H_0^{λ} or H_A^{λ} holds, we have

$$\mathrm{LM}_{\rho}^{\mathrm{m}\star}(\tilde{\theta}) = n^* L_{\rho}^{\star'}(\tilde{\theta})\left[I_{\rho\cdot\gamma}(\tilde{\theta}) - I_{\rho\lambda\cdot\gamma}(\tilde{\theta})I_{\lambda\cdot\gamma}^{-1}(\tilde{\theta})I_{\rho\lambda\cdot\gamma}^{'}(\tilde{\theta})\right]^{-1} L_{\rho}^{\star}(\tilde{\theta}) \xrightarrow{d} \chi_1^2, \tag{4.16}$$

   where $L_{\rho}^{\star}(\tilde{\theta}) = \left[L_{\rho}(\tilde{\theta}) - \mathcal{I}_{\rho\lambda\cdot\gamma}(\tilde{\theta})I_{\lambda\cdot\gamma}^{-1}(\tilde{\theta})L_{\lambda}(\tilde{\theta})\right]$ is the adjusted score function, $I_{\rho\lambda\cdot\gamma}(\tilde{\theta}) = I_{\rho\lambda}(\tilde{\theta}) - I_{\rho\gamma}(\tilde{\theta})I_{\gamma\gamma}^{-1}(\tilde{\theta})I_{\gamma\lambda}(\tilde{\theta})$ and $I_{\lambda\cdot\gamma}(\tilde{\theta}) = I_{\lambda\lambda}(\tilde{\theta}) - I_{\lambda\gamma}(\tilde{\theta})I_{\gamma\gamma}^{-1}(\tilde{\theta})I_{\gamma\lambda}(\tilde{\theta})$.

3. Under H_A^{ρ} and irrespective of whether H_0^{λ} or H_A^{λ} holds, we have

$$\mathrm{LM}_{\rho}^{\mathrm{m}\star}(\tilde{\theta}) \xrightarrow{d} \chi_1^2(\varphi_2), \tag{4.17}$$

   where $\varphi_2 = \delta_{\rho}^2\left(I_{\rho\cdot\gamma} - I_{\rho\lambda\cdot\gamma}I_{\lambda\cdot\gamma}^{-1}I_{\rho\lambda\cdot\gamma}^{'}\right)$.

Proof. See Appendix D. □

Now, we consider the null hypothesis $\mathrm{H}_0^{\lambda}: \lambda_0 = \lambda_{\star}$, when $\mathrm{H}_0^{\rho}: \rho_0 = \rho_{\star}$ holds. Then, the one-directional LM test for this hypothesis can be expressed as

$$\mathrm{LM}_{\lambda}^{\mathrm{m}}(\tilde{\theta}) = n^* L_{\lambda}^{'}(\tilde{\theta})\left[I_{\lambda\cdot\gamma}(\tilde{\theta})\right]^{-1} L_{\lambda}(\tilde{\theta}), \tag{4.18}$$

where $I_{\lambda\cdot\gamma}(\tilde{\theta}) = I_{\lambda\lambda}(\tilde{\theta}) - I_{\lambda\gamma}(\tilde{\theta})I_{\gamma\gamma}^{-1}(\tilde{\theta})I_{\gamma\lambda}(\tilde{\theta})$. The distribution of (4.18) can be investigated from the first order Taylor expansion of $L_{\lambda}(\tilde{\theta})$ and $L_{\gamma}(\tilde{\theta})$ around θ_0 when H_A^{λ} and H_A^{ρ} hold. It can be shown that these first order expansions are

$$\sqrt{n^*}L_{\lambda}(\tilde{\theta}) = \sqrt{n^*}L_{\lambda}(\theta_0) - L_{\lambda\rho}(\theta_0)\,\delta_{\rho} - L_{\lambda\lambda}(\theta_0)\,\delta_{\lambda} + \sqrt{n^*}L_{\lambda\gamma}(\theta_0)\left(\tilde{\gamma}-\gamma_0\right) + o_p(1) \tag{4.19}$$

$$\sqrt{n^*}L_{\gamma}\left(\tilde{\theta}\right) = \sqrt{n^*}L_{\gamma}(\theta_0) - L_{\gamma\rho}(\theta_0)\,\delta_{\rho} - L_{\gamma\lambda}(\theta_0)\,\delta_{\lambda} + \sqrt{n^*}L_{\gamma\gamma}(\theta_0)\left(\tilde{\gamma}-\gamma_0\right) + o_p(1). \tag{4.20}$$

[16] See Corollary 5.1.5 of Fuller (1996).

Then, using (4.19), (4.20) and Lemma 3, we can obtain

$$\sqrt{n^*}L_\lambda(\tilde{\theta}) = \left[1,\ -I_{\rho\gamma}I_{\gamma\gamma}^{-1}\right] \times \begin{bmatrix}\sqrt{n^*}L_\lambda(\theta_0)\\ \sqrt{n^*}L_\gamma(\theta_0)\end{bmatrix} + \left[I_{\lambda\rho} - I_{\lambda\gamma}I_{\gamma\gamma}^{-1}I_{\gamma\rho}\right]\delta_\rho + \left[I_{\lambda\lambda} - I_{\lambda\gamma}I_{\gamma\gamma}^{-1}I_{\gamma\lambda}\right]\delta_\lambda + o_p(1). \quad (4.21)$$

The asymptotic distribution of $\sqrt{n^*}L_\lambda(\tilde{\theta})$ in (4.21) can be determined from the asymptotic distribution of score functions in the right hand side of (4.21) (see Lemma 3). Hence, we can show that $\sqrt{n^*}L_\lambda(\tilde{\theta}) \xrightarrow{d} N\left[I_{\lambda\cdot\gamma}\delta_\lambda + I_{\lambda\rho\cdot\gamma}\delta_\rho, I_{\lambda\cdot\gamma}\right]$, where $I_{\lambda\cdot\gamma} = \left[I_{\lambda\lambda} - I_{\lambda\gamma}I_{\gamma\gamma}^{-1}I_{\gamma\lambda}\right]$, and $I_{\lambda\rho\cdot\gamma} = \left[I_{\lambda\rho} - I_{\lambda\gamma}I_{\gamma\gamma}^{-1}I_{\gamma\rho}\right]$. This last result together with (4.21) implies the following proposition.

Proposition 4. — Under our Assumptions 5–7, the following results hold.

1. Under H_A^λ and H_A^ρ, we have

$$\mathrm{LM}_\lambda^{\mathrm{m}}(\tilde{\theta}) \xrightarrow{d} \chi_1^2(\psi_1), \quad (4.22)$$

where $\psi_1 = \delta_\lambda^2 I_{\lambda\cdot\gamma} + \delta_\rho I_{\lambda\rho\cdot\gamma}\delta_\lambda + \delta_\lambda I_{\lambda\rho\cdot\gamma}^{'}\delta_\rho + \delta_\rho^2 I_{\lambda\rho\cdot\gamma}^{'}I_{\lambda\cdot\gamma}^{-1}I_{\lambda\rho\cdot\gamma}$.

2. For the robust test $\mathrm{LM}_\lambda^{\mathrm{m}\star}(\tilde{\theta})$, under H_0^λ and irrespective of whether H_0^ρ or H_A^ρ holds, we have

$$\mathrm{LM}_\lambda^{\mathrm{m}\star}(\tilde{\theta}) = n^* L_\lambda^{\star'}(\tilde{\theta})\left[I_{\lambda\cdot\gamma}(\tilde{\theta}) - \mathcal{I}_{\lambda\rho\cdot\gamma}(\tilde{\theta})I_{\rho\cdot\gamma}^{-1}(\tilde{\theta})I_{\lambda\rho\cdot\gamma}^{'}(\tilde{\theta})\right]^{-1} L_\rho^\star(\tilde{\theta}) \xrightarrow{d} \chi_1^2, \quad (4.23)$$

where $L_\lambda^\star(\tilde{\theta}) = \left[L_\lambda(\tilde{\theta}) - I_{\lambda\rho\cdot\gamma}(\tilde{\theta})I_{\rho\cdot\gamma}^{-1}(\tilde{\theta})L_\rho(\tilde{\theta})\right]$ is the adjusted gradient, and $I_{\lambda\rho\cdot\gamma}(\tilde{\theta}) = \left[I_{\lambda\rho}(\tilde{\theta}) - I_{\lambda\gamma}(\tilde{\theta})I_{\gamma\gamma}^{-1}(\tilde{\theta})I_{\gamma\rho}(\tilde{\theta})\right]$.

3. Under H_A^λ and irrespective of whether H_0^ρ or H_A^ρ holds, we have

$$\mathrm{LM}_\lambda^{\mathrm{m}\star}(\tilde{\theta}) \xrightarrow{d} \chi_1^2(\psi_2), \quad (4.24)$$

where $\psi_2 = \delta_\lambda^2\left(I_{\lambda\cdot\gamma} - I_{\lambda\rho\cdot\gamma}I_{\rho\cdot\gamma}^{-1}I_{\lambda\rho\cdot\gamma}^{'}\right)$.

Proof. See Appendix D. □

Note that Propositions 3 and 4 show that the robust versions of tests have less asymptotic power than the corresponding one directional tests when there is no parametric misspecification in the model.

5. The Test Statistics for Contextual Effects

The social interaction effects through observed peers' characteristics is known as the contextual effects and is measured by $k_2 \times 1$ parameter vector β_{02} in our model. In spatial econometric literature, the associated matrix $W_r X_{2r}$ is called the spatial Durbin term. On motivations for specifications that include spatial Durbin terms, see LeSage and Pace (2009), Elhorst (2014), Halleck Vega and Elhorst (2015) and Burridge et al. (2016). In this section, we consider the GMM gradient tests and the ML score tests for hypotheses about β_{02}.

First, we state the test statistics in the GMM framework. For notational simplicity, let $\psi_0 = \beta_{02}$, $\phi_0 = (\rho_0, \lambda_0)^{'}$ and $\gamma_0 = \beta_{01}$ be true parameter vectors. Then, ψ, ϕ and γ denote arbitrary parameter values in the parameter space. Let $\theta_0 = (\psi_0^{'}, \phi_0^{'}, \gamma_0^{'})^{'}$ be the parameter vector of the model. We assume that $G(\theta)$, $C(\theta)$, $B(\theta)$ and $\mathcal{H}$, which are defined in Section 3.2, are partitioned according to dimensions of ψ, ϕ and γ. Consider $H_0^\psi : \psi_0 = \psi_\star$ and $H_0^\phi : \phi_0 = \phi_\star$, where $\psi_\star$ and $\phi_\star$ are hypothesized values under the null. The sequence of local alternatives are $H_A^\psi : \psi_0 = \psi_\star + \delta_\psi/\sqrt{n}$ and $H_A^\phi : \phi_0 = \phi_\star + \delta_\phi/\sqrt{n}$, where δ_ψ and δ_ϕ are bounded vectors. We can determine the GMM gradient test statistics for $H_0^\psi : \psi_0 = \psi_\star$ by following the similar arguments used for Proposition 1. The GMM gradient test for H_0^ψ when H_0^ϕ holds is

$$\mathrm{LM}_\psi^{\mathrm{g}}(\tilde{\theta}) = n\, C_\psi^{'}(\tilde{\theta})\left[B_{\psi\cdot\gamma}(\tilde{\theta})\right]^{-1} C_\psi(\tilde{\theta}), \quad (5.1)$$

where $B_{\psi\cdot\gamma}(\tilde{\theta}) = B_{\psi\psi}(\tilde{\theta}) - B_{\psi\gamma}(\tilde{\theta})B_{\gamma\gamma}^{-1}(\tilde{\theta})B_{\gamma\psi}(\tilde{\theta})$ and $\tilde{\theta} = (\psi_\star^{'}, \phi_\star^{'}, \tilde{\gamma})^{'}$ is the restricted optimal GMME. In the following proposition, we summarize the asymptotic results for $\mathrm{LM}_\psi^{\mathrm{g}}(\tilde{\theta})$ and its robust version.

Proposition 5. — Under Assumptions 1–4, the following results hold.

1. Under H_A^{ψ} and H_A^{ϕ}, we have

$$\mathrm{LM}_{\psi}^{\mathrm{g}}(\tilde{\theta}) \xrightarrow{d} \chi_{k_2}^2(\eta_1), \tag{5.2}$$

   where $\eta_1 = \delta_{\psi}' \mathcal{H}_{\psi\cdot\gamma}\delta_{\psi} + \delta_{\psi}'\mathcal{H}_{\psi\phi\cdot\gamma}\delta_{\phi} + \delta_{\phi}'\mathcal{H}_{\psi\phi\cdot\gamma}'\delta_{\psi} + \delta_{\phi}'\mathcal{H}_{\psi\phi\cdot\gamma}'\mathcal{H}_{\psi\cdot\gamma}^{-1}\mathcal{H}_{\psi\phi\cdot\gamma}\delta_{\phi}$.

2. Under H_0^{ψ} and irrespective of whether H_0^{ϕ} or H_A^{ϕ} holds, we have

$$\mathrm{LM}_{\psi}^{\mathrm{g}\star}(\tilde{\theta}) = n\, C_{\psi}^{\star\prime}(\tilde{\theta}) \left[B_{\psi\cdot\gamma}(\tilde{\theta}) - B_{\psi\phi\cdot\gamma}(\tilde{\theta})B_{\phi\cdot\gamma}^{-1}(\tilde{\theta})B_{\psi\phi\cdot\gamma}'(\tilde{\theta})\right]^{-1} C_{\psi}^{\star}(\tilde{\theta}) \xrightarrow{d} \chi_{k_2}^2, \tag{5.3}$$

   where $C_{\psi}^{\star}(\tilde{\theta}) = \left[C_{\psi}(\tilde{\theta}) - B_{\psi\phi\cdot\gamma}(\tilde{\theta})B_{\phi\cdot\gamma}^{-1}(\tilde{\theta})C_{\phi}(\tilde{\theta})\right]$ is the adjusted pseudo-gradient, and $B_{\psi\cdot\gamma}(\tilde{\theta}) = \left[B_{\psi\psi}(\tilde{\theta}) - B_{\psi\gamma}(\tilde{\theta})B_{\gamma\gamma}^{-1}(\tilde{\theta})B_{\gamma\psi}(\tilde{\theta})\right]$.

3. Under H_A^{ψ} and irrespective of whether H_0^{ϕ} or H_A^{ϕ} holds, we have

$$\mathrm{LM}_{\psi}^{\mathrm{g}\star}(\tilde{\theta}) \xrightarrow{d} \chi_{k_2}^2(\eta_2), \tag{5.4}$$

   where $\eta_2 = \delta_{\psi}'\big(\mathcal{H}_{\psi\cdot\gamma} - \mathcal{H}_{\psi\phi\cdot\gamma}\mathcal{H}_{\phi\cdot\gamma}^{-1}\mathcal{H}_{\psi\phi\cdot\gamma}'\big)\delta_{\psi}$.

Proof. See Appendix D. □

Next, we state the test statistics in the ML framework. Let $\psi_0 = \beta_{02}$, $\phi_0 = (\rho_0, \lambda_0)'$ and $\gamma_0 = (\beta_{01}', \sigma_0^2)'$ be true parameter vectors. The combined parameter vector is denoted by $\theta_0 = (\psi_0', \phi_0', \gamma_0')'$. We assume that $I(\theta)$ and I defined in Section 4.2 are partitioned according to dimensions of ψ, ϕ and γ. The LM test statistic for H_0^{ψ} when H_0^{ϕ} holds is, then, given by

$$\mathrm{LM}_{\psi}^{\mathrm{m}}(\tilde{\theta}) = n^* L_{\psi}'(\tilde{\theta}) \left[I_{\psi\cdot\gamma}\left(\tilde{\theta}\right)\right]^{-1} L_{\psi}(\tilde{\theta}), \tag{5.5}$$

where $I_{\psi\cdot\gamma}(\tilde{\theta}) = I_{\psi\psi}(\tilde{\theta}) - I_{\psi\gamma}(\tilde{\theta})I_{\gamma\gamma}^{-1}(\tilde{\theta})I_{\gamma\psi}(\tilde{\theta})$ and $\tilde{\theta} = (\psi_{\star}', \phi_{\star}', \tilde{\gamma})'$ is the restricted MLE. The next proposition summarizes asymptotic results for this test statistic and its robust version.

Proposition 6. — Under our Assumptions 5-7, the following results hold.

1. Under H_A^{ψ} and H_A^{ϕ}, we have

$$\mathrm{LM}_{\psi}^{\mathrm{m}}(\tilde{\theta}) \xrightarrow{d} \chi_{k_2}^2(\mu_1), \tag{5.6}$$

   where $\mu_1 = \delta_{\psi}' I_{\psi\cdot\gamma}\delta_{\psi} + \delta_{\psi} I_{\psi\phi\cdot\gamma}\delta_{\phi} + \delta_{\phi} I_{\psi\phi\cdot\gamma}'\delta_{\psi} + \delta_{\phi}' I_{\psi\phi\cdot\gamma}' I_{\psi\cdot\gamma}^{-1} I_{\psi\phi\cdot\gamma}\delta_{\phi}$.

2. Under H_0^{ψ} and irrespective of whether H_0^{ϕ} or H_A^{ϕ} holds, the distribution of the robust test $\mathrm{LM}_{\psi}^{\mathrm{m}\star}\big(\tilde{\theta}\big)$ is given by

$$\mathrm{LM}_{\psi}^{\mathrm{m}\star}(\tilde{\theta}) = n^* L_{\psi}^{\star\prime}(\tilde{\theta}) \left[I_{\psi\cdot\gamma}(\tilde{\theta}) - I_{\psi\phi\cdot\gamma}(\tilde{\theta})I_{\phi\cdot\gamma}^{-1}(\tilde{\theta})I_{\psi\phi\cdot\gamma}'(\tilde{\theta})\right]^{-1} L_{\psi}^{\star}(\tilde{\theta}) \xrightarrow{d} \chi_{k_2}^2, \tag{5.7}$$

   where $L_{\psi}^{\star}(\tilde{\theta}) = \left[L_{\psi}(\tilde{\theta}) - I_{\psi\phi\cdot\gamma}(\tilde{\theta})I_{\phi\cdot\gamma}^{-1}(\tilde{\theta})L_{\phi}(\tilde{\theta})\right]$ is the adjusted score function, $I_{\psi\phi\cdot\gamma} = I_{\psi\phi}(\tilde{\theta}) - I_{\psi\gamma}(\tilde{\theta})I_{\gamma\gamma}^{-1}(\tilde{\theta})I_{\gamma\phi}(\tilde{\theta})$ and $I_{\phi\cdot\gamma}(\tilde{\theta}) = I_{\phi\phi}(\tilde{\theta}) - I_{\phi\gamma}(\tilde{\theta})I_{\gamma\gamma}^{-1}(\tilde{\theta})I_{\gamma\phi}(\tilde{\theta})$.

3. Under H_A^{ψ} and irrespective of whether H_0^{ϕ} or H_A^{ϕ} holds, we have

$$\mathrm{LM}_{\psi}^{\mathrm{m}\star}(\tilde{\theta}) \xrightarrow{d} \chi_{k_2}^2(\mu_2), \tag{5.8}$$

   where $\mu_2 = \delta_{\psi}'\big(I_{\psi\cdot\gamma} - I_{\psi\phi\cdot\gamma}I_{\phi\cdot\gamma}^{-1}I_{\psi\phi\cdot\gamma}'\big)\delta_{\psi}$.

Proof. See Appendix D. □

Under H_A^{ψ} and H_0^{ϕ}, Propositions 5 and 6, respectively, show that $\eta_1 - \eta_2 \geq 0$ and $\mu_1 - \mu_2 \geq 0$. That is, the robust versions of tests have less asymptotic power than the corresponding one directional tests when there is no parametric misspecification in the model, i.e., when $\phi_0 = \phi_\star$.

Remark 1. — The test statistics suggested in Propositions 5 and 6 are robust to the local presence of λ_0 and ρ_0. Note that Propositions 5 and 6 are general enough and can easily be adjusted to formulate the test statistics for some other hypotheses of interest. For example, the test statistic that is only robust to the local presence of λ_0 can be obtained simply by setting $\phi_0 = \lambda_0$ and $\gamma = (\rho_0, \beta_{01}^{'})^{'}$ in Proposition 5, and $\phi_0 = \lambda_0$ and $\gamma = (\rho_0, \beta_{01}^{'}, \sigma_0^2)^{'}$ in Proposition 6. Similarly, the test statistic that is only robust to the local presence of ρ_0 can be obtained by setting $\phi_0 = \rho_0$ and $\gamma = (\lambda_0, \beta_{01}^{'})^{'}$ in Proposition 5, and $\phi_0 = \rho_0$ and $\gamma = (\lambda_0, \beta_{01}^{'}, \sigma_0^2)^{'}$ in Proposition 6.

6. The Relationship Between Test Statistics

There are four important observations regarding to the robust tests. First, the robust tests introduced by Bera and Yoon (1993) and(Bera et al., 2010) share the optimality property of the Neyman's $C(\alpha)$ test. In particular, Bera and Yoon (1993) show that the robust test is asymptotically equivalent to Neyman's $C(\alpha)$ test under the null and the local alternatives. It is important to note that the motivation for both tests are different. In the case of the robust test, the one-directional test statistic is adjusted in such a way that it has a central chi-square distribution when the alternative model has a local parametric misspecification. On the other hand, the $C(\alpha)$ test is developed in a framework that involves several nuisance parameters. In such a framework, an optimal test is the one that has the highest power among the class of tests obtaining the same size. To achieve the optimality, the $C(\alpha)$ test statistic is constructed in such a way that it is orthogonal to the gradients with respect to the nuisance parameters. The $C(\alpha)$ test can be computed with any consistent estimator and it reduces to the standard LM test when it is formulated with the optimal restricted GMME or the restricted MLE.

Second, the robust tests are formulated by an estimator obtained under the joint null hypothesis $\mathrm{H}_0 : \rho_0 = \rho_\star, \lambda_0 = \lambda_\star$. Under the joint null, the model reduces to a one-way panel data type model $Y_r = X_{1r}\beta_{01} + W_r X_{2r}\beta_{02} + l_{m_r}\alpha_{0r} + \varepsilon_r$, which can be estimated by an OLSE. Therefore, the computation of test statistics does not require any nonlinear optimization routines. On the other hand, the conditional LM tests (see LM_{ρ}^{jA}, $\mathrm{LM}_{\lambda}^{jA}$ LM_{ψ}^{jA} and LM_{ψ}^{jA}, where $j = \mathrm{g}, \mathrm{m}$ in Tables 1 and 2) require the estimation of spatial parameters, which can be computationally involved.

Third, it is easy to check whether a robust test reduces to a one-directional test. Recall that the adjusted gradients are in the forms of $L_\lambda^\star(\tilde\theta) = \Big[L_\lambda(\tilde\theta) - I_{\lambda\rho\cdot\gamma}(\tilde\theta)I_{\rho\cdot\gamma}^{-1}(\tilde\theta)L_\rho(\tilde\theta)\Big]$ and $C_\lambda^\star(\tilde\theta) = \Big[C_\lambda(\tilde\theta) - \mathcal{B}_{\lambda\rho\cdot\beta}(\tilde\theta)B_{\rho\cdot\beta}^{-1}(\tilde\theta)C_\rho(\tilde\theta)\Big]$ for $\mathrm{H}_0 : \lambda_0 = \lambda_\star$. Hence, the robust tests formulated with these adjusted gradients reduce to the corresponding one-directional tests when $I_{\lambda\rho\cdot\gamma} = 0$ and $B_{\lambda\rho\cdot\beta} = 0$. In such cases, the one directional tests are valid under the local presence of ρ in the alternative model. Similarly, in the case of $H_0 : \psi_0 = \psi_\star$, the robust test statistics reduce to the corresponding one directional test statistics when $I_{\psi\phi\cdot\gamma} = 0$ and $B_{\psi\phi\cdot\gamma} = 0$.

Finally, the test statistic for the joint null $\mathrm{H}_0 : \rho_0 = \rho_\star$, $\lambda_0 = \lambda_\star$ can be decomposed into two orthogonal components: (i) the robust test statistic, and (ii) the one directional test statistic. In the context of the GMM framework, the joint test statistic is formulated with $\Big[\mathbf{B}_{1\cdot3}(\tilde\theta)\Big]^{-1}$ in (3.14). By the inverse of the partitioned matrix, we have

$$
\begin{aligned}
\Big[\mathbf{B}_{1\cdot3}(\tilde\theta)\Big]^{-1} &= \begin{bmatrix} \mathbf{A}_1^{-1} & -\mathbf{A}_1^{-1}B_{12\cdot3}B_{2\cdot3}^{-1} \\ -B_{\lambda\cdot\beta}^{-1}B_{\lambda\rho\cdot\beta}\mathbf{A}_1^{-1} & B_{\lambda\cdot\beta}^{-1} + B_{\lambda\cdot\beta}^{-1}B_{\lambda\rho\cdot\beta}\mathbf{A}_1^{-1}B_{\rho\lambda\cdot\beta}B_{\lambda\cdot\beta}^{-1} \end{bmatrix} \\
&= \begin{bmatrix} B_{\rho\cdot\beta}^{-1} + B_{\rho\cdot\beta}^{-1}B_{\rho\lambda\cdot\beta}\mathbf{A}_2^{-1}B_{\lambda\rho\cdot\beta}B_{\rho\cdot\beta}^{-1} & -B_{\rho\cdot\beta}B_{\rho\lambda\cdot\beta}^{-1}\mathbf{A}_2^{-1} \\ -\mathbf{A}_2^{-1}B_{\lambda\rho\cdot\beta}B_{\rho\cdot\beta}^{-1} & \mathbf{A}_2^{-1} \end{bmatrix},
\end{aligned}
\tag{6.1}
$$

where $\mathbf{A}_1 = \Big[B_{\rho\cdot\beta} - B_{\rho\lambda\cdot\beta}B_{\lambda\cdot\beta}^{-1}B_{\lambda\rho\cdot\beta}\Big]$ and $\mathbf{A}_2 = \Big[B_{\lambda\cdot\beta} - B_{\lambda\rho\cdot\beta}B_{\rho\cdot\beta}^{-1}B_{\rho\lambda\cdot\beta}\Big]$. A similar result can be obtained for $\Big[\mathbf{I}_{1\cdot3}(\tilde\theta)\Big]^{-1}$. These results can be used to establish relationships between the test statistics as shown in the next corollary.

Corollary 1. — In the GMM framework, we have the following relations.

$$\mathrm{LM}^{\mathrm{g}}_{\rho\lambda} = \mathrm{LM}^{\mathrm{g}}_{\lambda} + \mathrm{LM}^{\mathrm{g}\star}_{\rho} = \mathrm{LM}^{\mathrm{g}}_{\rho} + \mathrm{LM}^{\mathrm{g}\star}_{\lambda}. \tag{6.2}$$

Similarly, in the ML framework, the following relations hold.

$$\mathrm{LM}^{\mathrm{m}}_{\rho\lambda} = \mathrm{LM}^{\mathrm{m}}_{\lambda} + \mathrm{LM}^{\mathrm{m}\star}_{\rho} = \mathrm{LM}^{\mathrm{m}}_{\rho} + \mathrm{LM}^{\mathrm{m}\star}_{\lambda}. \tag{6.3}$$

Proof. See Appendix D. □

The results in (6.2) and (6.3) show that the robust tests can also be computed from the joint and the one directional tests.

7. Monte Carlo Simulations

To shed light on the performance of the proposed tests in finite samples, we conduct a Monte Carlo study based on two different data generating processes. Note that the computations of one directional and robust tests statistics require $\tilde{\theta}$, which is the OLS estimator when $\rho_0 = 0$ and $\lambda_0 = 0$ in the model. A summary of tests statistics is given in Tables 1 and 2. As indicated in these tables, all test statistics will be available when they are evaluated at $\tilde{\theta}$, except the conditional test statistics LM^{A}_{ρ}, $\mathrm{LM}^{A}_{\lambda}$ and LM^{A}_{ψ}. In Tables 1 and 2, the test statistic LM^{A}_{ρ} is for $\mathrm{H}_0 : \rho_0 = 0$ in the presence of λ_0, $\mathrm{LM}^{A}_{\lambda}$ is for $\mathrm{H}_0 : \lambda_0 = 0$ in the presence of ρ_0, and LM^{A}_{ψ} is for $\mathrm{H}_0 : \psi_0 = 0$ in the presence of ϕ_0. These test statistics can be calculated by using the general results in (3.9) and (4.8), and their computations require the estimation of the corresponding restricted models by the GMME and the MLE.

We consider two data generating processes:

$$\text{DGP 1:} \quad Y_r = S_r^{-1}X_{1r}\beta_{01} + S_r^{-1}W_rX_{2r}\beta_{02} + S_r^{-1}l_{m_r}\alpha_r + S_r^{-1}R_r^{-1}\varepsilon_r \tag{7.1}$$

$$\text{DGP 2:} \quad Y_r = S_r^{-1}X_{3r}\beta_{01} + S_r^{-1}W_rX_{3r}\beta_{02} + S_r^{-1}l_{m_r}\alpha_r + S_r^{-1}R_r^{-1}\varepsilon_r \tag{7.2}$$

In DGP 1, X_{1r} and X_{2r} are $m_r \times 1$ vectors of independent standard normal random variables with the associated coefficient vector $(\beta_{01}, \beta_{02})^{'} = (1.2, 0.6)^{'}$. In DGP 2, we use the U.S. county-level data set of Pace and Barry (1997) on the 1980 presidential election. More specifically, $X_{3r} = (X_{3r,1}, X_{3r,2})$, where $X_{3r,1}$ is the standardized value of log income per-capita and $X_{3r,2}$ is the standardized value of the homeownership rate. The data set describes 3107 U.S. counties, of which we use the first n observations in the Monte Carlo study. For the parameter values, we set $(\beta_{01}^{'}, \beta_{02}^{'})^{'} = (1.2, 0.6, -0.4, 0.1)^{'}$ in Model 2. For each group $r = 1, 2, \ldots, R$, α_r is a random draw from $N(0, 1)$. The disturbance terms ε_{ir}s are independently generated from two distributions: (i) $N(0, 1)$ and (ii) $\mathrm{Gamma}\big(1, 1\big) - 1$. The Gamma distribution generates disturbances with positive skewness and excess kurtosis.

For the interaction scenario, we consider an experiments where the number of groups is $R = 60$. We allow m_r to vary across R groups by randomly assigning a value from the set of integers $\{10, 11, \ldots, 15\}$ to each group size. The total number of observations n varies between 600 and 900. Following Liu and Lee (2010), the weight matrix W_r is generated in two steps. We first draw an integer value ϑ_{ir} uniformly from the set of integer values $\{1, 2, 3, 4\}$. Then, if $\vartheta_{ir} + i \leq m_r$, the $(i+1)$th, $\ldots$, $(i+\vartheta_{ir})$th elements of the ith row of W_r are set to one and the rest of the elements in the ith row are set to zero. On the other hand, if $\vartheta_{ir} + i > m_r$, the first $(\vartheta_{ir} + i - m_r)$ entries of the ith row are set to one and the others are set to zero. Then, W is generated as the row-normalized $\mathrm{D}\left(W_1, \ldots, W_R\right)$ and we let $M = W$.

For the size analysis of test statistics for endogenous effects and/or correlated effects in Table 1, we set $\lambda_0 = 0$ and $\rho_0 = 0$ in (7.1) and (7.2). Following Halleck Vega and Elhorst (2015), we refer to these models as the SLX models. For the power analysis of these test statistics, we consider three specifications for the alternative model. The first alternative is the spatial lag model ($\mathrm{SARAR}(1, 0)$) where we allow for spatial dependence in the dependent variable but not in the disturbance term, i.e., $\rho_0 = 0$. Note that $\mathrm{SARAR}(1, 0)$ specification can also be considered as a null model for LM_{ρ} statistics for testing $\mathrm{H}_0 : \rho_0 = 0$. The second alternative model is the spatial error model ($\mathrm{SARAR}(0, 1)$) which allows for spatial dependence in the disturbances but not in the dependent variable, i.e., $\lambda_0 = 0$. Similarly, $\mathrm{SARAR}(0, 1)$ can also be considered as another null model for the one-directional LM statistics for testing $\mathrm{H}_0 : \lambda_0 = 0$. Finally, the

Table 1: Summary of test statistics for spatial autoregressive parameters

GMM	Parameters		Test statistic
Null hypothesis	ρ_0	λ_0	
$\mathrm{H}_0 : \rho_0 = 0$	–	Set to zero	$\mathrm{LM}_{\rho}^{\mathrm{g}}$ in (3.14)
$\mathrm{H}_0 : \rho_0 = 0$	–	Unrestricted, estimated	$\mathrm{LM}_{\rho}^{\mathrm{g}A}$ in (3.9)
$\mathrm{H}_0 : \rho_0 = 0$	–	Unrestricted, not estimated	$\mathrm{LM}_{\rho}^{\mathrm{g}\star}$ in (4.17)
$\mathrm{H}_0 : \lambda_0 = 0$	Set to zero	–	$\mathrm{LM}_{\lambda}^{\mathrm{g}}$ in (3.21)
$\mathrm{H}_0 : \lambda_0 = 0$	Unrestricted, estimated	–	$\mathrm{LM}_{\lambda}^{\mathrm{g}A}$ in (3.9)
$\mathrm{H}_0 : \lambda_0 = 0$	Unrestricted, not estimated	–	$\mathrm{LM}_{\lambda}^{\mathrm{g}\star}$ in (4.24)
$\mathrm{H}_0 : \lambda_0 = 0,\ \rho_0 = 0$	–	–	$\mathrm{LM}_{\rho\lambda}^{\mathrm{g}}$ in (3.9)
ML			
$\mathrm{H}_0 : \rho_0 = 0$	–	Set to zero	$\mathrm{LM}_{\rho}^{\mathrm{m}}$ in (4.11)
$\mathrm{H}_0 : \rho_0 = 0$	–	Unrestricted, estimated	$\mathrm{LM}_{\rho}^{\mathrm{m}A}$ in (4.8)
$\mathrm{H}_0 : \rho_0 = 0$	–	Unrestricted, not estimated	$\mathrm{LM}_{\rho}^{\mathrm{m}\star}$ in (4.16)
$\mathrm{H}_0 : \lambda_0 = 0$	Set to zero	–	$\mathrm{LM}_{\lambda}^{\mathrm{m}}$ in (4.18)
$\mathrm{H}_0 : \lambda_0 = 0$	Unrestricted, estimated	–	$\mathrm{LM}_{\lambda}^{\mathrm{m}A}$ in (4.8)
$\mathrm{H}_0 : \lambda_0 = 0$	Unrestricted, not estimated	–	$\mathrm{LM}_{\lambda}^{\mathrm{m}\star}$ in (4.23)
$\mathrm{H}_0 : \lambda_0 = 0,\ \rho_0 = 0$	–	–	$\mathrm{LM}_{\rho\lambda}^{\mathrm{m}}$ in (4.10)

Table 2: Summary of test statistics for contextual effects

GMM	Parameters		Test statistic
Null hypothesis	ρ_0	λ_0	
$\mathrm{H}_0 : \beta_{02} = 0$	Set to zero	Set to zero	$\mathrm{LM}_{\psi}^{\mathrm{g}}$ in (5.1)
$\mathrm{H}_0 : \beta_{02} = 0$	Unrestricted, estimated	Unrestricted, estimated	$\mathrm{LM}_{\psi}^{\mathrm{g}A}$ in (3.9)
$\mathrm{H}_0 : \beta_{02} = 0$	Unrestricted, not estimated	Unrestricted, not estimated	$\mathrm{LM}_{\psi}^{\mathrm{g}\star}$ in (5.3)
ML			
$\mathrm{H}_0 : \beta_{02} = 0$	Set to zero	Set to zero	$\mathrm{LM}_{\psi}^{\mathrm{m}}$ in (5.5)
$\mathrm{H}_0 : \beta_{02} = 0$	Unrestricted, estimated	Unrestricted, estimated	$\mathrm{LM}_{\psi}^{\mathrm{m}A}$ in (4.8)
$\mathrm{H}_0 : \beta_{02} = 0$	Unrestricted, not estimated	Unrestricted, not estimated	$\mathrm{LM}_{\psi}^{\mathrm{m}\star}$ in (5.7)

third alternative model allows for both type of spatial dependence, namely SARAR$(1,1)$. In the relevant alternative models, we let spatial parameters λ_0 and ρ_0 take on values from 0.1 to 0.6 with an increment of 0.1.

In the case of tests for the contextual effects in Table 2, we only use DGP 2 to study the size and power properties of test statistics. For the size analysis, we set $\beta_{02} = 0_{2\times 1}$ and let λ_0 and ρ_0 vary between 0.1 to 0.6. For the power analysis, we set $\lambda_0 = 0.3$ and $\rho_0 = 0.2$, and let elements of β_{02} take on values from $\{-1, -0.5, 0.5, 1\}$. All Monte Carlo simulations are based on 1000 repetitions.

Finally, we need to specify the set of moment functions for the GMM approach. As we mentioned before, we are interested in the case where the number of instruments is kept fixed as the number of observations grows without a bound. Therefore, we choose a simple set of moment functions: $Q_{1r} = J_r\big(X_r,\, W_rX_r,\, W_r^2X_r\big)$, $U_{1r} = J_rW_rJ_r - \mathrm{tr}\big(J_rW_rJ_r\big)J_r/\mathrm{tr}\big(J_r\big)$ and $U_{2r} = J_rW_r^2J_r - \mathrm{tr}\big(J_rW_r^2J_r\big)J_r/\mathrm{tr}\big(J_r\big)$.

8. Results for Endogenous Effects and Correlated Effects

In this section, we investigate the finite sample properties of the test statistics for endogenous effects and correlated effects. In the following, we first evaluate the empirical rejection frequencies of each test under the null hypothesis, and then provide a power analysis for each test.

8.1. Results on Size Properties

To present simulation results on size properties, we use the P value discrepancy plots suggested in Davidson and MacKinnon (1998), which are based on the empirical distribution functions (edf) of p-values. Let τ be 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 $F(x)$ be the cumulative distribution function (cdf) of the asymptotic distribution of τ evaluated at x. 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 points denoted by x_i for $i = 1, \ldots, m$ from the interval (0, 1). Then an estimate of cdf of $p(\tau)$ is given by

$$\widehat{F}(x_i) = \frac{1}{\mathcal{R}} \sum_{j=1}^{\mathcal{R}} \mathbf{1}\big(p(\tau_j) \leq x_i\big). \tag{8.1}$$

As stated in Davidson and MacKinnon (1998), there is no decisive way to choose the sequence x_i from (0, 1). In practice, the main attention is typically paid to the Type-I errors which are set at levels smaller than or equal to 10%. We choose the following sequence and focus on levels smaller than or equal to 10%.

$$\{x_i\}_{i=1}^{m} = \{0.001 : 0.001 : 0.010 \quad 0.015 : 0.005 : 0.990 \quad 0.991 : 0.001 : 0.999\} \tag{8.2}$$

The P value discrepancy plot is defined as the plot of $\widehat{F}(x_i) - x_i$ against x_i under the assumption that the true data generating process is characterized by the null hypothesis. If $F(x)$ approximate to the finite sample distribution of τ well enough, then each $p(\tau_j)$ will have a uniform distribution over $(0, 1)$. Hence, the P value plot, obtained by a plot of $\widehat{F}(x_i)$ against x_i, should be close to the 45 degree line. Therefore, a P value discrepancy plot highlights the differences between the empirical distribution function and the 45 degree line. The discrepancies from the horizontal axis in a P value discrepancy plot suggest an empirical distribution that differs from the asymptotic distribution used to determine the critical values.

To asses 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. 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% point-wise confidence interval in a P value discrepancy plot. In the discrepancy plots, this interval will be represented by the red solid lines (for some examples, see Taspinar and Dogan (2016)).

To save space, the size results based on the SLX models will be presented through the P value discrepancy plots while the size results based on SARAR$(1, 0)$ or SARAR$(0, 1)$ are summarized in tables. When the null model is SARAR(1,0) or SARAR(0,1), we focus solely on the nominal size of 5% and provide size deviations at this level only.

The general observations on the size properties of tests from Figures 1-2 and Tables 3- 4 are listed below. For notational simplicity, if a superscript "g" or "m" is not specified for a test, it means that the observation made holds for both the GMM based test and the ML based test.

1. Figures 1 and 2 present the size properties of test statistics under $H_0 : \lambda_0 = \rho_0 = 0$. Figures 1 and 2 show that all LM tests based on GMM are generally over-sized regardless of the normality of the errors. In both figures, the maximum size distortion is always less than 0.03 and the size distortions generally lie inside the 95% point-wise confidence interval and therefore they are acceptable.
2. In Figure 1, $\text{LM}_{\rho}^{\text{g}A}$ and $\text{LM}_{\lambda}^{\text{g}A}$ are generally over-sized and their size discrepancies lie outside the 95% point-wise confidence interval. For example, for the nominal size of 5%, the actual rejection rate of $\text{LM}_{\rho}^{\text{g}A}$ is about 7%. In Figure 2, $\text{LM}_{\lambda}^{\text{g}A}$, $\text{LM}_{\lambda}^{\text{g}\star}$ and $\text{LM}_{\rho}^{\text{g}A}$ are over-sized especially in panel (a).
3. Figures 1 and 2 clearly indicate that the size distortions of all ML based tests generally lie inside the 95% point-wise confidence interval and are smaller compared to the GMM based tests. Surprisingly, the ML based tests perform in a similar fashion even when the errors are not normally distributed.

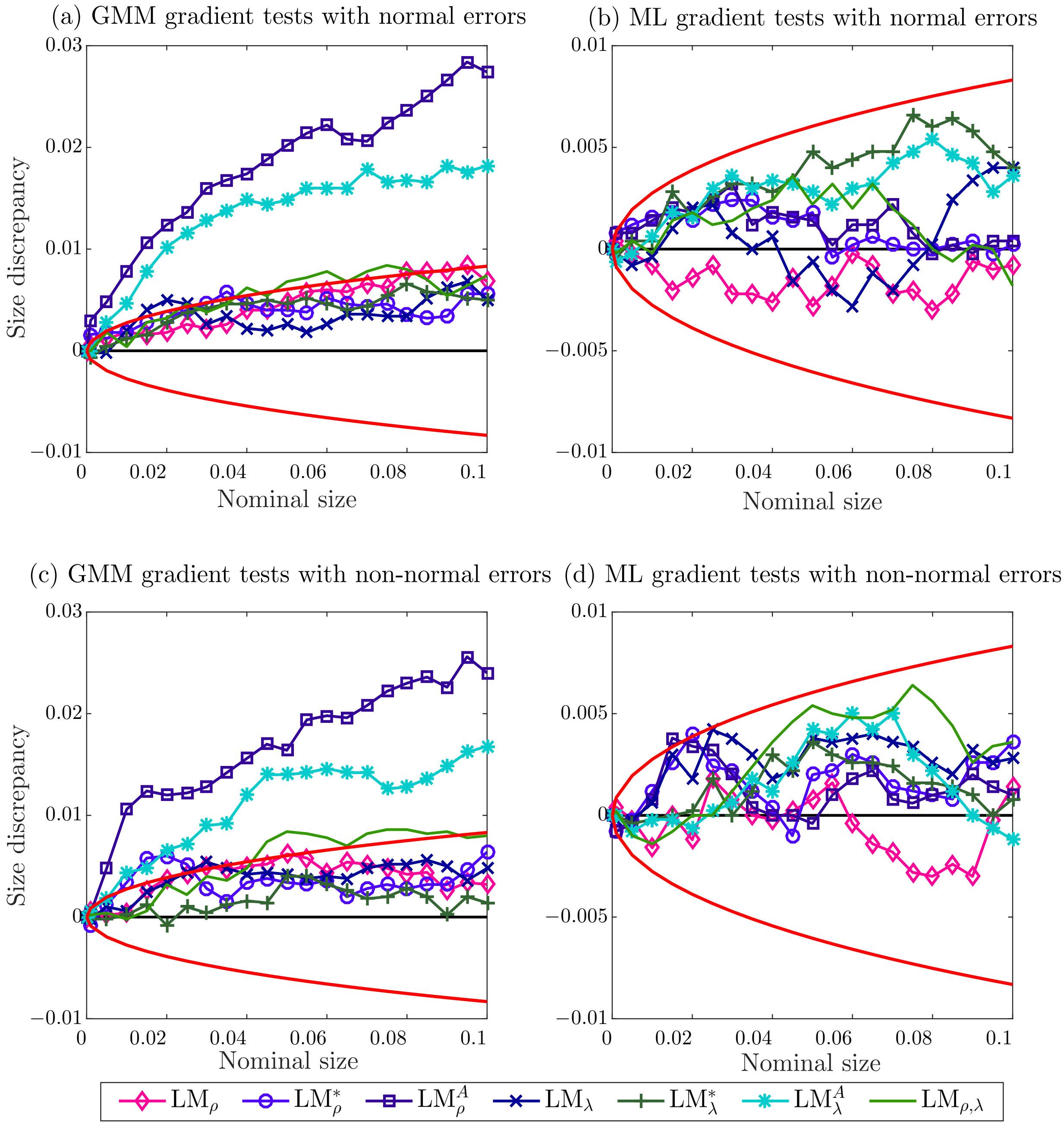

Figure 1: Size discrepancy plots under DGP 1

4. Table 3 and 4 provide some evidences on the magnitude of size distortions as a function of the size of local parametric misspecification in the alternative model. We would expect that the robust versions of one directional tests, $\mathrm{LM}_\rho^\star$ and $\mathrm{LM}_\lambda^\star$, to perform relatively better than LM_ρ and LM_λ, respectively. Overall, this seems to be the case, especially when the null model is SARAR$(0,1)$.
5. Tables 3 and 4 show that LM_ρ^A and LM_λ^A perform well in all cases. This is not surprising as these tests require the estimation of the spatial parameter λ_0 and ρ_0, respectively.
6. When the null model is SARAR (1,0) in Tables 3 and 4, $\mathrm{LM}_\rho^\star$ performs satisfactorily for small values of

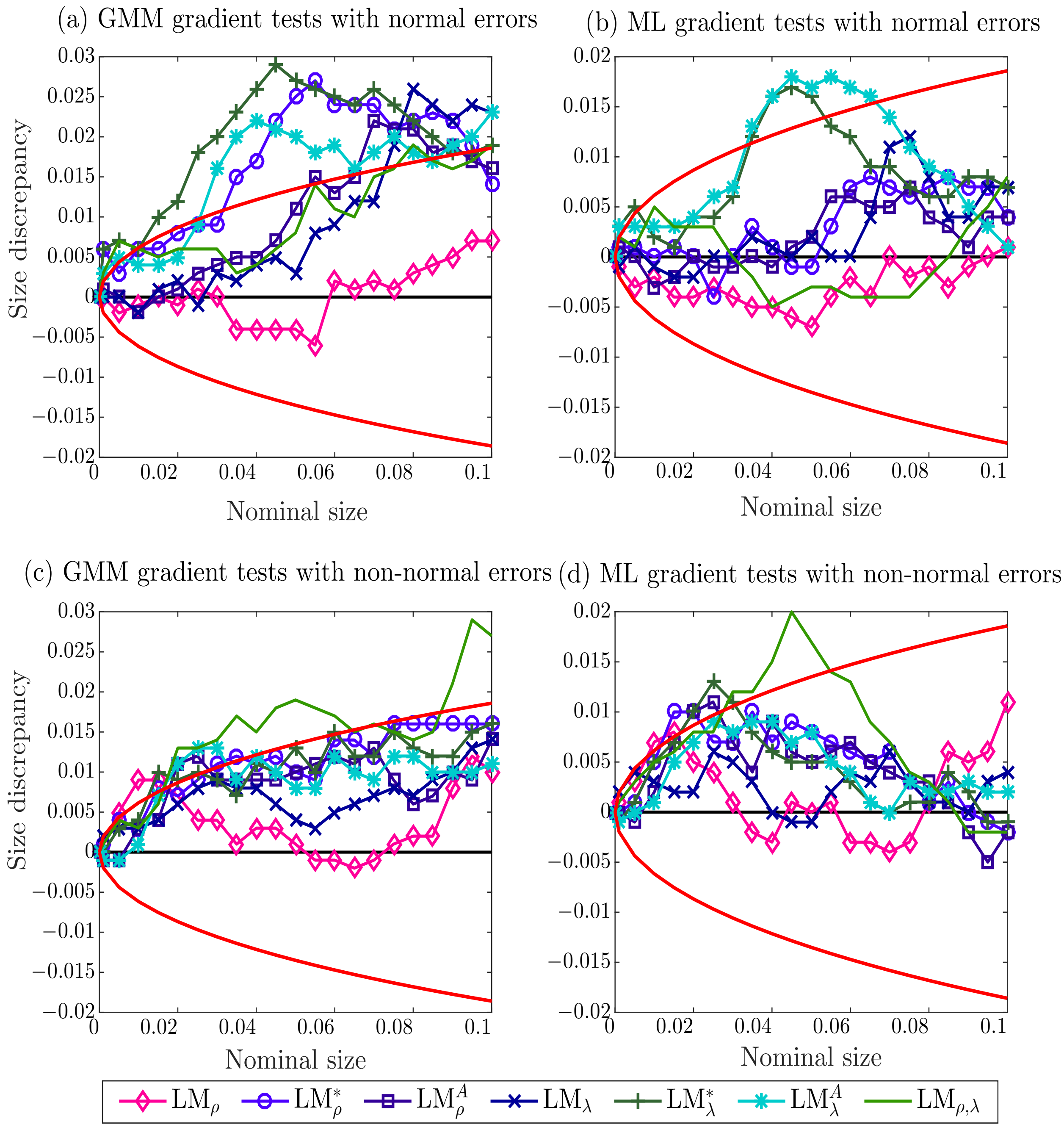

Figure 2: Size discrepancy plots under DGP 2

λ_0 in the alternative model. Indeed, when λ_0 is less than 0.3, $\mathrm{LM}_\rho^\star$ always performs better than LM_ρ. On the other hand, when the local misspecification deteriorates as λ_0 gets larger, $\mathrm{LM}_\rho^\star$ severely over rejects the null model, although still beats LM_ρ in all cases. Recall that $\mathrm{LM}_\rho^\star$ uses the least squares residuals from the transformed model and implements a correction on the test statistics for a local parametric misspecification of the alternative model, i.e., ignoring the spatial lag. The bias of the least squares residuals depends on the strength of spatial dependence as well as on the connectedness of the weights matrix. Therefore, we can expect poor performance for the robust tests as λ_0 deviates from

Table 3: Empirical size of tests at 5% level under DGP 1

H_0: SARAR$(1,0)$												
	Normal distribution			Gamma distribution			Normal distribution			Gamma distribution		
λ_0	$\mathrm{LM}^{\mathrm{g}}_{\rho}$	$\mathrm{LM}^{\mathrm{g}\star}_{\rho}$	$\mathrm{LM}^{\mathrm{g}A}_{\rho}$	$\mathrm{LM}^{\mathrm{g}}_{\rho}$	$\mathrm{LM}^{\mathrm{g}\star}_{\rho}$	$\mathrm{LM}^{\mathrm{g}A}_{\rho}$	$\mathrm{LM}^{\mathrm{m}}_{\rho}$	$\mathrm{LM}^{\mathrm{m}\star}_{\rho}$	$\mathrm{LM}^{\mathrm{m}A}_{\rho}$	$\mathrm{LM}^{\mathrm{m}}_{\rho}$	$\mathrm{LM}^{\mathrm{m}\star}_{\rho}$	$\mathrm{LM}^{\mathrm{m}A}_{\rho}$
0.1	0.455	0.055	0.068	0.441	0.055	0.072	0.415	0.050	0.051	0.410	0.054	0.056
0.2	0.946	0.061	0.070	0.937	0.057	0.067	0.936	0.053	0.051	0.929	0.052	0.050
0.3	1.000	0.088	0.070	0.999	0.099	0.065	1.000	0.069	0.049	0.999	0.073	0.048
0.4	1.000	0.188	0.073	1.000	0.201	0.068	1.000	0.103	0.049	1.000	0.120	0.049
0.5	1.000	0.392	0.065	1.000	0.406	0.069	1.000	0.166	0.043	1.000	0.186	0.048
0.6	1.000	0.655	0.069	1.000	0.646	0.065	1.000	0.244	0.050	1.000	0.258	0.046
H_0: SARAR$(0,1)$												
ρ_0	$\mathrm{LM}^{\mathrm{g}}_{\lambda}$	$\mathrm{LM}^{\mathrm{g}\star}_{\lambda}$	$\mathrm{LM}^{\mathrm{g}A}_{\lambda}$	$\mathrm{LM}^{\mathrm{g}}_{\lambda}$	$\mathrm{LM}^{\mathrm{g}\star}_{\lambda}$	$\mathrm{LM}^{\mathrm{g}A}_{\lambda}$	$\mathrm{LM}^{\mathrm{m}}_{\lambda}$	$\mathrm{LM}^{\mathrm{m}\star}_{\lambda}$	$\mathrm{LM}^{\mathrm{m}A}_{\lambda}$	$\mathrm{LM}^{\mathrm{m}}_{\lambda}$	$\mathrm{LM}^{\mathrm{m}\star}_{\lambda}$	$\mathrm{LM}^{\mathrm{m}A}_{\lambda}$
0.1	0.177	0.059	0.068	0.173	0.049	0.058	0.159	0.060	0.056	0.166	0.054	0.046
0.2	0.499	0.051	0.061	0.455	0.048	0.060	0.464	0.060	0.047	0.447	0.057	0.046
0.3	0.786	0.057	0.066	0.760	0.054	0.064	0.760	0.076	0.053	0.760	0.070	0.050
0.4	0.934	0.051	0.060	0.913	0.051	0.065	0.931	0.073	0.047	0.926	0.076	0.053
0.5	0.981	0.052	0.064	0.971	0.051	0.065	0.983	0.078	0.048	0.979	0.081	0.048
0.6	0.992	0.062	0.071	0.988	0.054	0.072	0.995	0.092	0.052	0.995	0.089	0.051

zero substantially. Note that when the null model is SARAR (1,0), $\mathrm{LM}^{\star}_{\rho}$ performs relatively better in Tables 3 than in Table 4. This results suggest that the performance of tests statistics should be investigated under realistic data generating processes.

7. In Table 3, $\mathrm{LM}^{\star}_{\lambda}$ perform satisfactorily regardless of the strength of spatial dependence in the alternative model and beats LM_{λ} in all cases. In In Table 3, $\mathrm{LM}^{\mathrm{g}\star}_{\lambda}$ performs generally better than $\mathrm{LM}^{\mathrm{g}A}_{\lambda}$, even though the latter requires the estimation of ρ_0. This result may seem surprising at first, but note that in this case the least squares residuals are still consistent under the parametric misspecification of the alternative model. The relative performances of $\mathrm{LM}^{\mathrm{m}\star}_{\lambda}$ and $\mathrm{LM}^{\mathrm{m}A}_{\lambda}$ are reversed, when we move from the GMM based tests to the ML based tests. That is, $\mathrm{LM}^{\mathrm{m}A}_{\lambda}$ performs relatively better than $\mathrm{LM}^{\mathrm{m}\star}_{\lambda}$ in most cases. In Table 4, the robust test based on the ML is performing relatively better than the one based on GMM. Also, the robust tests have relatively larger size distortions in Table 4 than in Table 3.

8.2. Results on Power Properties

The results on the power properties of tests are presented in Tables 5-8.[17] The general observations on the power properties of our proposed tests are listed as follows.

1. In Tables 5 and 6, the null model is the SLX model and the alternative model is either SARAR(1,0) or SARAR(0,1). When the alternative model is SARAR(1,0), the results in both tables indicate that all test statistics for λ_0 have satisfactory power. $\mathrm{LM}^{\star}_{\lambda}$ and $\mathrm{LM}^{A}_{\lambda}$ present very similar performance but the former has the computational advantage and is robust to local deviations of ρ_0 from zero. The power of tests are relatively slightly lower in Table 6.
2. In Tables 5 and 6, the test statistics for $H_0:\rho_0=0$ should lack of power when the alternative model is SARAR(1,0). The conditional test statistics, LM^{A}_{ρ} lack power in all cases. The robust test statistic, $\mathrm{LM}^{\star}_{\rho}$, lacks power when λ_0 locally deviate from zero. Both LM_{ρ} and $\mathrm{LM}_{\rho\lambda}$ have good powers against the positive spatial lag term in both tables. These results clearly show that the application of LM_{ρ} and $\mathrm{LM}_{\rho\lambda}$ can lead to the incorrect inference.

[17]For the sake of brevity, we only provide power results for the case where the disturbance terms are normally distributed. The results based on the gamma distribution are similar and available upon request.

Table 4: Empirical size of tests at 5% level under DGP 2

H_0: $\mathrm{SARAR}(1,0)$												
	Normal distribution			Gamma distribution			Normal distribution			Gamma distribution		
λ_0	$\mathrm{LM}_\rho^{\mathrm{g}}$	$\mathrm{LM}_\rho^{\mathrm{g}\star}$	$\mathrm{LM}_\rho^{\mathrm{g}A}$	$\mathrm{LM}_\rho^{\mathrm{g}}$	$\mathrm{LM}_\rho^{\mathrm{g}\star}$	$\mathrm{LM}_\rho^{\mathrm{g}A}$	$\mathrm{LM}_\rho^{\mathrm{m}}$	$\mathrm{LM}_\rho^{\mathrm{m}\star}$	$\mathrm{LM}_\rho^{\mathrm{m}A}$	$\mathrm{LM}_\rho^{\mathrm{m}}$	$\mathrm{LM}_\rho^{\mathrm{m}\star}$	$\mathrm{LM}_\rho^{\mathrm{m}A}$
0.1	0.372	0.059	0.053	0.358	0.069	0.059	0.347	0.061	0.050	0.346	0.059	0.057
0.2	0.873	0.200	0.067	0.899	0.170	0.041	0.865	0.075	0.056	0.890	0.060	0.042
0.3	0.992	0.354	0.059	0.995	0.366	0.060	0.992	0.054	0.050	0.995	0.068	0.066
0.4	1.000	0.497	0.064	1.000	0.500	0.060	1.000	0.063	0.049	1.000	0.066	0.063
0.5	1.000	0.590	0.063	1.000	0.549	0.071	1.000	0.067	0.058	1.000	0.075	0.063
0.6	1.000	0.711	0.071	1.000	0.671	0.063	1.000	0.077	0.046	1.000	0.099	0.052
H_0: $\mathrm{SARAR}(0,1)$												
ρ_0	$\mathrm{LM}_\lambda^{\mathrm{g}}$	$\mathrm{LM}_\lambda^{\mathrm{g}\star}$	$\mathrm{LM}_\lambda^{\mathrm{g}A}$	$\mathrm{LM}_\lambda^{\mathrm{g}}$	$\mathrm{LM}_\lambda^{\mathrm{g}\star}$	$\mathrm{LM}_\lambda^{\mathrm{g}A}$	$\mathrm{LM}_\lambda^{\mathrm{m}}$	$\mathrm{LM}_\lambda^{\mathrm{m}\star}$	$\mathrm{LM}_\lambda^{\mathrm{m}A}$	$\mathrm{LM}_\lambda^{\mathrm{m}}$	$\mathrm{LM}_\lambda^{\mathrm{m}\star}$	$\mathrm{LM}_\lambda^{\mathrm{m}A}$
0.1	0.360	0.043	0.069	0.349	0.046	0.049	0.332	0.065	0.065	0.348	0.051	0.044
0.2	0.851	0.065	0.068	0.865	0.061	0.058	0.839	0.073	0.054	0.866	0.061	0.059
0.3	0.992	0.111	0.048	0.990	0.095	0.058	0.991	0.062	0.043	0.989	0.073	0.055
0.4	1.000	0.191	0.045	1.000	0.184	0.052	1.000	0.055	0.044	1.000	0.070	0.063
0.5	1.000	0.282	0.067	1.000	0.257	0.053	1.000	0.086	0.047	1.000	0.070	0.061
0.6	1.000	0.357	0.065	1.000	0.342	0.064	1.000	0.089	0.056	1.000	0.088	0.064

3. There are similar findings in Tables 5 and 6 when the alternative model is SARAR(0,1). All one directional tests and the two directional tests for $H_0 : \rho_0 = 0$ have satisfactory power. In both tables, LM_ρ has more power than $\mathrm{LM}_\rho^\star$ and LM_ρ^A, and the difference in power levels get smaller when $\rho_0 = 0.2$ in the alternative model. In both tables, $\mathrm{LM}_{\rho\lambda}$ is indistinguishable from the one directional tests, but they cannot point the true alternative model and could lead to the wrong inference.
4. In Tables 5 and 6, when the alternative model is SARAR(0,1), the conditional test statistic, LM_λ^A, reports power around the nominal size level in all cases. The robust test statistics, $\mathrm{LM}_\lambda^\star$ indicate less powers only when ρ_0 locally deviates from zero, which is inline with our asymptotic results. Again, LM_λ and $\mathrm{LM}_{\lambda\rho}$ do not lack power and therefore can lead to incorrect inference.
5. In Tables 7 and 8, the alternative model is SARAR(1,1) and λ_0 and ρ_0 values vary from 0.1 to 0.6. Both GMM and ML based one directional test statistics relatively have higher power than the corresponding robust test statistics, especially when λ_0 and ρ_0 are close to zero. In all cases, $\mathrm{LM}_{\lambda\rho}$ has higher power and are indistinguishable from the one directional tests statistics LM_ρ and LM_λ. The conditional test statistics LM_ρ^A and LM_λ^A achieve higher power than the one-directional robust test in most cases, but not as much as the one-directional and two directional test statistics.

Table 5: Power of Tests at 5% level under DGP 1

	GMM							ML						
H_1: SARAR$(1,0)$														
λ_0	LM^{g}_{ρ}	$LM^{g\star}_{\rho}$	LM^{gA}_{ρ}	LM^{g}_{λ}	$LM^{g\star}_{\lambda}$	LM^{gA}_{λ}	$LM^{g}_{\rho\lambda}$	LM^{m}_{ρ}	$LM^{m\star}_{\rho}$	LM^{mA}_{ρ}	LM^{m}_{λ}	$LM^{m\star}_{\lambda}$	LM^{mA}_{λ}	$LM^{m}_{\rho\lambda}$
0.1	0.455	0.055	0.068	0.830	0.652	0.674	0.753	0.415	0.050	0.051	0.826	0.667	0.646	0.747
0.2	0.946	0.061	0.070	1.000	0.997	0.997	1.000	0.936	0.053	0.051	1.000	0.997	0.995	1.000
0.3	1.000	0.088	0.070	1.000	1.000	1.000	1.000	1.000	0.069	0.049	1.000	1.000	1.000	1.000
0.4	1.000	0.188	0.073	1.000	1.000	1.000	1.000	1.000	0.103	0.049	1.000	1.000	1.000	1.000
0.5	1.000	0.392	0.065	1.000	1.000	1.000	1.000	1.000	0.166	0.043	1.000	1.000	1.000	1.000
0.6	1.000	0.655	0.069	1.000	1.000	1.000	1.000	1.000	0.244	0.050	1.000	1.000	1.000	1.000
H_1: SARAR$(0,1)$														
ρ_0														
0.1	0.408	0.297	0.327	0.177	0.059	0.068	0.327	0.387	0.288	0.287	0.159	0.060	0.056	0.309
0.2	0.897	0.748	0.772	0.499	0.051	0.061	0.834	0.886	0.745	0.739	0.464	0.060	0.047	0.822
0.3	0.997	0.970	0.975	0.786	0.057	0.066	0.994	0.996	0.971	0.971	0.760	0.076	0.053	0.993
0.4	1.000	0.998	0.998	0.934	0.051	0.060	1.000	1.000	0.999	0.998	0.931	0.073	0.047	1.000
0.5	1.000	1.000	1.000	0.981	0.052	0.064	1.000	1.000	1.000	1.000	0.983	0.078	0.048	1.000
0.6	1.000	1.000	1.000	0.992	0.062	0.071	1.000	1.000	1.000	1.000	0.995	0.092	0.052	1.000

Table 6: Power of Tests at 5% level under DGP 2

	GMM							ML						
H_1: SARAR$(1,0)$														
λ_0	LM^{g}_{ρ}	$LM^{g\star}_{\rho}$	LM^{gA}_{ρ}	LM^{g}_{λ}	$LM^{g\star}_{\lambda}$	LM^{gA}_{λ}	$LM^{g}_{\rho\lambda}$	LM^{m}_{ρ}	$LM^{m\star}_{\rho}$	LM^{mA}_{ρ}	LM^{m}_{λ}	$LM^{m\star}_{\lambda}$	LM^{mA}_{λ}	$LM^{m}_{\rho\lambda}$
0.1	0.372	0.059	0.053	0.387	0.078	0.085	0.309	0.347	0.061	0.050	0.375	0.082	0.084	0.310
0.2	0.873	0.200	0.067	0.886	0.226	0.107	0.840	0.865	0.075	0.056	0.887	0.123	0.088	0.826
0.3	0.992	0.354	0.059	0.995	0.435	0.133	0.994	0.992	0.054	0.050	0.994	0.173	0.078	0.992
0.4	1.000	0.497	0.064	1.000	0.586	0.266	1.000	1.000	0.063	0.049	1.000	0.304	0.101	1.000
0.5	1.000	0.590	0.063	1.000	0.759	0.538	1.000	1.000	0.067	0.058	1.000	0.562	0.165	1.000
0.6	1.000	0.711	0.071	1.000	0.895	0.814	1.000	1.000	0.077	0.046	1.000	0.796	0.264	1.000
H_1: SARAR$(0,1)$														
ρ_0														
0.1	0.373	0.063	0.070	0.360	0.043	0.069	0.279	0.353	0.078	0.065	0.332	0.065	0.065	0.287
0.2	0.864	0.164	0.154	0.851	0.065	0.068	0.803	0.856	0.152	0.142	0.839	0.073	0.054	0.803
0.3	0.993	0.286	0.244	0.992	0.111	0.048	0.989	0.993	0.268	0.216	0.991	0.062	0.043	0.989
0.4	1.000	0.468	0.347	1.000	0.191	0.045	1.000	1.000	0.413	0.278	1.000	0.055	0.044	1.000
0.5	1.000	0.556	0.455	1.000	0.282	0.067	1.000	1.000	0.569	0.368	1.000	0.086	0.047	1.000
0.6	1.000	0.634	0.564	1.000	0.357	0.065	1.000	1.000	0.681	0.395	1.000	0.089	0.056	1.000

Table 7: Power of Tests at 5% level under DGP 1

H_1: SARAR$(1,1)$																
		GMM							ML							
λ_0	ρ_0	$\mathrm{LM}^{\mathrm{g}}_{\rho}$	$\mathrm{LM}^{\mathrm{g}\star}_{\rho}$	$\mathrm{LM}^{\mathrm{g}A}_{\rho}$	$\mathrm{LM}^{\mathrm{g}}_{\lambda}$	$\mathrm{LM}^{\mathrm{g}\star}_{\lambda}$	$\mathrm{LM}^{\mathrm{g}A}_{\lambda}$	$\mathrm{LM}^{\mathrm{g}}_{\rho\lambda}$	$\mathrm{LM}^{\mathrm{m}}_{\rho}$	$\mathrm{LM}^{\mathrm{m}\star}_{\rho}$	$\mathrm{LM}^{\mathrm{m}A}_{\rho}$	$\mathrm{LM}^{\mathrm{m}}_{\lambda}$	$\mathrm{LM}^{\mathrm{m}\star}_{\lambda}$	$\mathrm{LM}^{\mathrm{m}A}_{\lambda}$	$\mathrm{LM}^{\mathrm{m}}_{\rho\lambda}$
0.1	0.1	0.922	0.257	0.331	0.967	0.644	0.662	0.962	0.914	0.299	0.289	0.966	0.666	0.635	0.961
0.1	0.2	0.998	0.685	0.786	0.996	0.645	0.678	0.999	0.998	0.755	0.749	0.996	0.675	0.645	0.999
0.1	0.3	1.000	0.947	0.976	0.999	0.623	0.681	1.000	1.000	0.974	0.970	1.000	0.663	0.644	1.000
0.1	0.4	1.000	0.997	0.999	1.000	0.614	0.694	1.000	1.000	0.999	0.999	1.000	0.662	0.660	1.000
0.1	0.5	1.000	1.000	1.000	1.000	0.605	0.714	1.000	1.000	1.000	1.000	1.000	0.657	0.679	1.000
0.1	0.6	1.000	1.000	1.000	1.000	0.581	0.752	1.000	1.000	1.000	1.000	1.000	0.641	0.718	1.000
0.2	0.1	0.998	0.149	0.322	1.000	0.996	0.996	1.000	0.998	0.321	0.279	1.000	0.996	0.994	1.000
0.2	0.2	1.000	0.510	0.798	1.000	0.996	0.997	1.000	1.000	0.786	0.765	1.000	0.996	0.995	1.000
0.2	0.3	1.000	0.850	0.982	1.000	0.992	0.995	1.000	1.000	0.977	0.976	1.000	0.992	0.994	1.000
0.2	0.4	1.000	0.969	0.999	1.000	0.992	0.997	1.000	1.000	0.998	0.998	1.000	0.992	0.995	1.000
0.2	0.5	1.000	0.995	1.000	1.000	0.989	0.998	1.000	1.000	1.000	1.000	1.000	0.988	0.997	1.000
0.2	0.6	1.000	1.000	1.000	1.000	0.986	0.999	1.000	1.000	1.000	1.000	1.000	0.986	0.998	1.000
0.3	0.1	1.000	0.062	0.340	1.000	1.000	1.000	1.000	1.000	0.382	0.294	1.000	1.000	1.000	1.000
0.3	0.2	1.000	0.243	0.816	1.000	1.000	1.000	1.000	1.000	0.810	0.785	1.000	1.000	1.000	1.000
0.3	0.3	1.000	0.565	0.980	1.000	1.000	1.000	1.000	1.000	0.971	0.976	1.000	1.000	1.000	1.000
0.3	0.4	1.000	0.820	0.998	1.000	1.000	1.000	1.000	1.000	0.996	0.998	1.000	1.000	1.000	1.000
0.3	0.5	1.000	0.930	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.3	0.6	1.000	0.980	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.4	0.1	1.000	0.045	0.359	1.000	1.000	1.000	1.000	1.000	0.464	0.316	1.000	1.000	1.000	1.000
0.4	0.2	1.000	0.062	0.828	1.000	1.000	1.000	1.000	1.000	0.828	0.796	1.000	1.000	1.000	1.000
0.4	0.3	1.000	0.205	0.986	1.000	1.000	1.000	1.000	1.000	0.973	0.981	1.000	1.000	1.000	1.000
0.4	0.4	1.000	0.442	1.000	1.000	1.000	1.000	1.000	1.000	0.998	1.000	1.000	1.000	1.000	1.000
0.4	0.5	1.000	0.639	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.4	0.6	1.000	0.789	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.5	0.1	1.000	0.158	0.360	1.000	1.000	1.000	1.000	1.000	0.536	0.312	1.000	1.000	1.000	1.000
0.5	0.2	1.000	0.063	0.849	1.000	1.000	1.000	1.000	1.000	0.855	0.820	1.000	1.000	1.000	1.000
0.5	0.3	1.000	0.044	0.990	1.000	1.000	1.000	1.000	1.000	0.976	0.985	1.000	1.000	1.000	1.000
0.5	0.4	1.000	0.098	0.999	1.000	1.000	1.000	1.000	1.000	0.997	0.999	1.000	1.000	1.000	1.000
0.5	0.5	1.000	0.202	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.5	0.6	1.000	0.329	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.6	0.1	1.000	0.434	0.365	1.000	1.000	1.000	1.000	1.000	0.577	0.320	1.000	1.000	1.000	1.000
0.6	0.2	1.000	0.237	0.857	1.000	1.000	1.000	1.000	1.000	0.856	0.824	1.000	1.000	1.000	1.000
0.6	0.3	1.000	0.106	0.991	1.000	1.000	1.000	1.000	1.000	0.970	0.988	1.000	1.000	1.000	1.000
0.6	0.4	1.000	0.056	1.000	1.000	1.000	1.000	1.000	1.000	0.994	1.000	1.000	1.000	1.000	1.000
0.6	0.5	1.000	0.038	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000
0.6	0.6	1.000	0.060	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 8: Power of Tests at 5% level under DGP 2

H_1: SARAR(1,1)		GMM							ML						
λ_0	ρ_0	LM^g_ρ	$LM^{g\star}_\rho$	LM^{gA}_ρ	LM^g_λ	$LM^{g\star}_\lambda$	LM^{gA}_λ	$LM^g_{\rho\lambda}$	LM^m_ρ	$LM^{m\star}_\rho$	LM^{mA}_ρ	LM^m_λ	$LM^{m\star}_\lambda$	LM^{mA}_λ	$LM^m_{\rho\lambda}$
0.1	0.1	0.913	0.140	0.082	0.914	0.110	0.073	0.866	0.907	0.086	0.069	0.909	0.082	0.071	0.855
0.1	0.2	0.999	0.294	0.138	1.000	0.231	0.072	0.996	0.999	0.131	0.120	1.000	0.080	0.078	0.996
0.1	0.3	1.000	0.426	0.231	1.000	0.332	0.079	1.000	1.000	0.228	0.193	1.000	0.104	0.086	1.000
0.1	0.4	1.000	0.575	0.304	1.000	0.439	0.087	1.000	1.000	0.331	0.253	1.000	0.098	0.096	1.000
0.1	0.5	1.000	0.657	0.386	1.000	0.546	0.118	1.000	1.000	0.404	0.323	1.000	0.125	0.151	1.000
0.1	0.6	1.000	0.705	0.510	1.000	0.591	0.157	1.000	1.000	0.507	0.401	1.000	0.132	0.199	1.000
0.2	0.1	0.995	0.375	0.073	0.996	0.365	0.100	0.994	0.994	0.074	0.061	0.997	0.123	0.090	0.996
0.2	0.2	1.000	0.503	0.151	1.000	0.504	0.103	1.000	1.000	0.102	0.129	1.000	0.140	0.128	1.000
0.2	0.3	1.000	0.600	0.245	1.000	0.591	0.113	1.000	1.000	0.168	0.224	1.000	0.134	0.184	1.000
0.2	0.4	1.000	0.689	0.355	1.000	0.654	0.143	1.000	1.000	0.221	0.329	1.000	0.155	0.264	1.000
0.2	0.5	1.000	0.738	0.515	1.000	0.716	0.252	1.000	1.000	0.310	0.528	1.000	0.158	0.429	1.000
0.2	0.6	1.000	0.762	0.690	1.000	0.747	0.430	1.000	1.000	0.340	0.667	1.000	0.169	0.576	1.000
0.3	0.1	1.000	0.531	0.090	1.000	0.568	0.140	1.000	1.000	0.072	0.094	1.000	0.186	0.141	1.000
0.3	0.2	1.000	0.617	0.189	1.000	0.641	0.153	1.000	1.000	0.100	0.180	1.000	0.188	0.193	1.000
0.3	0.3	1.000	0.675	0.342	1.000	0.703	0.202	1.000	1.000	0.142	0.355	1.000	0.211	0.340	1.000
0.3	0.4	1.000	0.716	0.538	1.000	0.733	0.344	1.000	1.000	0.198	0.568	1.000	0.223	0.539	1.000
0.3	0.5	1.000	0.753	0.757	1.000	0.780	0.594	1.000	1.000	0.254	0.803	1.000	0.197	0.763	1.000
0.3	0.6	1.000	0.790	0.880	1.000	0.830	0.831	1.000	1.000	0.315	0.916	1.000	0.259	0.903	1.000
0.4	0.1	1.000	0.609	0.127	1.000	0.712	0.252	1.000	1.000	0.084	0.102	1.000	0.326	0.173	1.000
0.4	0.2	1.000	0.683	0.316	1.000	0.746	0.357	1.000	1.000	0.115	0.314	1.000	0.356	0.410	1.000
0.4	0.3	1.000	0.710	0.549	1.000	0.786	0.499	1.000	1.000	0.166	0.577	1.000	0.367	0.614	1.000
0.4	0.4	1.000	0.745	0.817	1.000	0.786	0.712	1.000	1.000	0.214	0.835	1.000	0.330	0.833	1.000
0.4	0.5	1.000	0.798	0.946	1.000	0.823	0.917	1.000	1.000	0.267	0.951	1.000	0.344	0.951	1.000
0.4	0.6	1.000	0.798	0.987	1.000	0.835	0.980	1.000	1.000	0.363	0.997	1.000	0.320	0.997	1.000
0.5	0.1	1.000	0.672	0.168	1.000	0.827	0.553	1.000	1.000	0.083	0.150	1.000	0.558	0.313	1.000
0.5	0.2	1.000	0.737	0.449	1.000	0.851	0.674	1.000	1.000	0.106	0.459	1.000	0.578	0.615	1.000
0.5	0.3	1.000	0.768	0.785	1.000	0.857	0.812	1.000	1.000	0.198	0.811	1.000	0.576	0.860	1.000
0.5	0.4	1.000	0.742	0.957	1.000	0.830	0.966	1.000	1.000	0.267	0.964	1.000	0.561	0.977	1.000
0.5	0.5	1.000	0.795	0.990	1.000	0.857	0.991	1.000	1.000	0.350	0.996	1.000	0.532	0.996	1.000
0.5	0.6	1.000	0.814	0.999	1.000	0.863	0.999	1.000	1.000	0.346	1.000	1.000	0.577	1.000	1.000
0.6	0.1	1.000	0.754	0.220	1.000	0.914	0.873	1.000	1.000	0.069	0.196	1.000	0.844	0.503	1.000
0.6	0.2	1.000	0.794	0.638	1.000	0.919	0.920	1.000	1.000	0.106	0.652	1.000	0.827	0.836	1.000
0.6	0.3	1.000	0.785	0.918	1.000	0.902	0.981	1.000	1.000	0.210	0.925	1.000	0.795	0.971	1.000
0.6	0.4	1.000	0.816	0.995	1.000	0.908	0.999	1.000	1.000	0.267	0.991	1.000	0.803	0.998	1.000
0.6	0.5	1.000	0.813	0.999	1.000	0.895	1.000	1.000	1.000	0.304	0.998	1.000	0.784	1.000	1.000
0.6	0.6	1.000	0.816	1.000	1.000	0.907	1.000	1.000	1.000	0.384	1.000	1.000	0.790	1.000	1.000

9. Results for Contextual Effects

In this section, we investigate the size and power properties of test statistics for the contextual effects. We consider the following test statistics: (i) the robust test statistics $\mathrm{LM}_{\psi}^{\star}$ of Proposition 5 and 6, (ii) the conditional test statistics in (3.9) and (4.8), and (iii) the F-statistic. The computation of $\mathrm{LM}_{\psi}^{\star}$ is based on the OLS estimator of $Y_r = X_{3r}\beta_{01} + l_{m_r}\alpha_r + \varepsilon_r$, while the computation of conditional test LM_{ψ}^{A} is based on the restricted ML estimation of $Y_r = S_r^{-1}X_{3r}\beta_{01} + S_r^{-1}l_{m_r}\alpha_r + S_r^{-1}R_r^{-1}\varepsilon_r$. Hence, the conditional test statistics require the estimation of both λ_0 and ρ_0. To compute the F-statistic, estimations of the restricted model $Y_r = X_{3r}\beta_{01} + l_{m_r}\alpha_r + \varepsilon_r$ and the unrestricted model $Y_r = X_{3r}\beta_{01} + W_r X_{3r}\beta_{02} + l_{m_r}\alpha_r + \varepsilon_r$ are needed. It is clear that the robust test statistic has computational advantage as it only requires a single OLS estimation. Note that we use only DGP 2 to investigate the size and power properties. Here, the hypothesis of interest is $\mathrm{H}_0 : \psi_0 = 0_{2\times 1}$, where $\psi_0 = \beta_{02}$. To investigate power properties, we vary the values of β_{02} between -1 to 1 in the alternative model $Y_r = S_r^{-1}X_{3r}\beta_{01} + S_r^{-1}W_r X_{3r}\beta_{02} + S_r^{-1}l_{m_r}\alpha_r + S_r^{-1}R_r^{-1}\varepsilon_r$, and set $\lambda_0 = 0.3$ and $\rho_0 = 0.2$. The results are presented in Tables 9 and 10. The main observations from these results are listed as follows.

1. The size properties are presented in Table 9. The conditional test statistic LM_{ψ}^{A} has proper sizes in all cases. The F-statistic is always over-sized and only report small size distortions in the first block of Table 9, where $\lambda_0 = 0$. In all other cases, it reports very large size distortions.
2. The robust test statistic is under-sized when λ_0 locally deviates from zero. As λ_0 get larger, the size distortions of robust test get larger. The presence of spatial lag dependence in the true data generating process relatively has more distorting effects on the size performance of the robust LM tests and the F test.
3. Overall, the robust test statistic outperforms the F-statistic in terms of size distortions. The performance of all test statistics seem to be not affected by the distribution of disturbance terms.
4. All test statistics have satisfactory power levels except for some negative combinations $\beta_{02,1}$ and $\beta_{02,2}$.
5. As expected, the robust test statistic has relatively lower power than other test statistics. The power of $\mathrm{LM}_{\psi}^{\star}$ increases asymmetrically as $\beta_{02,1}$ moves away from zero, and increases faster on the positive side.

Table 9: Size of Tests at 5% level: H_0: SARAR$(1,1)$

		Normal					Gamma				
		GMM		ML			GMM		ML		
λ_0	ρ_0	$\mathrm{LM}_{\psi}^{\mathrm{g}\star}$	$\mathrm{LM}_{\psi}^{\mathrm{g}A}$	$\mathrm{LM}_{\psi}^{\mathrm{m}\star}$	$\mathrm{LM}_{\psi}^{\mathrm{m}A}$	F	$\mathrm{LM}_{\psi}^{\mathrm{g}\star}$	$\mathrm{LM}_{\psi}^{\mathrm{g}A}$	$\mathrm{LM}_{\psi}^{\mathrm{m}\star}$	$\mathrm{LM}_{\psi}^{\mathrm{m}A}$	F
0.00	0.00	0.014	0.042	0.013	0.051	0.078	0.017	0.029	0.015	0.046	0.057
0.00	0.05	0.022	0.043	0.018	0.056	0.071	0.020	0.044	0.019	0.064	0.075
0.00	0.10	0.013	0.039	0.012	0.053	0.068	0.018	0.034	0.016	0.062	0.067
0.00	0.15	0.019	0.043	0.020	0.061	0.075	0.020	0.048	0.021	0.076	0.082
0.00	0.20	0.021	0.033	0.021	0.038	0.078	0.021	0.040	0.020	0.060	0.083
0.00	0.25	0.022	0.034	0.019	0.054	0.079	0.025	0.036	0.027	0.059	0.096
0.00	0.30	0.027	0.041	0.027	0.047	0.082	0.028	0.046	0.028	0.061	0.090
0.05	0.00	0.015	0.051	0.015	0.053	0.144	0.014	0.031	0.013	0.059	0.133
0.05	0.05	0.015	0.034	0.014	0.054	0.141	0.019	0.047	0.022	0.068	0.148
0.05	0.10	0.014	0.036	0.013	0.050	0.158	0.016	0.043	0.016	0.049	0.174
0.05	0.15	0.021	0.045	0.022	0.047	0.171	0.022	0.035	0.025	0.059	0.155
0.05	0.20	0.027	0.044	0.032	0.052	0.167	0.017	0.047	0.021	0.061	0.175
0.05	0.25	0.018	0.056	0.019	0.052	0.185	0.020	0.051	0.017	0.063	0.170
0.05	0.30	0.019	0.041	0.025	0.046	0.204	0.025	0.035	0.028	0.063	0.167
0.10	0.00	0.020	0.050	0.019	0.060	0.402	0.019	0.050	0.019	0.052	0.419
0.10	0.05	0.016	0.037	0.018	0.043	0.418	0.020	0.037	0.018	0.055	0.456
0.10	0.10	0.012	0.052	0.013	0.055	0.405	0.021	0.055	0.023	0.061	0.431
0.10	0.15	0.027	0.049	0.026	0.053	0.400	0.020	0.047	0.021	0.050	0.433
0.10	0.20	0.020	0.042	0.025	0.042	0.417	0.021	0.042	0.017	0.043	0.449
0.10	0.25	0.023	0.042	0.032	0.040	0.410	0.021	0.034	0.028	0.048	0.452
0.10	0.30	0.019	0.045	0.027	0.045	0.472	0.024	0.041	0.032	0.067	0.455
0.15	0.00	0.028	0.047	0.026	0.052	0.722	0.028	0.043	0.028	0.054	0.726
0.15	0.05	0.019	0.041	0.021	0.049	0.722	0.021	0.045	0.021	0.049	0.734
0.15	0.10	0.023	0.041	0.024	0.050	0.725	0.031	0.040	0.030	0.056	0.725
0.15	0.15	0.034	0.046	0.038	0.046	0.732	0.016	0.052	0.019	0.056	0.729
0.15	0.20	0.037	0.043	0.048	0.045	0.735	0.019	0.041	0.027	0.035	0.757
0.15	0.25	0.029	0.043	0.036	0.047	0.736	0.029	0.059	0.036	0.072	0.739
0.15	0.30	0.030	0.053	0.033	0.054	0.749	0.029	0.045	0.032	0.056	0.750
0.20	0.00	0.033	0.067	0.034	0.064	0.929	0.047	0.056	0.050	0.064	0.947
0.20	0.05	0.028	0.053	0.033	0.055	0.941	0.035	0.048	0.030	0.053	0.922
0.20	0.10	0.030	0.041	0.029	0.035	0.907	0.039	0.056	0.043	0.053	0.920
0.20	0.15	0.039	0.047	0.041	0.044	0.926	0.027	0.045	0.032	0.058	0.929
0.20	0.20	0.033	0.062	0.040	0.055	0.922	0.035	0.047	0.041	0.053	0.928
0.20	0.25	0.040	0.044	0.041	0.041	0.906	0.045	0.051	0.053	0.051	0.931
0.20	0.30	0.043	0.045	0.050	0.032	0.909	0.044	0.045	0.045	0.047	0.913
0.25	0.00	0.065	0.051	0.066	0.049	0.993	0.060	0.051	0.060	0.043	0.987
0.25	0.05	0.053	0.044	0.057	0.037	0.988	0.079	0.072	0.074	0.055	0.987
0.25	0.10	0.066	0.059	0.067	0.054	0.989	0.052	0.056	0.058	0.057	0.992
0.25	0.15	0.046	0.057	0.057	0.050	0.994	0.064	0.069	0.068	0.057	0.987
0.25	0.20	0.055	0.061	0.068	0.044	0.987	0.070	0.039	0.069	0.046	0.981
0.25	0.25	0.065	0.057	0.076	0.049	0.989	0.077	0.051	0.084	0.052	0.987
0.25	0.30	0.077	0.043	0.098	0.040	0.986	0.081	0.050	0.093	0.051	0.976
0.30	0.00	0.110	0.057	0.127	0.048	0.999	0.099	0.064	0.106	0.053	1.000
0.30	0.05	0.122	0.052	0.134	0.043	0.998	0.127	0.065	0.130	0.051	0.999
0.30	0.10	0.125	0.054	0.136	0.056	0.998	0.138	0.076	0.146	0.070	1.000
0.30	0.15	0.135	0.066	0.143	0.056	1.000	0.127	0.048	0.148	0.041	0.996
0.30	0.20	0.118	0.066	0.122	0.051	0.999	0.140	0.062	0.138	0.055	0.998
0.30	0.25	0.125	0.049	0.142	0.041	0.999	0.138	0.055	0.138	0.050	0.998
0.30	0.30	0.148	0.061	0.166	0.061	0.998	0.156	0.037	0.157	0.040	0.998

Table 10: Power of Tests at 5% level: H_1: SARAR$(1,1)$

		Normal					Gamma				
		GMM		ML			GMM		ML		
$\beta_{02,1}$	$\beta_{02,2}$	$\mathrm{LM}_{\psi}^{\mathrm{g}\star}$	$\mathrm{LM}_{\psi}^{\mathrm{g}A}$	$\mathrm{LM}_{\psi}^{\mathrm{m}\star}$	$\mathrm{LM}_{\psi}^{\mathrm{m}A}$	F	$\mathrm{LM}_{\psi}^{\mathrm{g}\star}$	$\mathrm{LM}_{\psi}^{\mathrm{g}A}$	$\mathrm{LM}_{\psi}^{\mathrm{m}\star}$	$\mathrm{LM}_{\psi}^{\mathrm{m}A}$	F
-1.0	-1.0	0.603	1.000	0.572	1.000	1.000	0.629	1.000	0.602	1.000	1.000
-1.0	-0.5	0.435	1.000	0.419	1.000	1.000	0.469	1.000	0.470	1.000	1.000
-1.0	0.5	0.838	1.000	0.860	1.000	1.000	0.847	1.000	0.865	1.000	1.000
-1.0	1.0	0.960	1.000	0.976	1.000	1.000	0.956	1.000	0.987	1.000	1.000
-0.5	-1.0	0.390	1.000	0.369	0.999	1.000	0.401	1.000	0.373	0.998	1.000
-0.5	-0.5	0.088	0.913	0.102	0.902	0.971	0.102	0.890	0.116	0.895	0.968
-0.5	0.5	0.620	1.000	0.645	1.000	1.000	0.669	1.000	0.680	1.000	1.000
-0.5	1.0	0.896	1.000	0.932	1.000	1.000	0.904	1.000	0.936	1.000	1.000
0.5	-1.0	0.712	1.000	0.692	1.000	1.000	0.732	1.000	0.726	1.000	1.000
0.5	-0.5	0.389	1.000	0.405	1.000	1.000	0.409	1.000	0.403	1.000	1.000
0.5	0.5	0.628	0.999	0.688	0.998	1.000	0.597	1.000	0.622	0.999	1.000
0.5	1.0	0.850	1.000	0.895	1.000	1.000	0.851	1.000	0.896	1.000	1.000
1.0	-1.0	0.935	1.000	0.926	1.000	1.000	0.928	1.000	0.923	1.000	1.000
1.0	-0.5	0.803	1.000	0.800	1.000	1.000	0.792	1.000	0.818	1.000	1.000
1.0	0.5	0.800	1.000	0.851	1.000	1.000	0.821	1.000	0.860	1.000	1.000
1.0	1.0	0.927	1.000	0.957	1.000	1.000	0.907	1.000	0.955	1.000	1.000

10. Conclusion

In this paper, we formulate robust LM tests within the GMM and the ML frameworks for a social interaction model with a network structure. These tests are robust in the sense that their null asymptotic distributions are still a central chi-square distribution when the alternative model has a local parametric misspecification. We show that the asymptotic null distribution of the standard LM test deviates from the central chi-square distribution when the alternative model is misspecified. Hence, the robust tests are size-resistant as they produce, asymptotically, correct size. Within the context of our social interaction model, we formally show 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 endogenous effects, the correlated effects, and the contextual effects in a social interaction model.

One attractive feature of our proposed tests is that their test statistics are easy to compute and only require the least squares estimates from a transformed linear regression model. Therefore, our proposed tests can easily be made available for practical applications using standard statistical software. In a Monte Carlo study, we investigate the size and power properties of our proposed tests. Our results show that the robust tests have good finite sample properties and can be useful for the detection of the source of dependence in a social interaction model. The Monte Carlo results show evidence for the analytical results that the robust tests are valid when the alternative model locally deviates from the true data generating process. Of course, more simulation work and empirical applications are needed to further confirm the finite sample properties of our suggested tests.

Appendices

Appendix A. Some Useful Lemmas

Lemma 1. — Let $\tilde{\theta} = \theta_0 + o_p(1)$ and $\tilde{\Omega}$ be a consistent estimate of Ω. Define $\bar{g}_2(\theta) = \frac{\mu_3}{\sigma_0^2}\omega' P_K \epsilon(\theta) - g_2(\theta)$. Then, when $\frac{K}{n} \to 0$, we have the following results:

1. $B(\tilde{\theta}) = \sigma_0^{-2}\,\mathrm{D}(0, \mathcal{X}(\rho_0)) + \lim_{n\to\infty} \frac{1}{n}\bar{D}_2' V_{22}\bar{D}_2 + o_p(1)$, where $\mathcal{X}(\rho_0) = \lim_{n\to\infty}\frac{1}{n} f'(\rho_0) f(\rho_0)$, $V_{22} = \left[\left(\mu_4 - 3\sigma_0^4\right)\omega'\omega + \sigma_0^4\Upsilon - \frac{\mu_3^2}{\sigma_0^2}\omega' P_K\omega\right]^{-1}$, and $\bar{D}_2 = D_2 - \frac{\mu_3}{\sigma_0^2}\left[0,\, \omega' f\right]$.
2. $-\frac{1}{\sqrt{n}} G'(\theta_0)\,\tilde{\Omega}^{-1} g(\theta_0) = \frac{1}{\sqrt{n}}\left[\mathrm{tr}\left(P_K M R^{-1}\right), \mathrm{tr}\left(P_K \bar{G}\right) e_1'\right]' + \frac{\sigma_0^2}{\sqrt{n}}\left[0,\, f'\epsilon\right] + \frac{1}{\sqrt{n}}\bar{D}_2' V_{22}\bar{g}_2(\theta_0) + o_p(1)$, where e_1 is the first unit column vector of dimension $k+1$.
3. $\frac{\sigma_0^2}{\sqrt{n}}\left[0,\, f'\epsilon\right] + \frac{1}{\sqrt{n}}\bar{D}_2' V_{22}\bar{g}_2(\theta_0) \xrightarrow{d} N\left[0, \sigma_0^2\,\mathrm{D}\left(0,\, \mathcal{X}(\rho_0)\right] + \lim_{n\to\infty}\frac{1}{n}\bar{D}_2' V_{22}\bar{D}_2\right]$, and $\frac{1}{\sqrt{n}}\left[\mathrm{tr}\left(P_K M R^{-1}\right), \mathrm{tr}\left(P_K\bar{G}\right) e_1'\right]' = O(\frac{K}{\sqrt{n}})$.

Proof. See Liu and Lee (2010, Propositions 4 & 5). □

Lemma 2. — Suppose that $W_r l_{m_r} = l_{m_r}$ and $M_r l_{m_r} = l_{m_r}$. Then,

1. $F_r' l_{m_r} = 0$, $F_r' F_r = I_{m_r - 1}$, and $F_r F_r' = J_r$.
2. $F_r' S(\lambda) = S_r^* F_r'$, $F_r' R_r W_r = R_r^* F_r' W_r = R_r^* W_r^* F_r'$, and $F_r' R_r Y_r = R_r^* F_r' Y_r = R_r^* Y_r^* F_r'$.
3. $|S^*(\lambda)| = |S_r(\lambda)|/(1-\lambda)$, and $|R^*(\rho)| = |R_r(\rho)|/(1-\rho)$.
4. $S^{*-1}(\lambda) = F_r' S^{-1}(\lambda) F_r$, $R^{*-1}(\rho) = F_r' R^{-1}(\rho) F_r$ and $G_r^*(\lambda) = S^{*-1}(\lambda) W_r^* = F_r' G(\lambda) F_r$.

Proof. See Lee et al. (2010, Lemma C.1). □

Lemma 3. — Suppose that $\tilde{\theta}$ is a consistent estimator of θ_0. Under Assumptions 5- 7, we have

1. $\sqrt{n^*} L(\theta_0) \xrightarrow{d} N\left[0, \lim_{n\to\infty}\Sigma\right]$, where $\Sigma = \mathrm{E}\left[-\frac{1}{n^*}\frac{\partial \ln L(\theta_0)}{\partial\theta\partial\theta'}\right]$ is stated in Appendix C.
2. $-L_{\theta\theta}(\tilde{\theta}) = \Sigma + o_p(1)$.

Proof. See Lee et al. (2010, Proposition 6.1). □

Appendix B. Detailed Expressions for GMM Gradient Tests

In this section, we provide explicit expressions for the components of test statistics. The variance matrix of $g(\theta_0)$ is

$$\Omega = \begin{bmatrix} \underbrace{\sigma_0^2 Q_K' Q_K}_{K\times K} & \underbrace{\mu_3 Q_K'\omega}_{K\times q} \\ \underbrace{\mu_3\omega' Q_K}_{q\times K} & \underbrace{(\mu_4 - 3\sigma_0^4)\omega'\omega + \sigma_0^4\Delta}_{q\times q} \end{bmatrix} \tag{B.1}$$

where $\omega = [\mathrm{vec}_D(T_1), \ldots, \mathrm{vec}_D(T_q)]$ and $\Delta = \frac{1}{2}\left[\mathrm{vec}(T_1^s), \ldots, \mathrm{vec}(T_q^s)\right]'\left[\mathrm{vec}(T_1^s), \ldots, \mathrm{vec}(T_q^s)\right]$. By the inverse of the partitioned matrix, we have

$$\Omega^{-1} = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix} \tag{B.2}$$

where

$$V_{11} = \sigma_0^{-2}\left(Q_K' Q_K\right)^{-1} + \left(\frac{\mu_3}{\sigma_0^2}\right)^2 \left(Q_K' Q_K\right)^{-1} Q_K'\omega V_{22}\omega' Q_K \left(Q_K' Q_K\right)^{-1}, \tag{B.3}$$

$$V_{21} = V_{12}' = -\frac{\mu_3}{\sigma_0^2} V_{22}\omega' Q_K\left(Q_K' Q_K\right)^{-1},\ V_{22} = \left[\left(\mu_4 - 3\sigma_0^4\right)\omega'\omega + \sigma_0^4\Delta - \frac{\mu_3^2}{\sigma_0^2}\omega' P_K\omega\right]^{-1}. \tag{B.4}$$

A consistent estimator of Ω can be obtained by replacing σ_0^2, μ_3 and μ_4 with their initial consistent counterparts.

The components of $C(\theta) = \frac{1}{n} G^{'}(\theta)\hat{\Omega}^{-1} g(\theta)$ can be explicitly stated in the following way:

$$\frac{\partial g(\theta)}{\partial \theta'} = -\left[\underbrace{G_\rho(\theta)}_{(q+K)\times 1} \quad \underbrace{G_\lambda(\theta)}_{(q+K)\times 1} \quad \underbrace{G_\beta(\theta)}_{(q+K)\times k} \right], \tag{B.5}$$

where

$$G_\rho(\theta) = \begin{bmatrix} Q_K^{'} M(Y - Z\delta) \\ \varepsilon_n^{'}(\theta) T_1^s M(Y - Z\delta) \\ \vdots \\ \varepsilon^{'}(\theta) T_q^s M(Y - Z\delta) \end{bmatrix}, \; G_\lambda(\theta) = \begin{bmatrix} Q_K^{'} R(\rho) W Y \\ \varepsilon^{'}(\theta) T_1^s R(\rho) W Y \\ \vdots \\ \varepsilon^{'}(\theta) T_q^s R(\rho) W Y \end{bmatrix}, \; G_\beta(\theta) = \begin{bmatrix} Q_K^{'} R(\rho) X \\ \varepsilon^{'}(\theta) T_1^s R(\rho) X \\ \vdots \\ \varepsilon^{'}(\theta) T_q^s R(\rho) X \end{bmatrix}.$$

Hence, the components of $C(\theta)$ can be determined as $C_j(\theta) = -\frac{1}{n} G_\rho^{'}(\theta) \begin{bmatrix} \hat{V}_{11} & \hat{V}_{12} \\ \hat{V}_{21} & \hat{V}_{22} \end{bmatrix} g(\theta)$ for $j \in \{\rho, \lambda, \beta\}$. Similarly, components of $B(\theta)$ can be determined as $B_{j,k}(\theta) = \frac{1}{n} G_j^{'}(\theta) \begin{bmatrix} \hat{V}_{11} & \hat{V}_{12} \\ \hat{V}_{21} & \hat{V}_{22} \end{bmatrix} G_k(\theta)$ for $j, k \in \{\rho, \lambda, \beta\}$. To calculate the relevant parts of the test statistics for spatial parameters, we simply evaluate $C(\theta)$ and $B(\theta)$ at $\tilde{\theta}$.

The relevant terms in the test statistics for the contextual effects can be determined from

$$G_\psi(\theta) = \begin{bmatrix} Q_K^{'} R(\rho) X_2 \\ \varepsilon^{'}(\theta) T_1^s R(\rho) X_2 \\ \vdots \\ \varepsilon^{'}(\theta) T_q^s R(\rho) X_2 \end{bmatrix}, \; G_\phi(\theta) = \begin{bmatrix} Q_K^{'} M(Y - Z\delta) & Q_K^{'} R(\rho) W Y \\ \varepsilon_n^{'}(\theta) T_1^s M(Y - Z\delta) & \varepsilon^{'}(\theta) T_1^s R(\rho) W Y \\ \vdots & \vdots \\ \varepsilon^{'}(\theta) T_q^s M(Y - Z\delta) & \varepsilon^{'}(\theta) T_q^s R(\rho) W Y \end{bmatrix}, \; G_\gamma(\theta) = \begin{bmatrix} Q_K^{'} R(\rho) X_1 \\ \varepsilon^{'}(\theta) T_1^s R(\rho) X_1 \\ \vdots \\ \varepsilon^{'}(\theta) T_q^s R(\rho) X_1 \end{bmatrix},$$

where $X_1 = \left(X_{11}^{'}, \ldots, X_{1R}^{'} \right)^{'}$ and $X_2 = \left(X_{21}^{'} W_1^{'}, \ldots, X_{2R}^{'} W_R^{'} \right)^{'}$. The components of $C(\theta)$ and $B(\theta)$ are calculated in a similar fashion as above for $j, k \in \{\psi, \phi, \gamma\}$.

Appendix C. Detailed Expressions for ML Tests

In this section, we state the explicit expressions for the relevant components of LM statistics. The first order derivatives of the log-likelihood function are given below.

$$1. \quad L_\rho(\theta) = \frac{1}{n^* \sigma^2} \left(\varepsilon^{'}(\theta) J H(\rho) \varepsilon(\theta) - \sigma^2 \mathrm{tr}(J H(\rho)) \right), \tag{C.1}$$

$$2. \quad L_\lambda(\theta) = \frac{1}{n^* \sigma^2} Y^{'} W^{'} R^{'}(\rho) J \varepsilon(\theta) - \frac{1}{n^*} \mathrm{tr}(J G(\lambda)), \; 3. \quad L_\gamma(\theta) = \begin{bmatrix} \frac{1}{n^* \sigma^2} \bar{X}^{'}(\rho) J \varepsilon(\theta) \\ \frac{1}{2 n^* \sigma^2} \left(\varepsilon^{'}(\theta) J \varepsilon(\theta) - n^* \sigma^2 \right) \end{bmatrix}, \tag{C.2}$$

where $H(\rho) = M R^{-1}(\rho)$, $\bar{X}(\rho) = R(\rho) X$. The components of $\Sigma(\theta)$ are given as

$$I_{\rho\rho}(\theta)=\frac{1}{n^*}\mathrm{tr}\left(H^s(\rho)JH(\rho)\right),\quad I_{\rho\lambda}(\theta)=\frac{1}{n^*}\mathrm{tr}\left(H^s(\rho)J\bar{G}(\lambda,\rho)\right),\tag{C.3}$$

$$I_{\rho\gamma}(\theta)=\left[0_{1\times k},\ \frac{1}{n^*\sigma^2}\mathrm{tr}\left(JH(\rho)\right)\right],\ I_{\lambda\rho}(\theta)=\frac{1}{n^*}\mathrm{tr}\left(H^s(\rho)J\bar{G}(\lambda,\rho)\right),\tag{C.4}$$

$$I_{\lambda\lambda}(\theta)=\frac{1}{n^*\sigma^2}\left(\bar{G}(\lambda,\rho)\bar{X}(\rho)\beta\right)^{'}J\left(\bar{G}(\lambda,\rho)\bar{X}(\rho)\beta\right)+\frac{1}{n^*}\mathrm{tr}\left(\bar{G}^s(\lambda,\rho)J\bar{G}(\lambda,\rho)\right),\tag{C.5}$$

$$I_{\lambda\gamma}(\theta)=\left[\frac{1}{n^*\sigma^2}\left(\bar{G}(\lambda,\rho)\bar{X}(\rho)\beta\right)^{'}J\bar{X}(\rho),\ \frac{1}{n^*\sigma^2}\mathrm{tr}\left(J\bar{G}(\lambda,\rho)\right)\right],\tag{C.6}$$

$$I_{\gamma\rho}(\theta)=\left[0_{1\times k},\ \frac{1}{n^*\sigma^2}\mathrm{tr}\left(JH(\rho)\right)\right]^{'},\ I_{\gamma\gamma}(\theta)=\begin{bmatrix}\frac{1}{n^*\sigma^2}\bar{X}^{'}J\bar{X} & 0_{k\times 1}\\ 0_{1\times k} & \frac{1}{2\sigma^4}\end{bmatrix},\tag{C.7}$$

$$I_{\gamma\lambda}(\theta)=\left[\frac{1}{n^*\sigma^2}\left(\bar{G}(\lambda,\rho)\bar{X}(\rho)\beta\right)^{'}J\bar{X}(\rho),\ \frac{1}{n^*\sigma^2}\mathrm{tr}\left(J\bar{G}(\lambda,\rho)\right)\right]^{'}.\tag{C.8}$$

where $\bar{G}(\lambda,\rho)=R(\rho)G(\lambda)R^{-1}(\rho)$ and $A^s=A+A^{'}$ for any square matrix A. To calculate the required parts of the test statistics, the first order derivatives and the components of $\Sigma(\theta)$ are evaluated at $\tilde{\theta}$.

The required parts in the test statistics for contextual effects are stated in the following.

$$1.\quad L_{\psi}(\theta)=\frac{1}{n^*\sigma^2}X_2^{'}R^{'}(\rho)J\varepsilon(\theta),\quad 2.\quad L_{\phi}(\theta)=\begin{bmatrix}\frac{1}{n^*\sigma^2}\left(\varepsilon^{'}(\theta)JH(\rho)\varepsilon(\theta)-\sigma^2\mathrm{tr}\left(JH(\rho)\right)\right)\\ \frac{1}{n^*\sigma^2}Y^{'}W^{'}R^{'}(\rho)J\varepsilon(\theta)-\frac{1}{n^*}\mathrm{tr}\left(JG(\lambda)\right)\end{bmatrix},\tag{C.9}$$

$$3.\quad L_{\gamma}(\theta)=\begin{bmatrix}\frac{1}{n^*\sigma^2}X_1^{'}R^{'}(\rho)J\varepsilon(\theta)\\ \frac{1}{2n^*\sigma^2}\left(\varepsilon^{'}(\theta)J\varepsilon(\theta)-n^*\sigma^2\right)\end{bmatrix}.\tag{C.10}$$

where $X_1=\left(X_{11}^{'},\ldots,X_{1R}^{'}\right)^{'}$ and $X_2=\left(X_{21}^{'}W_1^{'},\ldots,X_{2R}^{'}W_R^{'}\right)^{'}$. Then,

$$I_{\psi\psi}(\theta)=\frac{1}{n^*\sigma^2}X_2^{'}R^{'}(\rho)JR(\rho)X_2,\quad I_{\psi\phi}(\theta)=I_{\phi\psi}^{'}(\theta)=\left[0_{k_2\times 1}\quad \frac{1}{n^*\sigma^2}X_2^{'}R^{'}(\rho)J\left(\bar{G}(\lambda,\rho)\bar{X}(\rho)\beta\right)\right]\tag{C.11}$$

$$I_{\psi\gamma}(\theta)=I_{\gamma\psi}^{'}(\theta)=\left[\frac{1}{n^*\sigma^2}X_2^{'}R^{'}(\rho)JR(\rho)X_1\quad 0_{k_2\times 1}\right],\tag{C.12}$$

$$I_{\phi\phi}(\theta)=\begin{bmatrix}\frac{1}{n^*}\mathrm{tr}\left(H^s(\rho)JH(\rho)\right) & \frac{1}{n^*}\mathrm{tr}\left(H^s(\rho)J\bar{G}(\lambda,\rho)\right)\\ \frac{1}{n^*}\mathrm{tr}\left(H^s(\rho)J\bar{G}(\lambda,\rho)\right) & \frac{1}{n^*\sigma^2}\left(\bar{G}(\lambda,\rho)\bar{X}(\rho)\beta\right)^{'}J\left(\bar{G}(\lambda,\rho)\bar{X}(\rho)\beta\right)+\frac{1}{n^*}\mathrm{tr}\left(\bar{G}^s(\lambda,\rho)J\bar{G}(\lambda,\rho)\right)\end{bmatrix},\tag{C.13}$$

$$I_{\phi\gamma}(\theta)=I_{\gamma\phi}^{'}(\theta)=\begin{bmatrix}0_{1\times k_1} & \frac{1}{n^*\sigma^2}\mathrm{tr}\left(JH(\rho)\right)\\ \frac{1}{n^*\sigma^2}\left(\bar{G}(\lambda,\rho)\bar{X}(\rho)\beta\right)^{'}JR(\rho)X_1 & \frac{1}{n^*\sigma^2}\mathrm{tr}\left(J\bar{G}(\lambda,\rho)\right)\end{bmatrix},\tag{C.14}$$

$$I_{\gamma\gamma}(\theta)=\begin{bmatrix}\frac{1}{n^*\sigma^2}X_1^{'}R^{'}(\rho)JR(\rho)X_1 & 0_{k_1\times 1}\\ 0_{1\times k_1} & \frac{1}{2\sigma^4}\end{bmatrix}.\tag{C.15}$$

Appendix D. Proofs of Propositions

In this section, we only provide proofs for Propositions 1 and 2. Other propositions can be proved similarly, hence their proofs are omitted.

Proof Proposition 1. Let $\tilde{\theta}=(\rho_\star,\lambda_\star,\tilde{\gamma}^{'})^{'}$ be the restricted optimal GMME under H_0^{ρ} and H_0^{λ}. The first result directly follows from $\sqrt{n}\,C_\rho(\tilde{\theta})\xrightarrow{d}N\left[-\mathcal{H}_{\rho\cdot\beta}\delta_\rho-\mathcal{H}_{\rho\lambda\cdot\beta}\delta_\lambda,\mathcal{H}_{\rho\cdot\beta}\right]$, where $\mathcal{H}_{\rho\cdot\beta}=\left[\mathcal{H}_{\rho\rho}-\mathcal{H}_{\rho\beta}\mathcal{H}_{\beta\beta}^{-1}\mathcal{H}_{\beta\rho}\right]$, and $\mathcal{H}_{\rho\lambda\cdot\beta}=\left[\mathcal{H}_{\rho\lambda}-\mathcal{H}_{\rho\beta}\mathcal{H}_{\beta\beta}^{-1}\mathcal{H}_{\beta\lambda}\right]$. Therefore, we provide the proofs for the last two results in the following. To this end, we determine the joint distribution of $\mathbf{C}_{\rho\lambda}(\tilde{\theta})=\left[C_\rho^{'}(\tilde{\theta}),\,C_\lambda^{'}(\tilde{\theta})\right]^{'}$ under H_0^{ρ} and H_A^{λ}. When H_A^{ρ}

and H_A^{λ} hold, the first order Taylor expansions of the pseudo-scores $\mathbf{C}_{\rho\lambda}(\tilde{\theta})$ and $C_{\beta}(\tilde{\theta})$ around θ_0 can be stated as

$$\sqrt{n}\,\mathbf{C}_{\rho\lambda}\big(\tilde{\theta}\big) = \sqrt{n}\,\mathbf{C}_{\rho\lambda}\left(\theta_0\right) - \frac{1}{n}\mathbf{G}_{\rho\lambda}^{'}\left(\theta_0\right)\widehat{\Omega}^{-1}\mathbf{G}_{\rho\lambda}\left(\bar{\theta}\right)\left(\delta_\rho,\,\delta_\lambda\right)^{'} + \frac{1}{n}\,\mathbf{G}_{\rho\lambda}^{'}\left(\theta_0\right)\widehat{\Omega}^{-1}G_{\beta}\left(\bar{\theta}\right)\sqrt{n}\big(\tilde{\beta}-\beta_0\big) + o_p(1), \tag{D.1}$$

$$\sqrt{n}\,C_{\beta}\big(\tilde{\theta}\big) = \sqrt{n}\,C_{\beta}\left(\theta_0\right) - \frac{1}{n}G_{\beta}^{'}\left(\theta_0\right)\widehat{\Omega}^{-1}\mathbf{G}_{\rho\lambda}\left(\bar{\theta}\right)\left(\delta_\rho,\,\delta_\lambda\right)^{'} + \frac{1}{n}\,G_{\beta}^{'}\left(\theta_0\right)\widehat{\Omega}^{-1}G_{\beta}\left(\bar{\theta}\right)\sqrt{n}\big(\tilde{\beta}-\beta_0\big) + o_p(1), \tag{D.2}$$

where $\bar{\theta}$ lies between $\tilde{\theta}$ and θ_0, and $\mathbf{G}_{\rho\lambda}\left(\theta\right) = \left[G_{\rho}\left(\theta\right),\,G_{\lambda}\left(\theta\right)\right]$. Using (D.1) and (D.2) and Lemma 1, the following equation can be obtained.

$$\sqrt{n}\,\mathbf{C}_{\rho\lambda}\big(\tilde{\theta}\big) = \Big[-I_2,\,\boldsymbol{\mathcal{H}}_{\rho\lambda,\beta}\mathcal{H}_{\beta\beta}^{-1}\Big] \times \begin{bmatrix} -\frac{1}{\sqrt{n}}\mathbf{G}_{\rho\lambda}^{'}\left(\theta_0\right)\Omega^{-1}g\left(\theta_0\right) \\ -\frac{1}{\sqrt{n}}G_{\beta}^{'}\left(\theta_0\right)\Omega^{-1}g\left(\theta_0\right) \end{bmatrix} - \begin{bmatrix} \mathcal{H}_{\rho\cdot\beta} & \mathcal{H}_{\rho\lambda\cdot\beta} \\ \mathcal{H}_{\rho\lambda\cdot\beta}^{'} & \mathcal{H}_{\lambda\cdot\beta} \end{bmatrix} \times \begin{bmatrix} \delta_\rho \\ \delta_\lambda \end{bmatrix} + o_p(1) \tag{D.3}$$

where $\boldsymbol{\mathcal{H}}_{\rho\lambda,\beta} = \Big[\mathcal{H}_{\rho\beta}^{'},\,\mathcal{H}_{\lambda\beta}^{'}\Big]^{'}$. By Lemma 1, we have $\begin{bmatrix} \frac{1}{\sqrt{n}}\mathbf{G}_{\rho\lambda}\left(\theta_0\right)\Omega^{-1}g(\theta_0) \\ \frac{1}{\sqrt{n}}G_{\beta}\left(\theta_0\right)\Omega^{-1}g(\theta_0) \end{bmatrix} \xrightarrow{d} N\left[0,\,\mathcal{H}\right]$. Therefore, under H_0^{ρ} and H_A^{λ}, the result in (D.3) implies that

$$\sqrt{n}\,\mathbf{C}_{\rho\lambda}(\tilde{\theta}) \xrightarrow{d} N\left[-\begin{bmatrix} \mathcal{H}_{\rho\lambda\cdot\gamma}\delta_\lambda \\ \mathcal{H}_{\lambda\cdot\gamma}\delta_\lambda \end{bmatrix},\quad \begin{bmatrix} \mathcal{H}_{\rho\cdot\gamma} & \mathcal{H}_{\rho\lambda\cdot\gamma} \\ \mathcal{H}_{\rho\lambda\cdot\gamma}^{'} & \mathcal{H}_{\lambda\cdot\gamma} \end{bmatrix}\right]. \tag{D.4}$$

The result in (D.4) can be used to determine the asymptotic distribution of the adjusted pseudo-gradient $\sqrt{n}\,\Big[C_{\rho}(\tilde{\theta}) - \mathcal{H}_{\rho\lambda\cdot\beta}\mathcal{H}_{\lambda\cdot\beta}^{-1}C_{\lambda}(\tilde{\theta})\Big] = \Big[1, -\mathcal{H}_{\rho\lambda\cdot\beta}\mathcal{H}_{\lambda\cdot\beta}^{-1}\Big]\sqrt{n}\,\mathbf{C}_{\rho\lambda}(\tilde{\theta})$. Then, using (D.4), we can find that

$$\sqrt{n}\,\Big[C_{\rho}(\tilde{\theta}) - \mathcal{H}_{\rho\lambda\cdot\beta}\mathcal{H}_{\lambda\cdot\beta}^{-1}C_{\lambda}(\tilde{\theta})\Big] \xrightarrow{d} N\Big[0, \mathcal{H}_{1\cdot3} - \mathcal{H}_{12\cdot3}\mathcal{H}_{33}^{-1}\mathcal{H}_{12\cdot3}^{'}\Big]. \tag{D.5}$$

This last result and Lemma 1 imply that $\mathrm{LM}_{\rho}^{\mathrm{g}\star}\big(\tilde{\theta}\big) \xrightarrow{d} \chi_1^2$. Note that, (D.5) also holds under H_0^{ρ} and H_0^{λ}. This completes the proof of Proposition 1 (2).

The result in (D.3) can also be used to compare the asymptotic power of $\mathrm{LM}_1^{\mathrm{g}\star}\big(\tilde{\theta}\big)$ and $\mathrm{LM}_1^{\mathrm{g}}\big(\tilde{\theta}\big)$. Under H_A^{ρ} and H_A^{λ}, i.e., when there is no local parametric misspecification in the alternative model, the result in (D.3) implies that

$$\sqrt{n}\,C_{\rho}^{\star}(\tilde{\theta}) \xrightarrow{d} N\Big[-\big(\mathcal{H}_{\rho\cdot\beta} - \mathcal{H}_{\rho\lambda\cdot\beta}\mathcal{H}_{\lambda\cdot\beta}^{-1}\mathcal{H}_{\rho\lambda\cdot\beta}^{'}\big)\delta_\rho,\,\mathcal{H}_{\rho\cdot\beta} - \mathcal{H}_{\rho\lambda\cdot\beta}\mathcal{H}_{\lambda\cdot\beta}^{-1}\mathcal{H}_{\rho\lambda\cdot\beta}^{'}\Big]. \tag{D.6}$$

Therefore $\mathrm{LM}_{\rho}^{\mathrm{g}\star}(\tilde{\theta}) \xrightarrow{d} \chi_1^2(\vartheta_2)$, where $\vartheta_2 = \delta_\rho^2\left(\mathcal{H}_{\rho\cdot\beta} - \mathcal{H}_{\rho\lambda\cdot\beta}\mathcal{H}_{\lambda\cdot\beta}^{-1}\mathcal{H}_{\rho\lambda\cdot\beta}^{'}\right)$. It follows that $\vartheta_2 - \vartheta_4 \geq 0$, which shows that $\mathrm{LM}_{\rho}^{\mathrm{g}\star}\big(\tilde{\theta}\big)$ has less asymptotic power than $\mathrm{LM}_{\rho}^{\mathrm{g}}\big(\tilde{\theta}\big)$ when there is no local parametric misspecification. Note that the result in (D.6) also holds under H_A^{ρ} and H_A^{λ}. This completes the proof of Proposition 1 (3). □

Proof of Proposition 2. The first result directly follows from $\sqrt{n}\,C_{\lambda}(\tilde{\theta}) \xrightarrow{d} N\left[-\mathcal{H}_{\lambda\cdot\beta}\delta_\lambda - \mathcal{H}_{\lambda\rho\cdot\beta}\delta_\rho, \mathcal{H}_{\lambda\cdot\beta}\right]$, where $\mathcal{H}_{\lambda\cdot\beta} = \Big[\mathcal{H}_{\lambda\lambda} - \mathcal{H}_{\lambda\beta}\mathcal{H}_{\beta\beta}^{-1}\mathcal{H}_{\beta\lambda}\Big]$, and $\mathcal{H}_{\lambda\rho\cdot\beta} = \Big[\mathcal{H}_{\lambda\rho} - \mathcal{H}_{\lambda\beta}\mathcal{H}_{\beta\beta}^{-1}\mathcal{H}_{\beta\rho}\Big]$. Here, we provide only proofs for the last two results of the proposition. We will determine the limiting distribution of $\mathbf{C}_{\lambda\rho}(\tilde{\theta}) = \Big[C_{\lambda}^{'}(\tilde{\theta}), C_{\rho}^{'}(\tilde{\theta})\Big]^{'}$ under H_0^{λ} and H_A^{ρ}. A result similar to (D.3) can be derived as

$$\sqrt{n}\,\mathbf{C}_{\lambda\rho}\big(\tilde{\theta}\big) = \Big[-I_2,\,\boldsymbol{\mathcal{H}}_{\lambda\rho,\beta}\mathcal{H}_{\beta\beta}^{-1}\Big] \times \begin{bmatrix} -\frac{1}{\sqrt{n}}\mathbf{G}_{\lambda\rho}^{'}\left(\theta_0\right)\Omega^{-1}g\left(\theta_0\right) \\ -\frac{1}{\sqrt{n}}G_{\beta}^{'}\left(\theta_0\right)\Omega^{-1}g\left(\theta_0\right) \end{bmatrix} - \begin{bmatrix} \mathcal{H}_{\lambda\cdot\beta} & \mathcal{H}_{\lambda\rho\cdot\beta} \\ \mathcal{H}_{\rho\lambda\cdot\beta}^{'} & \mathcal{H}_{\rho\cdot\beta} \end{bmatrix} \times \begin{bmatrix} \delta_\lambda \\ \delta_\rho \end{bmatrix} + o_p(1) \tag{D.7}$$

where $\mathbf{G}_{\lambda\rho}(\theta) = [G_{\lambda}(\theta), G_{\rho}(\theta)]$ and $\boldsymbol{\mathcal{H}}_{\lambda\rho,\beta} = \left[\mathcal{H}_{\lambda\beta}^{'}, \mathcal{H}_{\rho\beta}^{'}\right]^{'}$. By Lemma 1, we have

$$\begin{bmatrix} -\frac{1}{\sqrt{n}}\mathbf{G}_{\lambda\rho}^{'}(\theta_0)\,\Omega^{-1}g(\theta_0) \\ -\frac{1}{\sqrt{n}}G_{\beta}^{'}(\theta_0)\,\Omega^{-1}g(\theta_0) \end{bmatrix} \xrightarrow{d} N\left[0, \begin{bmatrix} \mathcal{H}_{\lambda\lambda} & \mathcal{H}_{\lambda\rho} & \mathcal{H}_{\lambda\beta} \\ \mathcal{H}_{\rho\lambda} & \mathcal{H}_{\rho\rho} & \mathcal{H}_{\rho\beta} \\ \mathcal{H}_{\beta\lambda} & \mathcal{H}_{\beta\rho} & \mathcal{H}_{\beta\beta} \end{bmatrix}\right]. \tag{D.8}$$

Using (D.8) in (D.7), we obtain the following result under H_0^{λ} and H_A^{ρ}.

$$\sqrt{n}\,\mathbf{C}_{\lambda\rho}(\tilde{\theta}) \xrightarrow{d} N\left[-\begin{bmatrix} \mathcal{H}_{\lambda\rho\cdot\gamma}\delta_{\rho} \\ \mathcal{H}_{\rho\cdot\gamma}\delta_{\rho} \end{bmatrix}, \begin{bmatrix} \mathcal{H}_{\lambda\cdot\gamma} & \mathcal{H}_{\lambda\rho\cdot\gamma} \\ \mathcal{H}_{\lambda\rho\cdot\gamma}^{'} & \mathcal{H}_{\rho\cdot\gamma} \end{bmatrix}\right]. \tag{D.9}$$

Then, our assumptions and Lemma 1 ensure that

$$\begin{aligned} C_{\lambda}^{\star}(\tilde{\theta}) &= \left[C_{\lambda}(\tilde{\theta}) - B_{\lambda\rho\cdot\beta}(\tilde{\theta})B_{\rho\cdot\beta}^{-1}(\tilde{\theta})C_{\rho}(\tilde{\theta})\right] \\ &= \left[C_{\lambda}(\tilde{\theta}) - \mathcal{H}_{\lambda\rho\cdot\beta}\mathcal{H}_{\rho\cdot\beta}^{-1}C_{\rho}(\tilde{\theta})\right] + o_p(1) \xrightarrow{d} N\left[0, \mathcal{H}_{\lambda\cdot\beta} - \mathcal{H}_{\lambda\rho\cdot\beta}\mathcal{H}_{\rho\cdot\beta}^{-1}\mathcal{H}_{\lambda\rho\cdot\beta}^{'}\right]. \end{aligned} \tag{D.10}$$

This last result and Lemma 1 imply that $\mathrm{LM}_{\lambda}^{g\star}(\tilde{\theta}) \xrightarrow{d} \chi_1^2$. Since (D.10) also holds under H_0^{λ} and H_0^{ρ}, the result in Proposition 2 (2) follows.

Under H_A^{λ} and H_0^{ρ}, i.e., when there is no parametric misspecification in the alternative model, the result in (D.7) implies that

$$\sqrt{n}\,C_{\lambda}^{\star}(\tilde{\theta}) \xrightarrow{d} N\left[-(\mathcal{H}_{\lambda\cdot\beta} - \mathcal{H}_{\lambda\rho\cdot\beta}\mathcal{H}_{\rho\cdot\beta}^{-1}\mathcal{H}_{\lambda\rho\cdot\beta}^{'})\delta_{\lambda}, \,, \mathcal{H}_{\lambda\cdot\beta} - \mathcal{H}_{\lambda\rho\cdot\beta}\mathcal{H}_{\rho\cdot\beta}^{-1}\mathcal{H}_{\lambda\rho\cdot\beta}^{'}\right]. \tag{D.11}$$

Therefore, $\mathrm{LM}_{\lambda}^{g\star}(\tilde{\theta}) \xrightarrow{d} \chi_1^2(\zeta_2)$, where $\zeta_2 = \delta_{\lambda}^2(\mathcal{H}_{\lambda\cdot\beta} - \mathcal{H}_{\lambda\rho\cdot\beta}\mathcal{H}_{\rho\cdot\beta}^{-1}\mathcal{H}_{\lambda\rho\cdot\beta}^{'})$. It follows that $\zeta_1 - \zeta_2 \geq 0$ under H_A^{λ} and H_0^{ρ}. This result indicates that $\mathrm{LM}_{\lambda}^{g\star}(\tilde{\theta})$ has less asymptotic power than $\mathrm{LM}_{\lambda}^{g}(\tilde{\theta})$ when there is no local parametric misspecification in the model. Since (D.11) also holds under H_A^{λ} and H_A^{ρ}, the last result in Proposition 2 follows. □

Proof Corollary 1. These results directly follow by applying the inverse of the partitioned matrix formula to $\left[\mathbf{B}_{1\cdot 3}(\tilde{\theta})\right]^{-1}$ and $\left[\mathbf{I}_{1\cdot 3}(\tilde{\theta})\right]^{-1}$ in (3.13) and (4.10), respectively. □

References

Anderson, T., 2003. An Introduction to Multivariate Statistical Analysis, 3rd Edition. Wiley Series in Probability and Statistics. Wiley-Interscience.

Anselin, L., 1988. Spatial econometrics: Methods and Models. Springer, New York.

Anselin, L., Bera, A. K., Florax, R., Yoon, M. J., 1996. Simple diagnostic tests for spatial dependence. Regional Science and Urban Economics 26 (1), 77–104.

Bera, A. K., Bilias, Y., 2001. Rao's score, neyman's c(α) and silvey's lm tests: an essay on historical developments and some new results. Journal of Statistical Planning and Inference 97 (1), 9 – 44.

Bera, A. K., Montes-Rojas, G., Sosa-Escudero, W., 12 2010. General specification testing with locally misspecified models. Econometric Theory 26.

Bera, A. K., Yoon, M. J., 1993. Specification testing with locally misspecified alternatives. Econometric Theory 9 (4).

Bramoullé, Y., Djebbari, H., Fortin, B., 2009. Identification of peer effects through social networks. Journal of Econometrics 150 (1), 41 – 55.

Brock, W. A., Durlauf, S. N., 2001. Chapter 54 - interactions-based models. Vol. 5 of Handbook of Econometrics. Elsevier, pp. 3297 – 3380.

Burridge, P., Elhorst, J. P., Zigova, K., 2016. Group interaction in research and the use of general nesting spatial models. In: Spatial Econometrics: Qualitative and Limited Dependent Variables. Ch. 8, pp. 223–258.

Calvo-Armengol, A., Patacchini, E., Zenou, Y., 2009. Peer effects and social networks in education. The Review of Economic Studies 76 (4), 1239–1267.

Davidson, R., MacKinnon, J. G., 1987. Implicit alternatives and the local power of test statistics. Econometrica 55 (6).

Davidson, R., MacKinnon, J. G., 1998. Graphical methods for investigating the size and power of hypothesis tests. The Manchester School 66 (1), 1–26.

Elhorst, J. P., 2010. Applied spatial econometrics: Raising the bar. Spatial Economic Analysis 5 (1), 9–28.

Elhorst, J. P., 2014. Spatial Econometrics: From Cross-Sectional Data to Spatial Panels. SpringerBriefs in Regional Science. Springer Berlin Heidelberg.

Fuller, W. A., 1996. Introduction to Statistical Time Series. Wiley Series in Probability and Statistics. Wiley.

Goldsmith-Pinkham, P., Imbens, G. W., 2013. Social networks and the identification of peer effects. Journal of Business & Economic Statistics 31 (3), 253–264.

Halleck Vega, S., Elhorst, J. P., 2015. The slx model. Journal of Regional Science 55 (3), 339–363.

Hsieh, C.-S., Lee, L. F., 2014. A social interactions model with endogenous friendship formation and selectivity. Journal of Applied Econometrics.

Kelejian, H. H., Prucha, I. R., 2001. On the asymptotic distribution of the moran I test statistic with applications. Journal of Econometrics 104 (2), 219–257.

Kelejian, H. H., Prucha, I. R., 2010. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. Journal of Econometrics 157, 53–67.

Lee, L.-f., November 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. Econometrica 72 (6), 1899–1925.

Lee, L.-f., 2007. Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. Journal of Econometrics 140, 333–374.

Lee, L.-f., Liu, X., Lin, X., 2010. Specification and estimation of social interaction models with network structures. The Econometrics Journal 13, 145–176.

LeSage, J., Pace, R. K., 2009. Introduction to Spatial Econometrics (Statistics: A Series of Textbooks and Monographs. Chapman and Hall/CRC, London.

Lin, X., 2010. Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables. Journal of Labor Economics 28 (4), 825–860.

Liu, X., Lee, L.-f., 2010. Gmm estimation of social interaction models with centrality. Journal of Econometrics 159 (1), 99 – 115.

Liu, X., Patacchini, E., Zenou, Y., 2014. Endogenous peer effects: local aggregate or local average? Journal of Economic Behavior & Organization 103 (0), 39 – 59.

Manski, C. F., 1993. Identification of endogenous social effects: The reflection problem. Review of Economic Studies 60, 531–542.

Newey, W. K., 1985a. Generalized method of moments specification testing. Journal of Econometrics 29 (3), 229 – 256.

Newey, W. K., 1985b. Maximum likelihood specification testing and conditional moment tests. Econometrica 53 (5), 1047–1070.

Newey, W. K., 1985c. Maximum likelihood specification testing and conditional moment tests. Econometrica 53 (5), 1047–1070.

Newey, W. K., McFadden, D., 1994. Chapter 36 large sample estimation and hypothesis testing. Vol. 4 of Handbook of Econometrics. Elsevier, pp. 2111 – 2245.

Newey, W. K., West, K. D., 1987. Hypothesis testing with efficient method of moments estimation. International Economic Review 28 (3), pp. 777–787.

Pace, R. K., Barry, R., 1997. Quick computation of spatial autoregressive estimators. Geographical Analysis 29 (3), 232–247.

Saikkonen, P., 1989. Asymptotic relative efficiency of the classical test statistics under misspecification. Journal of Econometrics 42 (3), 351 – 369.

Smith, R. J., 1987. Alternative asymptotically optimal tests and their application to dynamic specification. The Review of Economic Studies 54 (4), 665–680.

Taspinar, S., Dogan, O., 2016. Teaching size and power properties of hypothesis tests through simulations. Journal of Econometric Methods 5 (1).

Tauchen, G., 1985. Diagnostic testing and evaluation of maximum likelihood models. Journal of Econometrics 30 (1), 415 – 443.

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