Robust LM Tests for Spatial Dynamic Panel Data Models under both Parametric and Distributional Misspecifications*

Yufan Leiluo†

November 18, 2019

Abstract

In this paper, I introduce Rao's score test statistics (Lagrange multiplier (LM) tests) for a spatial dynamic panel data (SDPD) model which includes a contemporaneous spatial lag, a time lag and a spatial-time lag. The tests are robust to both parametric and distributional misspecifications, and can be used to test the significance of each of the three lag terms. It can also test any combination of them jointly. The quasi maximum likelihood estimator (QMLE) for the SDPD models suffers from an incidental parameter problem due to the individual and time fixed effects in the model. The score functions then can have asymptotic bias, and are not centered on zero. This paper shows how to make correction to the score functions, so they are centered at 0 using both transformation approach and direct approach for estimation of the model. Then, I derive LM test statistics that are valid under distributional misspecification, and LM tests that are robust to local parametric misspecification. Finally, the two derived tests are combined to construct new LM tests, which are robust to both parametric and distributional misspecifications. I illustrate the performance of the suggested test in a Monte Carlo study and empirical applications.

JEL-Classification: C13, C21, C31.
Keywords: Spatial dynamic panel data model, SDPD, Rao's score tests, LM tests, MLE, Spatial dependence, Robust LM tests.

*I would like to thank my advisor Anil K. Bera for his guidance and support. I am also grateful for the valuable comment from my committee membersg Eun Yi Chung, JiHyung Lee, and Xiaofeng Shao.

†Department of Economics, University of Illinois at Urbana-Champaign, IL, USA, email: leiluo2@illinois.edu

1 Introduction

Mahalanobis (1944) and Moran (1950) bring the attention to the interaction between different locations that are spatially close to each other. Cliff and Ord (1972) introduces the spatial autoregressive (SAR) model which is a major tool to analyze data that shows spatial correlation. To allow for more modeling flexibility, the SAR model is then extended into a spatial panel data model. Elhorst (2003), Baltagi et al. (2003, 2007), Kapoor et al. (2007), Parent and LeSage (2010, 2011, 2012) studied the spatial panel data models with random effects. For fixed effect models, Korniotis (2010), Su and Yang (2015), Yu et al. (2008), Lee and Yu (2010, 2012) have investigated the static or dynamic models under various model structures.

In my paper, I consider the SDPD model setting in Lee and Yu (2010). The model includes a contemporaneous spatial lag term, a time lag term and a spatial-time lag term with both the individual and time fixed effects. For the estimation of the SDPD model, there are two common approaches, direct approach and transformation approach. With direct approach, the fixed effects are concentrated out by is QMLE or MLE, while for the transformation approach, the model is first transformed to wipe out the time fixed effect, and then the individual fixed effects are concentrated out by the corresponding QMLE or MLE. However, with either approach, the MLE or QMLE suffers from an incidental problem caused by the fixed effects. The limiting distribution of score functions derived from the concentrated log-likelihood function of the direct approach or from the log-likelihood function of the transformed model may not be centered on zero even when the number of individuals (n), and time periods (T) are large. The incidental parameter problem could cause a bias with an order $max\{O(1/T), O(1/n)\}$ for direct approach, and $O(1/T)$ for transformation approach. Lee and Yu (2010) and Yu et al. (2008) suggest to add a bias correction term to the score function to improve inference in the context of the SDPD models.

The estimation issues discussed in the above are also central concerns in developing score based test statistics for SDPD models. The classical LM test (Rao, 1948) does not take into account of the bias correction term, so the asymptotic distribution of the score function will be invalid, and the classic LM test will fail in SDPD model. So far, most of testing procedures have been considered for cross-sectional spatial models. For example, among others, see Anselin (1988, 2001), Anselin and Moreno (2003), Anselin and Rey (1991), Anselin et al. (1996), Baltagi and Li (2001), Baltagi and Yang (2013), Bera et al. (2018a,b), Born and Breitung (2011), Doğan et al. (2018), Kelejian and Robinson (1992), Pinkse (2004), Robinson (2008), Taşpınar et al. (2018), and Yang (2010). As for the spatial static and dynamic panel data models, the bulk of literature focuses on the estimation issues, while there are only few studies on testing and model specification issues for these models. We are aware of three studies on testing hypotheses for these models, namely Taşpınar et al. (2017), Yang (2016) and Bera et al. (2019). Taşpınar et al. (2017) use the GMM approach suggested in Lee and Yu (2014) and develop GMM gradient test statistics for a higher order SDPD model that includes a time lag term, spatial time lag terms and contemporaneous spatial lag terms. Their adjusted tests are robust to parametric misspecification in the alternative model and are asymptotically valid under both small and large T cases. Yang (2016) considers an extended SDPD model that also includes a spatial lag in the disturbance term. He uses the unified-M estimation method developed in Yang (2018) to derive joint and marginal LM test statistics for hypotheses about various spatial and time effects. All test statistics suggested in Yang (2016) are free from specifications of the initial conditions and are consistent when T is fixed. To achieve these desirable properties, the mean of score functions are subtracted from score functions to get the adjusted score functions that have a well defined limiting distribution. Yang (2016) then uses these adjusted score functions to formulate joint and marginal tests for spatial and time effects. Bera et al. (2019) develop LM test using the method suggested by Bera and Yoon (1993) and Bera et al. (2017) based on direct approach. They develop LM tests that is robust to parametric misspecification in the alternative model. The tests require both n and T go to ∞, and n/T should be non-zero and finite. This paper considers a more general test comparing to Bera et al. (2019) in the sense that distributional misspecification is included, and the tests are also extended to the transformation approach. The LM test statistics are constructed in the following steps. First, I derive the asymptotic distributions of the bias corrected score functions of SDPD model assuming no parametric or distributional misspecification, and form non-robust LM test statistics. Then, I consider the case when the normal distribution assumption is violated, and derive the LM tests that are robust to this distributional misspecification. I adjust the score functions so that the resulting LM statistics are valid when there is local parametric misspecification in the alternative models. Finally, I combine the two robust tests and obtain the LM tests that are robust to both parametric and distributional misspecifications. I prove that under a quite general restriction, the tests that are robust to both distributional and parametric misspecifications are asymptotic equivalent to the tests that are only robust to parametric misspecification, so the LM test that is only robust to parametric misspecification will provide reliable inference, even when the error is not normally distributed. The tests are constructed with both direct and transformation approaches. The direct approach can be applied to a weight matrix that is not row-normalized but requires limit of n/T being finite and non zero, while the transformation requires only $n/T < \infty$. Each has certain advantages.

The LM tests can be applied for testing the presence of the contemporaneous spatial lag, the time lag and the spatial-time lag in the SDPD model. One desired feature of LM tests is that these tests require only the estimation of the restricted model under the null hypothesis, with which all the spatial and time lag terms are restricted to be 0. The procedure circumvents a major part of estimation burden. With all the lag terms set to 0, the tests in this paper are simple and their computations only require ordinary least squared (OLS) estimates from a conventional (non-spatial) two-way error panel data model.

I design a Monte Carlo study to investigate the finite sample size and power properties of the proposed test statistics. The simulation results show that the robust test statistics have good size and power properties. Also, I demonstrate advantages of applying my robust tests in inference to empirical examples. Finally, to facilitate the application of my suggested test statistics in a specification search, an R software package (`sdpdlm`) is developed and available upon request.

The rest of this paper is organized as follows. Section 2 presents the SDPD model and its assumptions. Section 3 lays out the details of the QML estimation of the model. Sections 4 to 7 present the methods to derive the robust LM test statistics. Section 8 provides the details to compute suggested tests. Section 9 lays out the details of the Monte Carlo design and presents the results. It also presents empirical applications to illustrate the use of my proposed test methodology in practice. Section 10 ends the paper with concluding remarks. Some of the technical derivations and simulation results are relegated to an appendix.

2 The Model Specification and Assumptions

The SDPD model considered is

$$Y_{nt} = \lambda_0 W_n Y_{nt} + \gamma_0 Y_{n,t-1} + \rho_0 W_n Y_{n,t-1} + X_{nt}\beta_0 + \mathbf{c}_{n0} + \alpha_{t0} l_n + V_{nt}, \qquad t = 1, \ldots, T \tag{2.1}$$

where $Y_{nt} = (y_{1t}, y_{2t}, \ldots, y_{nt})'$ and $V_{nt} = (v_{1t}, v_{2t}, \ldots, v_{nt})'$ are both $n \times 1$ vector. v_{it}'s are *i.i.d.* distributed across t and i with mean of 0 and variance of σ_0^2. X_{nt} is the $n \times k$ matrix of exogenous variables with a conformable parameter vector β_0, W_n is the $n \times n$ known exogenous spatial weight matrix with zero diagonal elements. $\mathbf{c}_{n0}$ is the $n \times 1$ vector individual fixed effects, and α_{t0} is the time effect for period t,and l_n is the $n \times 1$ vector of ones. Let Θ be the parameter space. In order to distinguish the true parameter vector from other possible values in Θ, the true parameter vector is denoted by $\theta_0 = (\gamma_0, \rho_0, \beta_0', \lambda_0, \sigma_0^2)'$. An arbitrary value of parameter vector is denoted by $\theta = (\gamma, \rho, \beta', \lambda, \sigma^2)'$. Denote $S_n(\lambda) = (I_n - \lambda W_n)$ and $G_n(\lambda) = W_n S_n^{-1}(\lambda)$ for an arbitrary value λ. At the true parameter value, these terms are denoted by $S_n(\lambda_0) = S_n$ and $G_n(\lambda_0) = G_n$ for simplicity. Let $\overline{\Upsilon}_{nt}$ be an $n \times 1$ vector for $t = 1, \ldots, T$. I define $\overline{\Upsilon}_{nT} = \frac{1}{T}\sum_{t=1}^{T} \Upsilon_{nt}$, $\overline{\Upsilon}_{nT,-1} = \frac{1}{T}\sum_{t=1}^{T} \Upsilon_{n,t-1}$, $\tilde{\Upsilon}_{nt} = \Upsilon_{nt} - \overline{\Upsilon}_{nT}$, and $\tilde{\Upsilon}_{n,t-1} = \Upsilon_{n,t-1} - \overline{\Upsilon}_{nT,-1}$.
For the analysis of the asymptotic properties of estimators and the LM tests, I consider the model (2.1) under the following assumptions.

Assumption 1. *(i) W_n is a non-stochastic spatial weights matrix with zero diagonal elements, and it satisfies $W_n l_n = l_n$, (ii) W_n is a non-stochastic spatial weights matrix with zero diagonal elements, and it may or may not satisfy $W_n l_n = l_n$.*

Assumption 2. *(i) The disturbance terms v_{it}'s, $i = 1, ..., n$ and $t = 1, ..., T$, are independent and identically normally distributed across all i and t with mean zero and variance σ^2. (ii) The disturbance terms v_{it}'s, $i = 1, \ldots, n$ and $t = 1, \ldots, T$, are independent and identically distributed across all i and t with mean zero and variance σ_0^2, and $E|v_{it}|^{4+\eta} < \infty$ for some $\eta > 0$.*[1]

Assumption 3. *(i) $S_n(\lambda)$ is invertible for all $\lambda \in \Delta_\lambda$, where Δ_λ is a compact parameter space, and $\lambda_0 \in Int(\Delta_\lambda)$. (ii) The row and column sums of W_n and $S_n^{-1}(\lambda)$ are bounded in absolute*

[1] Note that the object of the central limit theorem (CLT) for this model involves the linear and quadratic forms of V_{nt}. When v_{it}'s are simply i.i.d, the CLT in Kelejian and Prucha (2001, 2010) requires the existence of $(4 + \eta_v)$th moments for disturbance terms, where $\eta_v > 0$.

value uniformly in n *and* $\lambda \in \Delta_\lambda$. *(iii) The sum* $\sum_{h=1}^{\infty} abs(A_n^h)$ *is uniformly bounded, where* $A_n = S_n^{-1}(\gamma_0 I_n + \rho_0 W_n)$ *and* $[abs(A_n)]_{ij} = |A_{n,ij}|$.[2]

Assumption 4. X_{nt} *has non-stochastic uniformly bounded elements for all* n *and* t*, and* $\lim_{T\to\infty} \frac{1}{nT}\sum_{t=1}^{T} \tilde{X}_{nt}' J_n \tilde{X}_{nt}$ *exists and is nonsingular, where* $J_n = I_n - \frac{1}{n} l_n l_n'$ *be the deviation from the group mean transformation.*

Assumption 5. *(i)* n *is a non-decreasing function of* T*, and* T *goes to infinity. (ii)* n *is an increasing function of* T*, and* T *goes to infinity.*

Most of the assumptions adopted for the model are the usual regularity conditions assumed in the literature (for example, see Kelejian and Prucha (2001, 2010) and Lee and Yu (2012)). Assumption 2 (ii) provides regularity assumptions for disturbance terms such that central limit theorem is applicable. Assumptions 3(iii) and 5 are consider by Lee and Yu (2010) for (2.1). Assumption 3(iii) is suggested for the dynamic spatial panel data models to limit the dependence over time and over the cross-sectional units. The asymptotic setting is characterized by Assumption 5. Assumptions 5(i) allows two cases: (i) $n \to \infty$ as $T \to \infty$, and (ii) n is fixed and $T \to \infty$, while Assumption 5(ii) only allows $n \to \infty$ as $T \to \infty$.

3 The QML Estimation Approach

Under the asymptotic settings in Assumption 5, Lee and Yu (2010) suggest two estimation approaches in the QML or ML framework. In the first approach, the model is transformed under Assumption 1(i), the model is first transformed with J_n to eliminate the time fixed effects. Then the transformed model is further transformed with the orthonormal eigenvectors matrix of J_n to remove the linear dependence among transformed disturbance terms, and the model is then estimated with a QMLE. In the second approach, direct approach, both individual and time fixed effects are estimated along with the other parameters. The fixed effects are directly concentrated out. Transformation approach requires that $n/T^3 \to 0$ for the estimator to be consistent; direct approach requires both $n/T^3 \to 0$ and $T/n^3 \to 0$ to have a consistent estimation. Hence, the transformation approach has an advantage over the direct approach especially when n is relatively small. However, the direct approach does not require the weight matrix to be row normalized. In this paper, I consider constructing the test statistics using both approaches, so researchers can choose either base on their needs.

For the transformation approach, under Assumption 1(i), $J_n W_n = J_n W_n J_n$, hence the transformation with J_n yields

$$J_n Y_{nt} = \lambda_0 J_n W_n J_n Y_{nt} + \gamma_0 J_n Y_{n,t-1} + \rho_0 J_n W_n J_n Y_{n,t-1} + J_n X_{nt}\beta_0 + J_n \mathbf{c}_{n0} + J_n V_{nt}. \quad (3.1)$$

[2] A sufficient condition for this assumption is that $\|A_n\| < 1$, where $\|\cdot\|$ is any matrix norm (Lee and Yu, 2010). When this condition holds, we have $(I_n - A)^{-1} = \sum_{h=1}^{\infty} A_n^h$.

Let $(F_{n,n-1},\, l_n/\sqrt{n})$ be the orthonormal eigenvectors matrix of J_n, where the $n\times(n-1)$ matrix $F_{n,n-1}$ corresponds to eigenvalue one, and $l_n/\sqrt{n}$ corresponds to eigenvalue zero.[3] Denote $Y_{nt}^{*}=F_{n,n-1}^{'}J_nY_{nt}$, $W_n^{*}=F_{n,n-1}^{'}W_nF_{n,n-1}$, $X_{nt}^{*}=F_{n,n-1}^{'}J_nX_{nt}$, $\mathbf{c}_{n0}^{*}=F_{n,n-1}^{'}J_n\mathbf{c}_{n0}$, and $V_{nt}^{*}=F_{n,n-1}^{'}J_nV_{nt}$. Then, transforming (3.1) into Y_{nt}^{*} yields

$$Y_{nt}^{*}=\lambda_0W_n^{*}Y_{nt}^{*}+\gamma_0Y_{n,t-1}^{*}+\rho_0W_n^{*}Y_{n,t-1}^{*}+X_{nt}^{*}\beta_0+\mathbf{c}_{n0}^{*}+V_{nt}^{*}. \tag{3.2}$$

The log-likelihood function for (3.2) can be written as

$$\begin{aligned}\ln L_{nT}^{t}\left(\theta,\mathbf{c}_n^{*}\right)=&-\frac{(n-1)T}{2}\ln 2\pi-\frac{(n-1)T}{2}\ln\sigma^2+T\ln|S_n^{*}(\lambda)|\\&-\frac{1}{2\sigma^2}\sum_{t=1}^{T}V_{nt}^{*'}\left(\theta\right)V_{nt}^{*}\left(\theta\right),\end{aligned} \tag{3.3}$$

where $S_n^{*}(\lambda)=(I_{n-1}-\lambda W_n^{*})$, $V_{nt}^{*}=S_n^{*}(\lambda)Y_{nt}^{*}-Z_{nt}^{*}\delta-\mathbf{c}_{n0}^{*}$, $Z_{nt}^{*}=\left(Y_{n,t-1}^{*},\,W_n^{*}Y_{n,t-1}^{*},\,X_{nt}^{*}\right)$ and $\delta=\left(\gamma,\,\rho,\,\beta^{'}\right)^{'}$. By concentrating out $\mathbf{c}_{n0}^{*}$ and using properties in Footnote 3, (3.3) can be written as

$$\begin{aligned}\ln L_{nT}^{t}\left(\theta\right)=&-\frac{(n-1)T}{2}\ln 2\pi-\frac{(n-1)T}{2}\ln\sigma^2+T\ln|S_n(\lambda)|-T\ln(1-\lambda)\\&-\frac{1}{2\sigma^2}\sum_{t=1}^{T}\tilde{V}_{nt}^{'}\left(\theta\right)J_n\tilde{V}_{nt}\left(\theta\right),\end{aligned} \tag{3.4}$$

where $\tilde{V}_{nt}(\theta)=S_n(\lambda)\tilde{Y}_{nt}-\tilde{Z}_{nt}\delta$ and $\tilde{Z}_{nt}=\left(\tilde{Y}_{n,t-1},\,W_n\tilde{Y}_{n,t-1},\,\tilde{X}_{nt}\right)$.

For direct approach, the log-likelihood function for (2.1) is written directly. For this purpose, let $\boldsymbol{\alpha}_T=(\alpha_1,\ldots,\alpha_T)^{'}$ be the vector of time effects. The log-likelihood function of (2.1) is written as

$$\begin{aligned}\ln L_{nT}^{d}\left(\theta,\mathbf{c}_n,\boldsymbol{\alpha}_T\right)=&-\frac{nT}{2}\ln 2\pi-\frac{nT}{2}\ln\sigma^2+T\ln|S_n(\lambda)|\\&-\frac{1}{2\sigma^2}\sum_{t=1}^{T}V_{nt}^{'}\left(\theta,\mathbf{c}_n,\boldsymbol{\alpha}_T\right)V_{nt}\left(\theta,\mathbf{c}_n,\boldsymbol{\alpha}_T\right),\end{aligned} \tag{3.5}$$

where $V_{nt}\left(\theta,\mathbf{c}_n,\boldsymbol{\alpha}_T\right)=\left[S_n(\lambda)Y_{nt}-Z_{nt}\delta-\mathbf{c}_n-\alpha_tl_n\right]$, $Z_{nt}=(Y_{n,t-1},\,W_nY_{n,t-1},\,X_{nt})$ and $\delta=\left(\gamma,\rho,\beta^{'}\right)^{'}$. In practice, the fixed effects are concentrated out from (3.5) to reduce the dimension of parameter vector before considering an optimization algorithm. The first-order condition of (3.5) with respect to α_t is $\frac{\partial\ln L_{nT}^{d}(\theta,\mathbf{c}_n,\boldsymbol{\alpha}_T)}{\partial\alpha_t}=\frac{1}{\sigma^2}l_n^{'}V_{nt}(\theta,\mathbf{c}_n,\boldsymbol{\alpha}_T)=0$ which implies that $\hat{\alpha}_t(\theta,\mathbf{c}_n)=\frac{1}{n}l_n^{'}\left(S_n(\lambda)Y_{nt}-Z_{nt}\delta-\mathbf{c}_n\right)$. Then, it follows that $V_{nt}\left(\theta,\mathbf{c}_n,\hat{\boldsymbol{\alpha}}_T(\theta,\mathbf{c}_n)\right)=J_n\left[S_n(\lambda)Y_{nt}-Z_{nt}\delta-\mathbf{c}_n\right]$. Hence, the likelihood function with $\hat{\boldsymbol{\alpha}}_T$ concentrated out is

[3] The basic properties of this orthonormal matrix are (i) $F_{n,n-1}^{'}l_n=0_{n\times 1}$, (ii) $F_{n,n-1}^{'}F_{n,n-1}=I_{n-1}$, (iii) $F_{n,n-1}F_{n,n-1}^{'}=J_n$, (iv) $F_{n,n-1}^{'}W_nl_n=0$, (v) $|I_{n-1}-\lambda W_n^{*}|=\frac{1}{1-\lambda}|I_n-\lambda W_n|$, where $W_n^{*}=F_{n,n-1}^{'}W_nF_{n,n-1}$, and (vii) $(I_{n-1}-\lambda W_n^{*})^{-1}=F_{n,n-1}^{'}(I_n-\lambda W_n)^{-1}F_{n,n-1}$.

$$\ln L_{nT}^{d}\left(\theta,\mathbf{c}_n\right) = -\frac{nT}{2}\ln 2\pi - \frac{nT}{2}\ln\sigma^2 + T\ln\left|S_n(\lambda)\right| - \frac{1}{2\sigma^2}\sum_{t=1}^{T} V_{nt}^{'}\left(\theta,\mathbf{c}_n,\hat{\boldsymbol{\alpha}}_T(\theta,\mathbf{c}_n)\right) V_{nt}\left(\theta,\mathbf{c}_n,\hat{\boldsymbol{\alpha}}_T(\theta,\mathbf{c}_n)\right). \tag{3.6}$$

The first-order condition of (3.6) with respect to $\mathbf{c}_n$ is $\frac{\partial \ln L_{nT}^{d}(\theta,\mathbf{c}_n)}{\partial \mathbf{c}_n} = \frac{1}{\sigma^2}\sum_{t=1}^{T} V_{nt}(\theta,\mathbf{c}_n,\hat{\boldsymbol{\alpha}}_T) = 0_{n\times 1}$, which implies that $\hat{\mathbf{c}}_n(\theta) = J_n\left[S_n(\lambda)\overline{Y}_{nT} - \overline{Z}_{nT}\delta\right]$. By substituting $\hat{\mathbf{c}}_n(\theta)$ into (3.6), I get the following concentrated log-likelihood function:

$$\ln L_{nT}^{d}\left(\theta\right) = -\frac{nT}{2}\ln 2\pi - \frac{nT}{2}\ln\sigma^2 + T\ln\left|S_n(\lambda)\right| - \frac{1}{2\sigma^2}\sum_{t=1}^{T} \tilde{V}_{nt}^{'}\left(\theta\right) J_n \tilde{V}_{nt}\left(\theta\right), \tag{3.7}$$

where $\tilde{V}_{nt} = S_n(\lambda)\tilde{Y}_{nt} - \tilde{Z}_{nt}\delta$.

The asymptotic distribution of the suggested robust test is based on the asymptotic distribution of score functions. In order to state asymptotic distribution of score functions, define the following bias correction terms:

$$\Delta_{nT,1}(\theta_0) = \frac{1}{\sqrt{(n-1)T}} \times \begin{bmatrix} \operatorname{tr}\left(J_n(I_n - A_n)^{-1}S_n^{-1}\right) \\ \operatorname{tr}\left(W_n J_n(I_n - A_n)^{-1}S_n^{-1}\right) \\ 0_{k_x\times 1} \\ \gamma_0\operatorname{tr}\left(G_n J_n(I_n - A_n)^{-1}S_n^{-1}\right) + \rho_0\operatorname{tr}\left(G_n W_n J_n(I_n - A_n)^{-1}S_n^{-1}\right) + \operatorname{tr}\left(J_n G_n\right) \\ \frac{n-1}{2\sigma_0^2} \end{bmatrix}, \tag{3.8}$$

$$\Delta_{nT,2}(\theta_0) = \sqrt{\frac{(n-1)}{n}}\Delta_{nT,1}(\theta_0), \tag{3.9}$$

$$\Delta_{nT,3}(\theta_0) = \sqrt{\frac{T}{n}}\begin{bmatrix} 0_{(k_x+2)\times 1} \\ \frac{1}{n}l_n^{'}G_n l_n \\ \frac{1}{2\sigma_0^2} \end{bmatrix}. \tag{3.10}$$

Under the assumptions, $\Delta_{nT,1}(\theta_0) = O\left(\sqrt{\frac{n}{T}}\right)$, $\Delta_{nT,2}(\theta_0) = O\left(\sqrt{\frac{n}{T}}\right)$, and $\Delta_{nT,3}(\theta_0) = O\left(\sqrt{\frac{T}{n}}\right)$. For notational simplification, define

$$\Sigma(\theta_0) = \lim_{T\to\infty} \mathrm{E}\left[-\frac{1}{(n-1)T}\frac{\partial^2 \ln L^t_{nT}(\theta_0)}{\partial\theta\partial\theta'}\right] \; or \; \lim_{T\to\infty} \mathrm{E}\left[-\frac{1}{nT}\frac{\partial^2 \ln L^d_{nT}(\theta_0)}{\partial\theta\partial\theta'}\right] \tag{3.11}$$

$$\begin{aligned}\Omega(\theta_0) = &\lim_{T\to\infty} \mathrm{E}\left[\frac{1}{(n-1)T}\left(\frac{\partial \ln L^t_{nT}(\theta_0)}{\partial\theta} + \Delta_{nT,1}(\theta_0)\right)\cdot\left(\frac{\partial \ln L^t_{nT}(\theta_0)}{\partial\theta} + \Delta_{nT,1}(\theta_0)\right)'\right] \\ or &\lim_{T\to\infty} \mathrm{E}\left[\frac{1}{nT}\left(\frac{\partial \ln L^d_{nT}(\theta_0)}{\partial\theta} + \Delta_{nT,2}(\theta_0) + \Delta_{nT,3}(\theta_0)\right)\cdot\left(\frac{\partial \ln L^d_{nT}(\theta_0)}{\partial\theta} + \Delta_{nT,2}(\theta_0) + \Delta_{nT,3}(\theta_0)\right)'\right]\end{aligned} \tag{3.12}$$

Here, the two sets of notations correspond to the context of transformation approach or direct approach respectively. Let $\bar{\theta} = \theta_0 + o_p(1)$. I collect the asymptotic results for analyzing the distribution of my test statistics in the following proposition.

Proposition 1. *For the transformation and direct approaches, we have the following results.*

1. *For the transformation approach, under Assumptions 1(i), 2(ii), 3, 4 and 5(i), the following results hold.*
   (a) $\frac{1}{\sqrt{(n-1)T}}\frac{\partial \ln L^t_{nT}(\theta_0)}{\partial\theta} + \Delta_{nT,1}(\theta_0) \xrightarrow{d} N\left(0, \Sigma(\theta_0)\right)$.
   (b) $-\frac{1}{(n-1)T}\frac{\partial^2 \ln L^t_{nT}(\bar{\theta})}{\partial\theta\partial\theta'} = \Sigma(\theta_0) + o_p(1)$.
   (c) $\frac{1}{nT}\left(\frac{\partial \ln L^t_{nT}(\bar{\theta})}{\partial\theta} + \Delta_{nT,1}(\bar{\theta})\right)\cdot\left(\frac{\partial \ln L^t_{nT}(\bar{\theta})}{\partial\theta} + \Delta_{nT,1}(\bar{\theta})\right)' = \Omega(\theta_0) + o_p(1)$.
   (d) $\Sigma(\theta_0) = \Omega(\theta_0)$.

2. *For the transformation approach, under Assumptions 1(i), 2(i), 3, 4 and 5(i), the following results hold.*
   (a) $\frac{1}{\sqrt{(n-1)T}}\frac{\partial \ln L^t_{nT}(\theta_0)}{\partial\theta} + \Delta_{nT,1}(\theta_0) \xrightarrow{d} N\left(0, \Sigma(\theta_0)\right)$.
   (b) $-\frac{1}{(n-1)T}\frac{\partial^2 \ln L^t_{nT}(\bar{\theta})}{\partial\theta\partial\theta'} = \Sigma(\theta_0) + o_p(1)$.
   (c) $\frac{1}{(n-1)T}\left(\frac{\partial \ln L^t_{nT}(\bar{\theta})}{\partial\theta} + \Delta_{nT,1}(\bar{\theta})\right)\cdot\left(\frac{\partial \ln L^t_{nT}(\bar{\theta})}{\partial\theta} + \Delta_{nT,1}(\bar{\theta})\right)' = \Omega(\theta_0) + o_p(1)$.

3. *For the direct approach, under Assumptions Assumptions 1(ii), 2(i), 3, 4 and 5(ii), the following results hold.*
   (a) $\frac{1}{\sqrt{nT}}\frac{\partial \ln L^d_{nT}(\theta_0)}{\partial\theta} + \Delta_{nT,2}(\theta_0) + \Delta_{nT,3}(\theta_0) \xrightarrow{d} N\left(0, \Sigma(\theta_0)\right)$.
   (b) $-\frac{1}{nT}\frac{\partial^2 \ln L^d_{nT}(\bar{\theta})}{\partial\theta\partial\theta'} = \Sigma_{nT}(\theta_0) + o_p(1)$.

(c) $\frac{1}{nT}\left(\frac{\partial \ln L^{d}_{nT}(\bar{\theta})}{\partial\theta}+\Delta_{nT,2}(\bar{\theta})+\Delta_{nT,3}(\bar{\theta})\right)\cdot\left(\frac{\partial \ln L^{d}_{nT}(\bar{\theta})}{\partial\theta}+\Delta_{nT,2}(\bar{\theta})+\Delta_{nT,3}(\bar{\theta})\right)'=\Omega(\theta_0)+o_p(1)$.

(d) $\Sigma(\theta_0)=\Omega(\theta_0)$.

4. *For the direct approach ,under Assumptions 1(ii), 2(ii), 3, 4 and 5(ii), the following results hold.*

   (a) $\frac{1}{\sqrt{nT}}\frac{\partial \ln L^{d}_{nT}(\theta_0)}{\partial\theta}+\Delta_{nT,2}(\theta_0)+\Delta_{nT,3}(\theta_0)\xrightarrow{d}N\left(0,\,\Omega(\theta_0)\right)$.

   (b) $-\frac{1}{nT}\frac{\partial^2 \ln L^{d}_{nT}(\bar{\theta})}{\partial\theta\partial\theta'}=\Sigma_{nT}(\theta_0)+o_p(1)$.

   (c) $\frac{1}{nT}\left(\frac{\partial \ln L^{d}_{nT}(\bar{\theta})}{\partial\theta}+\Delta_{nT,2}(\bar{\theta})+\Delta_{nT,3}(\bar{\theta})\right)\cdot\left(\frac{\partial \ln L^{d}_{nT}(\bar{\theta})}{\partial\theta}+\Delta_{nT,2}(\bar{\theta})+\Delta_{nT,3}(\bar{\theta})\right)'=\Omega(\theta_0)+o_p(1)$.

Proof. See Appendix C. □

There are three important observations in Proposition 1 for each of the two approaches, which are true under both ML and QML settings. Starting with the transformation approach, first when $\frac{n}{T}\to 0$, i.e., when T grows faster than n, $\Delta_{nT,1}(\theta_0)$ vanishes, and the estimation is consistent. Second, when $\frac{n}{T}\to k$, where $0<k<\infty$, i.e., when n is asymptotically proportional to T, $\Delta_{nT,1}(\theta_0)$ does not vanish and therefore MLE or QMLE under the transformation approach has an asymptotic bias of $1/\sqrt{(n-1)T}\times O\left(\sqrt{\frac{n}{T}}\right)=O(T^{-1})$. Third, when $\frac{n}{T}\to\infty$, i.e., when n grows faster than T, the presence of $\Delta_{nT,1}(\theta_0)$ causes the score functions to have a degenerate distribution: $\frac{1}{n-1}\frac{\partial \ln L_{nT}(\theta_0)}{\partial\theta}+\sqrt{\frac{T}{n-1}}\Delta_{nT,2}(\theta_0)\xrightarrow{p}0$. For the direct approach, first, when $\frac{n}{T}\to k$, where $0<k<\infty$, the limiting distribution of score functions is not centered on 0, because $\Delta_{nT,2}(\theta_0)=O\left(\sqrt{\frac{n}{T}}\right)$ and $\Delta_{nT,3}(\theta_0)=O\left(\sqrt{\frac{T}{n}}\right)$ do not vanish. Therefore, the estimator under the direct approach has a bias of $O(\max\{n^{-1},\,T^{-1}\})$. The second observation is that when $\frac{n}{T}\to\infty$, the score functions have a degenerate limiting distribution: $\frac{1}{n}\frac{\partial \ln L^{d}_{nT}(\theta_0)}{\partial\theta}+\sqrt{\frac{T}{n}}\Delta_{nT,2}(\theta_0)\xrightarrow{p}0$. The final observation on the direct approach is that when $\frac{n}{T}\to 0$, the score functions again have a degenerate limiting distribution: $\frac{1}{T}\frac{\partial \ln L^{d}_{nT}(\theta_0)}{\partial\theta}+\sqrt{\frac{n}{T}}\Delta_{nT,3}(\theta_0)\xrightarrow{p}0$.
Based on the above observations, I assume $\frac{n}{T}<\infty$ for constructing tests with transformation approach, and assume $\frac{n}{T}\to k$, $0<k<\infty$ for direct approach. Again, transformation approach has less restrictive assumption on n and T, but it requires the weight matrix to be row normalized.

4 LM tests under Ideal Condition

For the following sections, I provide the properties of my tests based on Bera et al. (2019). To make my discussion more general in the sense that the derivations are valid for both approaches, I will denote $n^\star$ as the effective sample size, so for direct approach $n^\star=n^d=nT$, for transformation approach $n^\star=n^t=(n-1)T$; denote the bias correction term as Δ_{nT}, so $\Delta_{nT}=\Delta_{nT,1}$ for transformation approach and $\Delta_{nT} = \Delta_{nT,2} + \Delta_{nT,3}$ for direct approach; denote the likelihood function as L_{nT}, so $L_{nT} = L^t_{nT}$ or L^d_{nT}.

In this section, I consider LM tests of the SDPD model assuming the ideal condition which means there is neither distributional nor parametric misspecification. Denote $\theta = (\varphi', \psi', \phi')'$ which is the parameter vector, where φ, ψ, and ϕ are $p \times 1$, $r \times 1$ and $s \times 1$ vectors, respectively. In the SDPD model, $\varphi = (\beta', \sigma^2)'$, which are the estimated parameters under restrictions that (ρ, λ, γ) take value $(\rho_\star, \lambda_\star, \gamma_\star)$, and denote the estimation of θ under such restrictions as $\tilde{\theta} = (\tilde{\varphi}', \psi'_\star, \phi'_\star)'$. ψ and ϕ could be any combinations of (ρ, λ, γ) and can be considered as the parameters of interest. In this context, ψ is the testing parameter, meaning we are testing whether it is true that $\psi = 0$ ($H_0 : \psi_0 = 0$), while ϕ is the vector of pure nuisance parameters. ϕ is restricted to be $\phi^\star$ when estimating the model, while $\phi = \phi^\star$ is not part of the null hypothesis. Then, even given null is true, $\phi = \phi^\star$ may not hold. For now, I consider the ideal condition, which means either $\psi = (\rho, \lambda, \gamma)$, $\phi = \emptyset$ where we are testing all the parameters of interest jointly; or ψ is a subset of (ρ, λ, γ), and $\phi = \psi^\star$ is correctly specified. The second scenario might seem too restrictive, while it can be viewed as a test for possible sub models under the SDPD framework. For example, when one is confident that there will be no spatial-time lag ($\phi = \rho = 0$), and wants to test whether there is possible time lag and spatial lag ($\psi = (\rho, \lambda) = 0$), this kind of tests can be applied.

Let $D_a(\theta) = \frac{1}{n^\star} \frac{\partial \ln L_{nT}(\theta)}{\partial a}$ and $D_{aa}(\theta) = \frac{1}{n^\star} \frac{\partial^2 \ln L_{nT}(\theta)}{\partial a \partial a'}$, where $a = \{\varphi, \psi, \phi\}$. As Proposition 1, $\Sigma(\theta) = E\left(-D_{\theta\theta}(\theta)\right)$, consider the following partitions:

$$\Sigma(\theta) = \begin{bmatrix} \Sigma_\varphi(\theta) & \Sigma_{\varphi\psi}(\theta) & \Sigma_{\varphi\phi}(\theta) \\ \Sigma_{\psi\varphi}(\theta) & \Sigma_\psi(\theta) & \Sigma_{\psi\phi}(\theta) \\ \Sigma_{\phi\varphi}(\theta) & \Sigma_{\phi\psi}(\theta) & \Sigma_\phi(\theta) \end{bmatrix}, \quad \Delta_{nT}(\theta) = \begin{bmatrix} \Delta_{nT,\varphi}(\theta) \\ \Delta_{nT,\psi}(\theta) \\ \Delta_{nT,\phi}(\theta) \end{bmatrix}, \qquad \text{for} \quad i = 2, 3. \tag{4.1}$$

Consider the local alternative $H^\psi_A : \psi_0 = \psi_\star + \zeta/\sqrt{n^\star}$, where ζ is bounded constant vector. The Taylor expansions for $\sqrt{n^\star} D_\psi(\tilde{\theta})$ and $\sqrt{n^\star} D_\varphi(\tilde{\theta})$ are:

$$\sqrt{n^\star} D_\psi(\tilde{\theta}) = \sqrt{n^\star} D_\psi(\theta_0) - D_{\psi\psi}(\theta_0)\zeta + \sqrt{n^\star} D_{\psi\varphi}(\theta_0)(\tilde{\varphi} - \varphi_0) + o_p(1), \tag{4.2}$$

$$\sqrt{n^\star} D_\varphi(\tilde{\theta}) = \sqrt{n^\star} D_\varphi(\theta_0) - D_{\varphi\psi}(\theta_0)\zeta + \sqrt{n^\star} D_{\varphi\varphi}(\theta_0)(\tilde{\varphi} - \varphi_0) + o_p(1). \tag{4.3}$$

Since $\tilde{\theta}$ is an MLE for φ under restricting $\psi = \psi_\star$, $\phi = \phi_\star$ $D_\varphi(\tilde{\theta}) = 0$. (5.2) and (5.3) would then imply:

$$\sqrt{n^\star} D_\psi(\tilde{\theta}) = \left[-\Sigma_{\psi\varphi}\Sigma_\varphi^{-1},\, I_r\right] \times \begin{bmatrix} \frac{1}{\sqrt{n^\star}} \frac{\partial \ln L_{nT}(\theta_0)}{\partial \varphi} \\ \frac{1}{\sqrt{n^\star}} \frac{\partial \ln L_{nT}(\theta_0)}{\partial \psi} \end{bmatrix} + \left[\Sigma_\psi - \Sigma_{\psi\varphi}\Sigma_\varphi^{-1}\Sigma_{\varphi\psi}\right]\zeta + o_p(1). \tag{4.4}$$

By Proposition 1,

$$\begin{bmatrix} \frac{1}{\sqrt{n^\star}} \frac{\partial \ln L_{nT}(\theta_0)}{\partial \varphi} \\ \frac{1}{\sqrt{n^\star}} \frac{\partial \ln L_{nT}(\theta_0)}{\partial \psi} \end{bmatrix} + \begin{bmatrix} \Delta_{nT,\varphi} \\ \Delta_{nT,\psi} \end{bmatrix} \xrightarrow{d} N\left(0_{(p+r)\times 1}, \begin{bmatrix} \Sigma_\varphi & \Sigma_{\varphi\psi} \\ \Sigma_{\psi\varphi} & \Sigma_\psi \end{bmatrix}\right). \tag{4.5}$$

(4.4) implies that:

$$\sqrt{n^\star}D_\psi(\tilde{\theta}) - \Sigma_{\psi\varphi}\Sigma_\varphi^{-1}\Delta_{nT,\varphi} + \Delta_{nT,\psi} \xrightarrow{d} N\left(\Sigma_{\psi\cdot\varphi}\zeta,\, \Sigma_{\psi\cdot\varphi}\right), \tag{4.6}$$

where $\Sigma_{\psi\cdot\varphi} = \left[\Sigma_\psi - \Sigma_{\psi\varphi}\Sigma_\varphi^{-1}\Sigma_{\varphi\psi}\right]$ and $\Sigma_{\psi\phi\cdot\varphi} = \left[\Sigma_{\psi\phi} - \Sigma_{\psi\varphi}\Sigma_\varphi^{-1}\Sigma_{\varphi\phi}\right]$. Let:

$$C_\psi(\tilde{\theta}) = D_\psi(\tilde{\theta}) - \frac{1}{\sqrt{n^\star}}\Sigma_{\psi\varphi}(\tilde{\theta})\Sigma_\varphi^{-1}(\tilde{\theta})\Delta_{nT,\varphi}(\tilde{\theta}) + \frac{1}{\sqrt{n^\star}}\Delta_{nT,\psi}(\tilde{\theta}), \tag{4.7}$$

such $C_\psi(\tilde{\theta})$ can be seen as the bias-corrected score function. Under H_A^ψ, we have

$$\sqrt{n^\star}C_\psi(\tilde{\theta}) \xrightarrow{d} N\left(\Sigma_{\psi\cdot\varphi}\zeta,\, \Sigma_{\psi\cdot\varphi}\right). \tag{4.8}$$

The LM test for $H_0 : \psi = 0$ is then given by:

$$LM_\psi = n^\star C_\psi'(\tilde{\theta})\Sigma_{\psi\cdot\varphi}^{-1}(\tilde{\theta})C_\psi(\tilde{\theta}), \tag{4.9}$$

With correct specification, LM_ψ is locally optimal and haswellknown asymptotic distributions under the null and a sequence of local alternatives, summarized as follows:

Proposition 2. *Under Assumptions 1(ii), 2(i), 3, 4, 5(ii) and $\frac{n}{T} \to k$, where $0 < k < \infty$, for direct approach; and under Assumptions 1(i), 2(i), 3, 4, 5(i) and $\frac{n}{T} < \infty$, for transformation approach, the following results hold.*

1. *Under H_0^ψ, we have*

$$LM_\psi \xrightarrow{d} \chi_r^2, \tag{4.10}$$

2. *Under H_A^ψ, we have*

$$LM_\psi \xrightarrow{d} \chi_r^2\left(\xi_1\right), \tag{4.11}$$

where $\zeta \neq 0$, and $\xi_1 = \zeta'\Sigma_{\psi\cdot\varphi}\zeta$ is the non-centrality parameter.

Proof. See Appendix C. □

Given ideal conditions, when null is true, the test follows a centeral chi-square distribution with r degrees of freedom. When the alternative is true the test statistic follows a non-centeral chi-square distribution and thus gives the power of the tests.

5 LM Tests under Distributional Misspecification

In previous section, normality of error terms is assumed. This assumption may not be true. In this section, I consider testing when one is not sure whether normality holds. For notational simplification, I use θ_0 to denote the pseudo true value of θ which minimizes Kullback–Leibler distance between the assumed normal distribution and the unknown true distribution of error terms (Kullback and Leibler, 1951). Consider the partition to the $\Omega(\theta)$ matrix

$$\Omega(\theta) = \begin{pmatrix} \Omega_{\varphi}(\theta) & \Omega_{\varphi\psi}(\theta) & \Omega_{\varphi\phi}(\theta) \\ \Omega_{\psi\varphi}(\theta) & \Omega_{\psi}(\theta) & \Omega_{\psi\phi}(\theta) \\ \Omega_{\phi\varphi}(\theta) & \Omega_{\phi\psi}(\theta) & \Omega_{\phi}(\theta) \end{pmatrix} \tag{5.1}$$

and let Ω denote $\Omega(\theta_0)$.
For previous section, where normality is assumed, we have

$$\frac{1}{\sqrt{n^{\star}}} \frac{\partial \ln L_{nT}(\theta_0)}{\partial \theta} + \Delta_{nT} \xrightarrow{d} N\left(0,\, \Sigma(\theta_0)\right), \tag{5.2}$$

and

$$\Sigma = \Omega. \tag{5.3}$$

When normality is misspecified, an immediate result is that the information matrix equation (5.3) does not hold. Under such QML setting, (5.3) should be, (White, 1982)

$$\frac{1}{\sqrt{n^{\star}}} \frac{\partial \ln L_{nT}(\theta_0)}{\partial \theta} + \Delta_{nT} \xrightarrow{d} N\left(0,\, \Omega(\theta_0)\right), \tag{5.4}$$

For this section, I still assume the restriction that $\varphi = \varphi_{\star}$ is true, so there is no parametric misspecification. However, the distribution of error is now allowed to be non-normal. Under the sequences of local alternatives $H_A^{\psi} : \psi_0 = \psi_{\star} + \zeta/\sqrt{n}$, and note that is still correct that $\Sigma(\theta_0) = E\left(-D_{\theta\theta}(\theta_0)\right)$, the Taylor expansion for $\sqrt{n^{\star}} D_{\psi}(\tilde{\theta})$ and $\sqrt{n^{\star}} D_{\varphi}(\tilde{\theta})$ are:

$$\sqrt{n^{\star}} D_{\psi}(\tilde{\theta}) = \sqrt{n^{\star}} D_{\psi}(\theta_0) - D_{\psi\psi}(\theta_0)\zeta + \sqrt{n^{\star}} D_{\psi\varphi}(\theta_0)(\tilde{\varphi} - \varphi_0) + o_p(1), \tag{5.5}$$
$$\sqrt{n^{\star}} D_{\varphi}(\tilde{\theta}) = \sqrt{n^{\star}} D_{\varphi}(\theta_0) - D_{\varphi\psi}(\theta_0)\zeta + \sqrt{n^{\star}} D_{\varphi\varphi}(\theta_0)(\tilde{\varphi} - \varphi_0) + o_p(1). \tag{5.6}$$

(5.5) and (5.6) imply:

$$\sqrt{n^{\star}} D_{\psi}(\tilde{\theta}) = \left[-\Sigma_{\psi\varphi}\Sigma_{\varphi}^{-1},\, I_r\right] \times \begin{bmatrix} \frac{1}{\sqrt{n^{\star}}} \frac{\partial \ln L_{nT}(\theta_0)}{\partial \varphi} \\ \frac{1}{\sqrt{n^{\star}}} \frac{\partial \ln L_{nT}(\theta_0)}{\partial \psi} \end{bmatrix} + \Sigma_{\psi\cdot\phi}\zeta + o_p(1). \tag{5.7}$$

Unlike under ideal conditions, the limiting distributions of score functions are now (5.4) instead of (4.5). Then, under H_A^{ψ},

$$\sqrt{n^{\star}} C_{\psi}(\tilde{\theta}) \xrightarrow{d} N\left(\Sigma_{\psi\cdot\varphi}\zeta, \mathcal{B}_{\psi\cdot\varphi}\right). \tag{5.8}$$

where $\mathcal{B}_{\psi\cdot\varphi} = \Omega_{\psi} - \Sigma_{\psi\varphi}\Sigma_{\varphi}^{-1}\Omega_{\varphi\psi} - \Omega_{\psi\varphi}\Sigma_{\varphi}^{-1}\Sigma_{\varphi\psi} + \Sigma_{\psi\varphi}\Sigma_{\varphi}^{-1}\Omega_{\varphi}\Sigma_{\varphi}^{-1}\Sigma_{\varphi\psi}$. Compare (4.8) with (5.8),

the distributional misspecification leads to a change in asymptotic variance of the $C(\tilde{\theta})$. Thus, the previous LM test statistic will be based on the wrong variance, and is invalid. However, when there is no distributional misspecification, $\Omega = \Sigma$, and we have $\mathcal{B}_{\psi\cdot\varphi} = \Sigma_{\psi\cdot\varphi}$. The variance degenerates to the ideal condition case which is as expected.
The LM test for $H_0 : \psi = 0$ that is robust to distributional misspecification is expressed by:

$$LM_{\psi}(D) = n^{\star} C'_{\psi}(\tilde{\theta}) \mathcal{B}^{-1}_{\psi\cdot\varphi}(\tilde{\theta}) C_{\psi}(\tilde{\theta}), \tag{5.9}$$

where $C_{\psi}(\tilde{\theta})$ is defined as in (4.8) and $\mathcal{B}_{\psi\cdot\varphi}(\tilde{\theta}) = \Omega_{\psi}(\tilde{\theta}) - \Sigma_{\psi\varphi}(\tilde{\theta})\Sigma_{\varphi}^{-1}(\tilde{\theta})\Omega_{\varphi\psi}(\tilde{\theta}) - \Omega_{\psi\varphi}(\tilde{\theta})\Sigma_{\varphi}^{-1}(\tilde{\theta})\Sigma_{\varphi\psi}(\tilde{\theta}) + \Sigma_{\psi\varphi}(\tilde{\theta})\Sigma_{\varphi}^{-1}(\tilde{\theta})\Omega_{\varphi}(\tilde{\theta})\Sigma_{\varphi}^{-1}(\tilde{\theta})\Sigma_{\varphi\psi}(\tilde{\theta})$. The results for this section are summarized as follows:

Proposition 3. *Under Assumptions 1(ii), 2(ii), 3, 4, 5(ii) and $\frac{n}{T} \to k$ for direct approach; and under Assumptions 1(i), 2(ii), 3, 4, 5(i) and $\frac{n}{T} < \infty$ for transformation approach, the following results hold for direct approach.*

1. *Under the distributional misspecification, the results in Proposition 2 are invalid. That is, LM_{ψ} does not has an asymptotic chi-square distribution.*
2. *Under H_0^{ψ}, we have*

$$LM_{\psi}(D) \xrightarrow{d} \chi^2_r, \tag{5.10}$$

3. *Under H_A^{ψ}, we have*

$$LM_{\psi}(D) \xrightarrow{d} \chi^2_r(\xi_2), \tag{5.11}$$

where $\zeta \neq 0$, and $\xi_2 = \zeta' \Sigma'_{\psi\cdot\varphi} \mathcal{B}^{-1}_{\psi\cdot\varphi} \Sigma_{\psi\cdot\varphi} \zeta$ is the non-centrality parameter.

Proof. See Appendix C. □

Comparing with LM_{ψ} in (4.9), $LM_{\psi}(D)$ is now weighted by $\mathcal{B}_{\psi\cdot\varphi}(\tilde{\theta})$ and thus have an asymptotic chi-square distribution when distributional misspecification is at presence. When there is no such distributional misspecification, we have $\Omega = \Sigma$, and thus $\mathcal{B}(\tilde{\theta}) = \Sigma(\tilde{\theta})$ asymptotically. $LM_{\psi}(D)$ will be asymptotically equivalent to LM_{ψ}.

6 LM Tests under Parametric Misspecification

In this section, I consider constructing the robust LM statistics with only parametric misspecification. Within this section I assume there is no distributional misspecification, and consider a local parametric misspecification i.e. $\phi_0 = \phi_{\star} + \delta/\sqrt{n^{\star}}$ where δ is bounded constant vector.
I provide a review of Bera et al. (2019) with a compact and simpler exposition. I investigate the asymptotic distribution of LM_{ψ} under the sequences of local alternatives $H_A^{\psi} : \psi_0 = \psi_{\star} + \zeta/\sqrt{n^{\star}}$

and $H_A^{\phi}: \phi_0 = \phi_\star + \delta/\sqrt{n^\star}$. The first-order Taylor expansions of scores $D_\psi(\tilde{\theta})$, $D_\varphi(\tilde{\theta})$ and $D_\phi(\tilde{\theta})$ are:

$$\sqrt{n^\star} D_\psi(\tilde{\theta}) = \sqrt{n^\star} D_\psi(\theta_0) - D_{\psi\psi}(\theta_0)\zeta - D_{\psi\phi}(\theta_0)\delta + \sqrt{n^\star} D_{\psi\varphi}(\theta_0)(\tilde{\varphi} - \varphi_0) + o_p(1), \tag{6.1}$$

$$\sqrt{n^\star} D_\varphi(\tilde{\theta}) = \sqrt{n^\star} D_\varphi(\theta_0) - D_{\varphi\psi}(\theta_0)\zeta - D_{\varphi\phi}(\theta_0)\delta + \sqrt{n^\star} D_{\varphi\varphi}(\theta_0)(\tilde{\varphi} - \varphi_0) + o_p(1). \tag{6.2}$$

$$\sqrt{n^\star} D_\phi(\tilde{\theta}) = \sqrt{n^\star} D_\phi(\theta_0) - D_{\phi\psi}(\theta_0)\zeta - D_{\phi\phi}(\theta_0)\delta + \sqrt{n^\star} D_{\phi\varphi}(\theta_0)(\tilde{\varphi} - \varphi_0) + o_p(1). \tag{6.3}$$

Solving the equations, we have the following results:

$$\sqrt{n^\star} D_\psi(\tilde{\theta}) = \left[-\Sigma_{\psi\varphi}\Sigma_\varphi^{-1},\, I_r\right] \times \begin{bmatrix} \frac{1}{\sqrt{n^\star}} \frac{\partial \ln L_{nT}(\theta_0)}{\partial \varphi} \\ \frac{1}{\sqrt{n^\star}} \frac{\partial \ln L_{nT}(\theta_0)}{\partial \psi} \end{bmatrix} + \Sigma_{\psi\cdot\varphi}\zeta + \Sigma_{\psi\phi\cdot\varphi}\delta + o_p(1). \tag{6.4}$$

$$\sqrt{n^\star} D_\phi(\tilde{\theta}) = \left[-\Sigma_{\phi\varphi}\Sigma_\varphi^{-1},\, I_s\right] \times \begin{bmatrix} \frac{1}{\sqrt{n^\star}} \frac{\partial \ln L_{nT}(\theta_0)}{\partial \varphi} \\ \frac{1}{\sqrt{n^\star}} \frac{\partial \ln L_{nT}(\theta_0)}{\partial \psi} \end{bmatrix} + \Sigma_{\phi\psi\cdot\varphi}\zeta + \Sigma_{\phi\cdot\varphi}\delta + o_p(1). \tag{6.5}$$

Where $\Sigma_{\psi\phi\cdot\varphi}(\theta) = \Sigma_{\psi\phi}(\theta) - \Sigma_{\psi\varphi}(\theta)\Sigma_\varphi^{-1}(\theta)\Sigma_{\varphi\phi}(\theta)$, $\Sigma_{\phi\psi\cdot\varphi}(\theta) = \Sigma_{\phi\psi}(\theta) - \Sigma_{\phi\varphi}(\theta)\Sigma_\varphi^{-1}(\theta)\Sigma_{\varphi\psi}(\theta)$, and $\Sigma_{\phi\cdot\varphi}(\theta) = \Sigma_\phi(\theta) - \Sigma_{\phi\varphi}(\theta)\Sigma_\varphi^{-1}(\theta)\Sigma_{\varphi\phi}(\theta)$. By asymptotic normality of the score function in Proposition 1,

$$\sqrt{n^\star} D_\psi(\tilde{\theta}) - \Sigma_{\psi\varphi}\Sigma_\varphi^{-1}\Delta_{nT,\varphi} + \Delta_{nT,\psi} \xrightarrow{d} N\left(\Sigma_{\psi\cdot\varphi}\zeta + \Sigma_{\psi\phi\cdot\varphi}\delta,\, \Sigma_{\psi\cdot\varphi}\right), \tag{6.6}$$

and

$$\sqrt{n^\star} C_\psi(\tilde{\theta}) \xrightarrow{d} N\left(\Sigma_{\psi\cdot\varphi}\zeta + \Sigma_{\psi\phi\cdot\varphi}\delta,\, \Sigma_{\psi\cdot\varphi}\right). \tag{6.7}$$

Again, compare (6.7) with (4.8) and (5.8). Different from distributional misspecification case, the asymptotic variance remains the same as under ideal conditions, while the parametric misspecification effects the asymptotic mean of the bias corrected score function. Now there is an additional $\Sigma_{\psi\phi\cdot\varphi}\delta$ term. When $\Sigma_{\psi\phi\cdot\varphi}$ is 0, the means of bias corrected score functions will be exactly the same as under ideal condition. When it is not 0, LM_ψ will over reject the null hypothesis. To further robustify the test, I adjust the score functions and get rid of the additional δ term no matter it is 0 or not, and we then have a LM test statistic robust to parametric misspecification.
Define adjusted score function for the nuisance parameters:

$$C_\phi(\tilde{\theta}) = D_\phi(\tilde{\theta}) - \frac{1}{\sqrt{n^\star}}\Sigma_{\phi,\varphi}(\tilde{\theta})\Sigma_\varphi^{-1}(\tilde{\theta})\Delta_{nT\varphi}(\tilde{\theta}) \; + \frac{1}{\sqrt{n^\star}}\Delta_{nT,\phi}(\tilde{\theta}), \tag{6.8}$$

(6.5) and Proposition 1 imply:

$$\sqrt{n^\star} C_\phi(\tilde{\theta}) \xrightarrow{d} N\left(\Sigma_{\phi\psi\cdot\varphi}\zeta + \Sigma_{\phi\cdot\varphi}\delta, \Sigma_{\phi\cdot\varphi}\right). \tag{6.9}$$

Using (6.9), the δ term can be cancelled from (6.7). In that way, the adjusted score function that has a 0 mean under the null with or without the presence of parametric misspeficication is:

$$C_{\psi}^{\star}(\tilde{\theta}) = C_{\psi}(\tilde{\theta}) - \Sigma_{\psi\phi\cdot\varphi}(\tilde{\theta})\Sigma_{\phi\cdot\varphi}^{-1}(\tilde{\theta})C_{\phi}(\tilde{\theta}) \tag{6.10}$$

It is obvious that $E\left(C_{\psi}^{\star}(\tilde{\theta})\right) = \left[\Sigma_{\psi\cdot\varphi} - \Sigma_{\psi\phi\cdot\varphi}\Sigma_{\phi\cdot\varphi}^{-1}\Sigma_{\phi\psi\cdot\varphi}\right]\zeta$ irrespective of value of δ. Calculation of variance of $C_{\psi}^{\star}(\tilde{\theta})$ will be shown in proof of the following Proposition 4.

$$\sqrt{n^{\star}}C_{\psi}^{\star}(\tilde{\theta}) \xrightarrow{d} N\left([\Sigma_{\psi\cdot\varphi} - \Sigma_{\psi\phi\cdot\varphi}\Sigma_{\phi\cdot\varphi}^{-1}\Sigma_{\phi\psi\cdot\varphi}]\zeta,\, \Sigma_{\psi\cdot\varphi} - \Sigma_{\psi\phi\cdot\varphi}\Sigma_{\phi\cdot\varphi}^{-1}\Sigma_{\phi\psi\cdot\varphi}\right). \tag{6.11}$$

The $LM_{\psi}(P)$ test that is robust to parametric misspeficication is then constructed based on (6.11). The asymptotic properties of the test is summarized in the following proposition.

Proposition 4. . *Under Assumptions 1(ii), 2(i), 3, 4, 5(ii) and* $\frac{n}{T} \to k$*, where* $0 < k < \infty$*; and under Assumptions 1(i), 2(ii), 3, 4, 5(i) and* $\frac{n}{T} < \infty$ *for transformation approach, the following results hold.*

1. *Under* H_A^{ψ} *and* H_A^{ϕ}*, we have*

$$LM_{\psi} \xrightarrow{d} \chi_r^2(\xi_3), \tag{6.12}$$

where $\xi_3 = \zeta'\Sigma_{\psi\cdot\varphi}\zeta + 2\zeta'\Sigma_{\psi\phi\cdot\varphi}\delta + \delta'\Sigma'_{\psi\phi\cdot\varphi}\Sigma_{\psi\cdot\varphi}^{-1}\Sigma_{\psi\phi\cdot\varphi}\delta$ *is the non-centrality parameter.*

2. *Under* H_0^{ψ} *and irrespective of whether* H_0^{ϕ} *or* H_A^{ϕ} *holds, let* $C_{\psi}^{\star}(\tilde{\theta}) = C_{\psi}(\tilde{\theta}) - \Sigma_{\psi\phi\cdot\varphi}(\tilde{\theta})\Sigma_{\phi\cdot\varphi}^{-1}(\tilde{\theta})C_{\phi}(\tilde{\theta})$*, we have*

$$LM_{\psi}(P) = n^{\star}C_{\psi}^{*\prime}(\tilde{\theta})\left[\Sigma_{\psi\cdot\varphi}(\tilde{\theta}) - \Sigma_{\psi\phi\cdot\varphi}(\tilde{\theta})\Sigma_{\phi\cdot\varphi}^{-1}(\tilde{\theta})\Sigma_{\phi\psi\cdot\varphi}(\tilde{\theta})\right]^{-1}C_{\psi}^{*}(\tilde{\theta}) \xrightarrow{d} \chi_r^2, \tag{6.13}$$

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

$$LM_{\psi}(P) = n^{\star}C_{\psi}^{*\prime}(\tilde{\theta})\left[\Sigma_{\psi\cdot\varphi}(\tilde{\theta}) - \Sigma_{\psi\phi\cdot\varphi}(\tilde{\theta})\Sigma_{\phi\cdot\varphi}^{-1}(\tilde{\theta})\Sigma_{\phi\psi\cdot\varphi}(\tilde{\theta})\right]^{-1}C_{\psi}^{*}(\tilde{\theta}) \xrightarrow{d} \chi_r^2(\xi_4), \tag{6.14}$$

where $\xi_4 = \zeta'\left(\Sigma_{\psi\cdot\varphi} - \Sigma_{\psi\phi\cdot\varphi}\Sigma_{\phi\cdot\varphi}^{-1}\Sigma_{\phi\phi\cdot\varphi}\right)\zeta$ *is the non-centrality parameter.*

Proof. See Appendix C. □

First, under null $\zeta = 0$ and with parametric misspecification $\delta \neq 0$, we have $\xi_3 = \delta'\Sigma'_{\psi\phi\cdot\varphi}\Sigma_{\psi\cdot\varphi}^{-1}\Sigma_{\psi\phi\cdot\varphi}\delta$. The LM_{ψ} test follows a non-central chi-square distribution and will then have wrong size; $LM_{\psi}(P)$ test follows central chi-square distribution under H_0^{ψ} regardless of whether $\delta = 0$ or not. By Bera and Bilias (2001), Bera and Yoon (1993) this test shares properties of Neyman's $C(\alpha)$ test under the null and local alternatives. When there is no parametric misspecification $\delta = 0$, $\xi_3 - \xi_4 = \zeta'\left(\Sigma_{\psi\phi\cdot\varphi}\Sigma_{\phi\cdot\varphi}^{-1}\Sigma_{\phi\psi\cdot\varphi}\right)\zeta \geq 0$. LM_{ψ} has a higher power than $LM_{\psi}(P)$.

7 LM Tests under Both Distributional and Parametric Misspecifications

In this section, I assume the existence of both distributional and parametric misspecifications. Again, the model is estimated $\tilde{\theta}$ under restrictions $H_0 : \psi_0 = \psi^\star$ and $\phi_0 = \psi^\star$. I consider local alternatives: $H_A^\psi : \psi_0 = \psi_\star + \zeta/\sqrt{n^\star}$ and $H_A^\phi : \phi_0 = \phi_\star + \delta/\sqrt{n^\star}$. By first-order Taylor expansions of scores $D_\psi(\tilde{\theta})$, $D_\varphi(\tilde{\theta})$ and $D_\phi(\tilde{\theta})$, we will have the same result as in Section 6:

$$\sqrt{n^\star} D_\psi(\tilde{\theta}) = \left[-\Sigma_{\psi\varphi}\Sigma_\varphi^{-1},\, I_r\right] \times \begin{bmatrix} \frac{1}{\sqrt{n^\star}} \frac{\partial \ln L_{nT}(\theta_0)}{\partial \varphi} \\ \frac{1}{\sqrt{n^\star}} \frac{\partial \ln L_{nT}(\theta_0)}{\partial \psi} \end{bmatrix} + \Sigma_{\psi\cdot\varphi}\zeta + \Sigma_{\psi\phi\cdot\varphi}\delta + o_p(1). \tag{7.1}$$

$$\sqrt{n^\star} D_\phi(\tilde{\theta}) = \left[-\Sigma_{\phi\varphi}\Sigma_\varphi^{-1},\, I_s\right] \times \begin{bmatrix} \frac{1}{\sqrt{n^\star}} \frac{\partial \ln L_{nT}(\theta_0)}{\partial \varphi} \\ \frac{1}{\sqrt{n^\star}} \frac{\partial \ln L_{nT}(\theta_0)}{\partial \psi} \end{bmatrix} + \Sigma_{\phi\psi\cdot\varphi}\zeta + \Sigma_{\phi\cdot\varphi}\delta + o_p(1). \tag{7.2}$$

The asymptotic distribution of $\frac{1}{\sqrt{n^\star}} \frac{\partial \ln L_{nT}(\theta_0)}{\partial \theta}$ is the same as in Section 5, then for the bias corrected function,

$$\sqrt{n^\star} C_\psi(\tilde{\theta}) \xrightarrow{d} N\left(\Sigma_{\psi\cdot\varphi}\zeta + \Sigma_{\psi\phi\cdot\varphi}\delta,\, \mathcal{B}_{\psi\cdot\varphi}\right). \tag{7.3}$$

$$\sqrt{n^\star} C_\phi(\tilde{\theta}) \xrightarrow{d} N\left(\Sigma_{\phi\psi\cdot\varphi}\zeta + \Sigma_{\phi\cdot\varphi}\delta, \mathcal{B}_{\phi\cdot\varphi}\right). \tag{7.4}$$

where $\mathcal{B}_{\phi\cdot\varphi} = \Omega_\phi - \Sigma_{\phi\varphi}\Sigma_\varphi^{-1}\Omega_{\varphi\phi} - \Omega_{\phi\varphi}\Sigma_\varphi^{-1}\Sigma_{\varphi\phi} + \Sigma_{\phi\varphi}\Sigma_\varphi^{-1}\Omega_\varphi\Sigma_\varphi^{-1}\Sigma_{\varphi\phi}$. The result shows the effects of each type of misspecification simultaneously on the bias-corrected score function. The parametric misspecification changes the mean, and distributional misspecification changes the variance in the same way as in previous sections. Then, construct the $C_\psi^\star(\tilde{\theta})$ adjusted score to cancel the effects of parametric misspecification:

$$C_\psi^\star(\tilde{\theta}) = C_\psi(\tilde{\theta}) - \Sigma_{\psi\phi\cdot\varphi}(\tilde{\theta})\Sigma_{\phi\cdot\varphi}^{-1}(\tilde{\theta})C_\phi(\tilde{\theta}) \tag{7.5}$$

In fact, (7.5) and (6.10) are the same. The expectation is still $[\Sigma_{\psi\cdot\varphi} - \Sigma_{\psi\phi\cdot\varphi}\Sigma_{\phi\cdot\varphi}^{-1}\Sigma_{\phi\psi\cdot\varphi}]\zeta$. The distributional misspecification does not affect the form or mean of adjusted score function. However, the asymptotic variance will be different from (6.11). The asymptotic variance of the adjusted score function is (see proof of the following Proposition 5):

$$\mathcal{D}_{\psi\cdot\varphi} = \mathcal{B}_{\psi\cdot\varphi} - \Sigma_{\psi\phi\cdot\varphi}\Sigma_{\phi\cdot\varphi}^{-1}\mathcal{B}_{\phi\psi\cdot\varphi} - \mathcal{B}_{\psi\phi\cdot\varphi}\Sigma_{\phi\cdot\varphi}^{-1}\Sigma_{\phi\psi\cdot\varphi} + \Sigma_{\psi\phi\cdot\varphi}\Sigma_{\phi\cdot\varphi}^{-1}\mathcal{B}_{\phi\cdot\varphi}\Sigma_{\phi\cdot\varphi}^{-1}\Sigma_{\phi\psi\cdot\varphi} \tag{7.6}$$

where $\mathcal{B}_{\psi\phi\cdot\varphi} = \Omega_{\psi\phi} - \Sigma_{\psi\varphi}\Sigma_\varphi^{-1}\Omega_{\varphi\phi} - \Omega_{\psi\varphi}\Sigma_\varphi^{-1}\Sigma_{\varphi\phi} + \Sigma_{\psi\varphi}\Sigma_\varphi^{-1}\Omega_\varphi\Sigma_\varphi^{-1}\Sigma_{\varphi\phi}$ and $\mathcal{B}_{\phi\psi\cdot\varphi}$ is similarly defined. All the above notations denote the terms evaluated at $\theta = \theta_0$.
Based on (7.5) and (7.6), the test $LM_\psi(DP)$ that is both robust to parametric and distributional misspecifications is constructed. The asymptotic properties of the proposed $LM_\psi(DP)$ statistic are provided in the following proposition.

Proposition 5. . *Under Assumptions 1(ii), 2(ii), 3, 4, 5(ii) and* $\frac{n}{T} \to k$*, where* $0 < k < \infty$*;*

and under Assumptions 1(i), 2(ii), 3, 4, 5(i) and $\frac{n}{T} < \infty$ for transformation approach, the following results hold.

1. *Under H_0^{ψ} and irrespective of whether H_0^{ϕ} or H_A^{ϕ} holds, let $C_{\psi}^{\star}(\tilde{\theta}) = C_{\psi}(\tilde{\theta}) - \Sigma_{\psi\phi\cdot\varphi}(\tilde{\theta})\Sigma_{\phi\cdot\varphi}^{-1}(\tilde{\theta})C_{\phi}(\tilde{\theta})$, we have*

$$LM_{\psi}(DP) = n^{\star}C_{\psi}^{*\prime}(\tilde{\theta})\mathcal{D}_{\psi\cdot\varphi}^{-1}(\tilde{\theta})C_{\psi}^{*}(\tilde{\theta}) \xrightarrow{d} \chi_r^2, \tag{7.7}$$

where

$$\mathcal{D}_{\psi\cdot\varphi}(\tilde{\theta}) = \mathcal{B}_{\psi\cdot\varphi}(\tilde{\theta}) - \Sigma_{\psi\phi\cdot\varphi}(\tilde{\theta})\Sigma_{\phi\cdot\varphi}^{-1}(\tilde{\theta})\mathcal{B}_{\phi\psi\cdot\varphi}(\tilde{\theta}) - \mathcal{B}_{\psi\phi\cdot\varphi}(\tilde{\theta})\Sigma_{\phi\cdot\varphi}^{-1}(\tilde{\theta})\Sigma_{\phi\psi\cdot\varphi}(\tilde{\theta}) \\ +\Sigma_{\psi\phi\cdot\varphi}(\tilde{\theta})\Sigma_{\phi\cdot\varphi}^{-1}(\tilde{\theta})\mathcal{B}_{\phi\cdot\varphi}(\tilde{\theta})\Sigma_{\phi\cdot\varphi}^{-1}(\tilde{\theta})\Sigma_{\phi\psi\cdot\varphi}(\tilde{\theta}) \tag{7.8}$$

$$\mathcal{B}_{\psi\cdot\varphi}(\tilde{\theta}) = \Omega_{\psi}(\tilde{\theta}) - \Sigma_{\psi\varphi}(\tilde{\theta})\Sigma_{\varphi}^{-1}(\tilde{\theta})\Omega_{\varphi\psi}(\tilde{\theta}) - \Omega_{\psi\varphi}(\tilde{\theta})\Sigma_{\varphi}^{-1}(\tilde{\theta})\Sigma_{\varphi\psi}(\tilde{\theta}) \\ +\Sigma_{\psi\varphi}(\tilde{\theta})\Sigma_{\varphi}^{-1}(\tilde{\theta})\Omega_{\varphi}(\tilde{\theta})\Sigma_{\varphi}^{-1}(\tilde{\theta})\Sigma_{\varphi\psi}(\tilde{\theta}) \tag{7.9}$$

$$\mathcal{B}_{\psi\phi\cdot\varphi}(\tilde{\theta}) = \Omega_{\psi\phi}(\tilde{\theta}) - \Sigma_{\psi\varphi}(\tilde{\theta})\Sigma_{\varphi}^{-1}(\tilde{\theta})\Omega_{\varphi\phi}(\tilde{\theta}) - \Omega_{\psi\varphi}(\tilde{\theta})\Sigma_{\varphi}^{-1}(\tilde{\theta})\Sigma_{\varphi\phi}(\tilde{\theta}) \\ +\Sigma_{\psi\varphi}(\tilde{\theta})\Sigma_{\varphi}^{-1}(\tilde{\theta})\Omega_{\varphi}(\tilde{\theta})\Sigma_{\varphi}^{-1}(\tilde{\theta})\Sigma_{\varphi\phi}(\tilde{\theta}) \tag{7.10}$$

and $\mathcal{B}_{\phi\psi\cdot\varphi}(\tilde{\theta})$ is similarly defined.

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

$$LM_{\psi}(DP) = n^{\star}C_{\psi}^{*\prime}(\tilde{\theta})\mathcal{D}_{\psi\cdot\varphi}^{-1}(\tilde{\theta})C_{\psi}^{*}(\tilde{\theta}) \xrightarrow{d} \chi_r^2(\xi_5) \tag{7.11}$$

where $\xi_5 = \zeta'\left(\Sigma_{\psi\cdot\varphi} - \Sigma_{\psi\phi\cdot\varphi}\Sigma_{\phi\cdot\varphi}^{-1}\Sigma_{\phi\psi\cdot\varphi}\right)'\mathcal{D}_{\psi\cdot\varphi}^{-1}\left(\Sigma_{\psi\cdot\varphi} - \Sigma_{\psi\phi\cdot\varphi}\Sigma_{\phi\cdot\varphi}^{-1}\Sigma_{\phi\psi\cdot\varphi}\right)\zeta$ is the non-centrality parameter.

Proof. See Appendix C. □

The expression of $LM_{\psi}(DP)$ is lengthy, while since it only requires the estimation of $\tilde{\theta}$, and for the SDPD model, the spatial lag, time lag and spatial-time lag are restricted to be 0. The estimation for $\tilde{\theta}$ is actually straight forward, and then makes $LM_{\psi}(DP)$ easy to compute.
By comparing Proposition 5 with Proposition 4 and Proposition 3, the adjustment for score is still the same as $LM_{\psi}(P)$ while when null is wrong, the non-centrality parameter is now affected by the distributional misspecification.

Corollary 1. *For SDPD model, the $LM_{\psi}(DP)$ simplifies in the following ways:*

1. *When ψ is any combination of (γ, ρ) we have:*

$$LM_{\psi}(D) = LM_{\psi} \tag{7.12}$$

2. *When ψ is any combination of (γ, ρ, λ), $n \to \infty$, and assume number of neighbors for each individual grows slower than $\sqrt{n}$ or is bounded, we have:*

$$LM_{\psi}(DP) \to LM_{\psi}(P) \tag{7.13}$$

$$LM_{\psi}(D) \to LM_{\psi} \tag{7.14}$$

Proof. See Appendix C. □

Here, the assumption on the number of neighbors can be seen as a restriction on W_n. The weight matrix should be sparse for the corollary to be valid. Since the corollary only requires the number of neighbors grows at slower rate than $\sqrt{n}$, it is satisfied in general.

By previous propositions, $LM_{\psi}(DP)$ and $LM_{\psi}(D)$ are robust to distributional misspecification. With the above corollary, we can conlude that for such SDPD model, the test that does not account for distributional misspecification will be quite robust to non-normal distributed disturbance terms. The simulation result shows that with relatively small number of individuals n, the effect of distributional misspecification to the test statistics will be subtle. Then, it will be safe to use the tests that are derived under MLE framework even in the non-normal case.

8 The Test Statistics

In this section, I mainly use Proposition 5 and Corollary 1 to derive test statistics for the following hypotheses when the normality is not guaranteed:

1. $H_0 : \rho_0 = \gamma_0 = \lambda_0 = 0$,
2. $H_0^{\lambda} : \lambda_0 = 0$ in the presence of γ_0 and ρ_0,
3. $H_0^{\gamma} : \gamma_0 = 0$ in the presence of λ_0 and ρ_0,
4. $H_0^{\rho} : \rho_0 = 0$ in the presence of λ_0 and γ_0.

A specification search may start with testing the joint null $H_0 : \rho_0 = \gamma_0 = \lambda_0 = 0$. If the joint null is accept, then it can be concluded that there is no need to use the SDPD model, while when the joint null is rejected, at least one of the three parameters is significant. We will then need to test H_0^{λ}, H_0^{γ}, and H_0^{ρ} to detect the source of rejection of the joint null. Let $\kappa = (\gamma,\, \rho,\, \lambda)'$ be the vector of testing parameters and $\tilde{\theta} = \left(0, 0, \tilde{\beta}', 0, \tilde{\sigma}^2\right)'$ be the restricted QMLE under the joint null $H_0 : \kappa_0 = 0$. The computation of all test statistics require $\tilde{\theta}$ only. Under the joint null $H_0 : \kappa_0 = 0$, (2.1) degenerates to:

$$Y_{nt} = X_{nt}\tilde{\beta} + \tilde{\mathbf{c}}_n + \tilde{\alpha}_t l_n + V_{nt}, \tag{8.1}$$

with the log likelihood function (after concentrating out fixed effects):

$$\ln L^d_{nT}(\tilde{\theta}) = -\frac{nT}{2}\ln 2\pi - \frac{nT}{2}\ln\tilde{\sigma}^2 - \frac{1}{2\tilde{\sigma}^2}\sum_{t=1}^{T}\tilde{V}'_{nt}(\theta)J_n\tilde{V}_{nt}(\theta), \tag{8.2}$$

where $\tilde{V}_{nt}(\theta) = \tilde{Y}_{nt} - \tilde{X}_{nt}\beta$. Then solving for restricted QMLE yields

$$\tilde{\beta} = \left(\sum_{t=1}^{T}\tilde{X}'_{nt}J_n\tilde{X}_{nt}\right)^{-1}\left(\sum_{t=1}^{T}\tilde{X}'_{nt}J_n\tilde{Y}_{nt}\right), \tag{8.3}$$

$$\tilde{\sigma}^2 = \frac{1}{nT}\sum_{t=1}^{T}\tilde{V}'_{nt}(\tilde{\theta})J_n\tilde{V}_{nt}(\tilde{\theta}), \tag{8.4}$$

where $\tilde{V}_{nt}(\tilde{\theta}) = \tilde{Y}_{nt} - \tilde{X}_{nt}\tilde{\beta}$.

Thus, the computation of my suggested test statistics does not require any nonlinear optimization or the application of the numerical search techniques. $\tilde{\theta}$ can also be computed with transformation approach. The likelihood function for transformation approach after concentrating out individual fixed effect and under joint null is:

$$\ln L_{nT}(\theta) = -\frac{(n-1)T}{2}\ln 2\pi - \frac{(n-1)T}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\tilde{V}'_{nt}(\theta)J_n\tilde{V}_{nt}(\theta). \tag{8.5}$$

It implies,

$$\tilde{\beta} = \left(\sum_{t=1}^{T}\tilde{X}'_{nt}J_n\tilde{X}_{nt}\right)^{-1}\left(\sum_{t=1}^{T}\tilde{X}'_{nt}J_n\tilde{Y}_{nt}\right), \tag{8.6}$$

$$\tilde{\sigma}^2 = \frac{1}{(n-1)T}\sum_{t=1}^{T}\tilde{V}'_{nt}(\tilde{\theta})J_n\tilde{V}_{nt}(\tilde{\theta}), \tag{8.7}$$

The $\tilde{\beta}$'s in (8.3) and (8.6) are exactly the same. Estimations for $\tilde{\sigma}^2$ in (8.4) and (8.7) are slightly different. First, both estimators are $\sqrt{nT}$ consistent, then by Neyman (1979), using either will result in the same asymptotic distribution for the suggested tests. Second, the model reduces to $Y_{nt} = X_{nt}\beta + \mathbf{c}_n + \alpha_t l_n + V_{nt}$ under the joint null hypothesis. The correct degrees of freedom for this model is $(n-1)(T-1)-k$, which can be used in the estimation of $\tilde{\sigma}^2$. The simulation results show that this adjustment can provide better results in finite samples comparing with either (8.4) or (8.7).

Terms for calculating all LM test statistics can be found at Appendix A. Start with the joint null

$H_0 : \kappa = 0$. Let $\mathbf{C}_\kappa(\tilde{\theta}) = \left(C_\gamma'(\tilde{\theta}),\, C_\rho'(\tilde{\theta}),\, C_\lambda'(\tilde{\theta})\right)'$ denote the vector of bias-corrected score function. By Proposition 2 and Proposition 3 we have two LM test statistics[4]:

$$LM_\kappa = n^\star \mathbf{C}_\kappa'(\tilde{\theta})\Sigma_{\kappa\cdot\varphi}^{-1}(\tilde{\theta})\mathbf{C}_\kappa(\tilde{\theta}) \tag{8.8}$$

$$LM_\kappa(D) = n^\star \mathbf{C}_\kappa'(\tilde{\theta})\mathcal{B}_{\kappa\cdot\varphi}^{-1}(\tilde{\theta})\mathbf{C}_\kappa(\tilde{\theta}) \tag{8.9}$$

The asymptotic null distributions are both χ_3^2, while $LM_\kappa(D)$ is robust to distributional misspecification comparing to LM_κ. For such joint test, there is no nuisance parameter, so the tests will not suffer from parametric misspecification problem. $LM_\kappa(D)$ takes care of distributional misspecification, but is equivalent to LM_κ with large n.
Let's continue considering testing one of parameters λ, γ, and ρ in the presence of the other two. I begin with the testing of $H_0^\gamma : \rho_0 = 0$. Testing for H_0^λ and H_0^γ will be similar. In my notations, $\psi = \rho$ and $\phi = (\lambda,\, \gamma)'$ for testing H_0^γ. Then, by Proposition 2, the one directional test in this case is stated as

$$LM_\rho = n^\star C_\rho'(\tilde{\theta})\Sigma_{\rho\cdot\varphi}^{-1}(\tilde{\theta})C_\rho(\tilde{\theta}), \tag{8.10}$$

where $\Sigma_{\rho\cdot\varphi}(\tilde{\theta}) = \left[\Sigma_\rho(\tilde{\theta}) - \Sigma_{\rho\varphi}(\tilde{\theta})\Sigma_\varphi^{-1}(\tilde{\theta})\Sigma_{\varphi\rho}(\tilde{\theta})\right]$. By Proposition 3, we have:

$$LM_\rho(D) = n^\star C_\rho'(\tilde{\theta})\mathcal{B}_{\rho\cdot\varphi}^{-1}(\tilde{\theta})C_\rho(\tilde{\theta}), \tag{8.11}$$

where $\mathcal{B}_{\rho\cdot\varphi}(\tilde{\theta}) = \Omega_\rho(\tilde{\theta}) - \Sigma_{\rho\varphi}(\tilde{\theta})\Sigma_\varphi^{-1}(\tilde{\theta})\Omega_{\varphi\rho}(\tilde{\theta}) - \Omega_{\rho\varphi}(\tilde{\theta})\Sigma_\varphi^{-1}(\tilde{\theta})\Sigma_{\varphi\rho}(\tilde{\theta}) + \Sigma_{\rho\varphi}(\tilde{\theta})\Sigma_\varphi^{-1}(\tilde{\theta})\Omega_\varphi(\tilde{\theta})\Sigma_\varphi^{-1}(\tilde{\theta})\Sigma_{\varphi\rho}(\tilde{\theta})$. By Corollary 1, (C.18) and (8.11) are identical, which means the test that derived under ideal condition will still be valid with non-normal distributed error disturbance terms. Such tests will be invalid when (λ_0, γ_0) deviates from $(0, 0)$. Then, construct the test by Proposition 4:

$$LM_\rho(P) = n^\star C_\rho^{*'}(\tilde{\theta})\left[\Sigma_{\rho\cdot\varphi}(\tilde{\theta}) - \Sigma_{\rho\phi\cdot\varphi}(\tilde{\theta})\Sigma_{\phi\cdot\varphi}^{-1}(\tilde{\theta})\Sigma_{\rho\phi\cdot\varphi}'(\tilde{\theta})\right]^{-1} C_\rho^*(\tilde{\theta}), \tag{8.12}$$

where $C_\rho^*(\tilde{\theta}) = \left[C_\rho(\tilde{\theta}) - \Sigma_{\rho\phi\cdot\varphi}(\tilde{\theta})\Sigma_{\phi\cdot\varphi}^{-1}(\tilde{\theta})C_\phi(\tilde{\theta})\right]$ is the adjusted score function, and by roposition 5:

$$LM_\rho(DP) = n^\star C_\rho^{*'}(\tilde{\theta})\mathcal{D}_{\rho\cdot\varphi}^{-1}(\tilde{\theta})C_\rho^*(\tilde{\theta}), \tag{8.13}$$

where

$$\begin{aligned}\mathcal{D}_{\rho\cdot\varphi}(\tilde{\theta}) = \mathcal{B}_{\rho\cdot\varphi}(\tilde{\theta}) - \Sigma_{\rho\phi\cdot\varphi}(\tilde{\theta})\Sigma_{\phi\cdot\varphi}^{-1}(\tilde{\theta})\mathcal{B}_{\phi\rho\cdot\varphi}(\tilde{\theta}) - \mathcal{B}_{\rho\phi\cdot\varphi}(\tilde{\theta})\Sigma_{\phi\cdot\varphi}^{-1}(\tilde{\theta})\Sigma_{\phi\rho\cdot\varphi}(\tilde{\theta}) \\ +\Sigma_{\rho\phi\cdot\varphi}(\tilde{\theta})\Sigma_{\phi\cdot\varphi}^{-1}(\tilde{\theta})\mathcal{B}_{\phi\cdot\varphi}(\tilde{\theta})\Sigma_{\phi\cdot\varphi}^{-1}(\tilde{\theta})\Sigma_{\phi\rho\cdot\varphi}(\tilde{\theta})\end{aligned} \tag{8.14}$$

[4]The explicit expressions are given in Sections A and B for all terms stated in this section.

$$\mathcal{B}_{\rho\cdot\varphi}(\tilde{\theta}) = \Omega_{\rho}(\tilde{\theta}) - \Sigma_{\rho\varphi}(\tilde{\theta})\Sigma_{\varphi}^{-1}(\tilde{\theta})\Omega_{\varphi\rho}(\tilde{\theta}) - \Omega_{\rho\varphi}(\tilde{\theta})\Sigma_{\varphi}^{-1}(\tilde{\theta})\Sigma_{\varphi\rho}(\tilde{\theta}) + \Sigma_{\rho\varphi}(\tilde{\theta})\Sigma_{\varphi}^{-1}(\tilde{\theta})\Omega_{\varphi}(\tilde{\theta})\Sigma_{\varphi}^{-1}(\tilde{\theta})\Sigma_{\varphi\rho}(\tilde{\theta}) \quad (8.15)$$

$$\mathcal{B}_{\rho\phi\cdot\varphi}(\tilde{\theta}) = \Omega_{\rho\phi}(\tilde{\theta}) - \Sigma_{\rho\varphi}(\tilde{\theta})\Sigma_{\varphi}^{-1}(\tilde{\theta})\Omega_{\varphi\phi}(\tilde{\theta}) - \Omega_{\rho\varphi}(\tilde{\theta})\Sigma_{\varphi}^{-1}(\tilde{\theta})\Sigma_{\varphi\phi}(\tilde{\theta}) + \Sigma_{\rho\varphi}(\tilde{\theta})\Sigma_{\varphi}^{-1}(\tilde{\theta})\Omega_{\varphi}(\tilde{\theta})\Sigma_{\varphi}^{-1}(\tilde{\theta})\Sigma_{\varphi\phi}(\tilde{\theta}) \quad (8.16)$$

and $\mathcal{B}_{\phi\rho\cdot\varphi}(\tilde{\theta})$ is similarly defined. Both $LM_\rho(P)$ and $LM_\rho(DP)$ follow χ_1^2 distribution irrespective whether there are parametric and distributional misspecificaions or not. Both tests can be used for significance test of ρ, and they should be robust to both parametric and distributional misspecifications.

Similarly, the robust tests for $H_0^\gamma : \gamma_0 = 0$ and $H_0^\lambda : \lambda_0 = 0$are:

$$LM_\gamma(DP) = n^\star C_\gamma^{*'}(\tilde{\theta})\mathcal{D}_{\gamma\cdot\varphi}^{-1}(\tilde{\theta})C_\gamma^*(\tilde{\theta}), \quad (8.17)$$

$$LM_\gamma(P) = n^\star C_\gamma^{*'}(\tilde{\theta})\left[\Sigma_{\gamma\cdot\varphi}(\tilde{\theta}) - \Sigma_{\gamma\phi\cdot\varphi}(\tilde{\theta})\Sigma_{\phi\cdot\varphi}^{-1}(\tilde{\theta})\Sigma_{\gamma\phi\cdot\varphi}'(\tilde{\theta})\right]^{-1} C_\gamma^*(\tilde{\theta}), \quad (8.18)$$

where $\phi = (\lambda, \rho)'$,

$$LM_\lambda(DP) = n^\star C_\lambda^{*'}(\tilde{\theta})\mathcal{D}_{\lambda\cdot\varphi}^{-1}(\tilde{\theta})C_\lambda^*(\tilde{\theta}), \quad (8.19)$$

$$LM_\lambda(P) = n^\star C_\lambda^{*'}(\tilde{\theta})\left[\Sigma_{\lambda\cdot\varphi}(\tilde{\theta}) - \Sigma_{\lambda\phi\cdot\varphi}(\tilde{\theta})\Sigma_{\phi\cdot\varphi}^{-1}(\tilde{\theta})\Sigma_{\lambda\phi\cdot\varphi}'(\tilde{\theta})\right]^{-1} C_\lambda^*(\tilde{\theta}) \quad (8.20)$$

where $\phi = (\gamma, \rho)'$. Other terms are defined in the same way as in (8.13). For the above three tests, we have $LM_\psi(DP) \rightarrow LM_\psi(P)$.

There are some comments on the tests. First, in general, we do not need to care about the distributional misspecification, while parametric misspecification can cause the tests to be invalid. Second, with a sparse weight matrix and large n, any $LM_\psi(DP)$ type tests are close to $LM_\psi(P)$ tests. Though $LM_\psi(DP)$ provides a safe inference since it is constructed under distributional misspecification, $LM_\lambda(P)$ will be very close to $LM_\psi(DP)$. MLE result will be valid enough in most cases with non-normal errors. The Corollary 1 does not hold when W_n is dense in the sense that number of neighbors grows at most at a faster rate than $\sqrt{n}$, or when n is small. In such cases, one has to apply the $LM_\psi(D)$ and $LM_\psi(DP)$ type tests. Third, when one is confident that the nuisance parameters are correctly specified, comparing with $LM_\psi(P)$ type tests, LM_ψ will provide higher power. Fourth, when n is small comparing to T, tests based on transformation performs better than direct approach based tests.

9 A Monte Carlo Study and Empirical Illustration

In this section, I describe the details of a Monte Carlo study to illustrate my tests. All the proposed tests in this paper are packaged into the companion R package `sdpdlm` in order to allow practitioners to conduct an extensive specification search in SDPD models.

9.1 Design

The design is based on Bera et al. (2019), Taşpınar et al. (2017) and Fang et al. (2014).

$$Y_{nt} = \lambda_0 W_n Y_{nt} + \gamma_0 Y_{n,t-1} + \rho_0 W_n Y_{n,t-1} + X_{nt}\beta_0 + \mathbf{c}_{n0} + \alpha_{t0} l_n + V_{nt}, \tag{9.1}$$

for $t = 1, \ldots, T$, I specify two weights matrices corresponding to the Rook contiguity case and the Queen contiguity case. Consider n spatial units that are randomly permuted and allocated into a lattice of $\sqrt{n} \times \sqrt{n}$ squares. In the Rook contiguity case, $w_{ij,n} = 1$ if the spatial unit j is in a square that is adjacent (left/right/above or below) to the square of the spatial unit i, so a typical unit has 4 neighboring units. In the Queen contiguity, $w_{ij,n} = 1$ if the spatial unit j is in a square that is adjacent to, or shares a corner with the square of the spatial unit i, then a typical unit has 8 neighboring units. In both cases, W_n is row normalized. Considering the two weight matrices actually give almost the same simulation results, I present only the Queen weight matrix here for brevity. The simulation on Rook contiguity is available upon request.

Two exogenous regressors are in the model. The first one is generated as $X_{1,nt} = \Psi_n + 0.01\, t\, l_n + U_{nt}$, where $U_{nt} = 0.5\, U_{n,t-1} + \varepsilon_{nt} + 0.5\, \varepsilon_{n,t-1}$ and $\varepsilon_{nt} \sim N(0_{n\times 1}, 2I_n)$. Furthermore, $\Psi_n = \Upsilon_n + 1/(T + m + 1)\sum_{t=-m}^{T} \varepsilon_{nt}$, where $\Upsilon_n \sim N(0_{n\times 1}, I_n)$ and $m = 20$. Then, $X_{nt} = \big(X_{1,nt}, W_n X_{2,nt}\big)$ where $X_{2,nt} \sim N(0_{n\times 1}, I_n)$. $\beta_0 = (1.2, 0.6)$. For the individual effects, $\mathbf{c}_{n0} = (1/T)\sum_{t=1}^{T} X_{1,nt}$, and α_{t0} is drawn from $N(0, 1)$.

For the error term v_{it}, we can see the differences between Ω and Σ are basically constants multiply $\frac{\mu_4 - 3\sigma^4}{\sigma^4}$ where μ_4 is the fourth moment of disturbance term. Then, the kurtosis of the error's distribution should be influential to the testing, while skewness should not. Thus, I specify four cases. The first case is $v_{it} \sim IID\ N(0, 1)$, which can be seen a benchmark model. The second case is a modified Student's t distribution. $v_{it} \sim IID\sqrt{\frac{3}{5}}t_5$ whose skewness and kurtosis are 0 and 9, respectively. This illustrates the performance of the suggested tests with a standard heavy tail distribution. The third case is a modified chi-square distribution, $v_{it} \sim 0.5(\chi^2_2 - 2)$ with skewness=2 and kurtosis=9. This shows the performance of the tests with highly skewed and heavy tailed distribution. The fourth case is a mixture normal with skewness 0 and kurtosis 16.5, $v_{it} \sim [\frac{20}{21}N(0, 1) + \frac{1}{21}N(3, 20)]/\sqrt{\frac{40}{21}}$. This provides the performance of tests with extremely high kurtosis.

The data generating process has $21 + T$ periods and the first 20 periods are discarded as warm up periods. For the size properties, I will use multiple combinations of (n, T). The tests suggested by Bera et al. (2019) are $LM^d_\psi(P)$ tests in my notation i.e. parametric misspecification robust tests based on direct approach. Bera et al. (2019) studied the performance of the tests with

$(n, T) = (100, 10)$, and the performance of the $LM^d_\psi(P)$ tests are acceptable. In this paper, I have shown that the direct approach requires n/T to be finite but none 0. Also, Corollary 1 requires $n \to \infty$ to have $LM_\psi(DP) \to LM_\psi(P)$. Then $LM^d_\psi(P)$ type tests could perform poorly when n is small. I purposely check the performance of the tests with small n. The simulation is with (n, T) being $(9, 111)$, $(25, 40)$, $(49, 20)$ and $(100, 10)$[5]. Some of combinations have very large T that are typically not realistic in empirical studies. I specify this to maintain a roughly $n \times T = 1000$ sample size, so the result is comparable. For other part, I will mainly consider $(n, T) = (49, 10)$ which should be a commonly seen moderate sample size.

Under the null (i.e., $\lambda_0 = \gamma_0 = \rho_0 = 0$), model (9.1) reduces to a two-way error model (2WE). For the alternative model, there are seven different specifications I can consider. Given their commonality in empirical studies, I chose to focus on the following four specifications. The first specification is a dynamic panel data model with no spatial effects (DPD), i.e., when $\lambda_0 = \rho_0 = 0$ and $\gamma_0 \neq 0$. The second specification is a spatial static panel model (SSPD), i.e., when $\lambda_0 \neq 0$ and $\rho_0 = \gamma_0 = 0$. The third specification is a spatial dynamic panel data model with no spatial-time lag (SDPDW), i.e., when $\rho_0 = 0$, $\lambda_0 \neq 0$ and $\gamma_0 \neq 0$. The final specification for the alternative models is the spatial dynamic panel data model (SDPD), i.e., when $\rho_0 \neq 0$, $\lambda_0 \neq 0$ and $\gamma_0 \neq 0$. Note that the first three alternative models can also be considered as the null models for the one-directional tests and their robust counterparts in the following way: (i) the DPD model for LM_ρ, $LM_\rho(D)$,$LM_\rho(P)$, $LM_\rho(DP)$, LM_λ, $LM_\lambda(D)$, $LM_\lambda(P)$ and $LM_\lambda(DP)$; (ii) the SSPD model for LM_ρ, $LM_\rho(D)$, $LM_\rho(P)$, $LM_\rho(DP)$, LM_γ, $LM_\gamma(D)$, $LM_\gamma(P)$ and $LM_\gamma(DP)$; (iii) the SDPDW model for LM_ρ, $LM_\rho(D)$, $LM_\rho(P)$ and $LM_\rho(DP)$. I let γ_0, λ_0 and ρ_0 take values from $\{0.01, 0.02, 0.03, 0.04, 0.05, 0.10, 0.15, 0.20, 0.30\}$ in the alternative models. Resampling is carried out for $5,000$ times. Note that for $LM_\psi(D)$ type tests, they are too restrictive on the nuisance parameter comparing to $LM_\psi(P)$ and $LM_\psi(DP)$, so is of less value. I am going to focus on $LM_\psi(P)$ and $LM_\psi(DP)$ tests. The one-directional ideal condition LM_ψ tests will also be included in the paper as a benchmark to make comparison. The neglected $LM_\psi(D)$ tests actually perform very similar to LM_ψ. Moreover, Bera et al. (2019) proposed parametric misspecification robust tests based on direct approach. In this paper, I will mainly analyze the transformation approach based tests to avoid repetition. The result for simulations with direct approach are in fact similar to what are presented for transformation approach.

9.2 Size Properties

I first analyze the size discrepancy plots, Figure 1 to Figure 4, to discuss the size properties of the suggested tests. The size plots are plotted in the following way. Consider a sequence of nominal size $\{x_i\}$. I generate $\mathcal{R}$ samples from Monte Carlo simulation under the null hypothesis. Then conduct tests on the generated samples. Let τ_j for $j = 1, \ldots, \mathcal{R}$ be the $\mathcal{R}$ realizations of the test statistics τ, and $p(\tau)$ be the p-value corresponding to τ. The empirical rejection rate

[5]For the sake of brevity, I only provide estimation results for $(n, T) = (100, 10)$, and $(9, 111)$ as one most extreme case v.s. the same sample size in Bera et al. (2019).

$\hat{F}x_i = \sum_{j=1}^{\mathcal{R}} \mathbf{1}\big(p(\tau_j) \leq x_i\big)/\mathcal{R}$. I draw $\widehat{F}(x_i) - x_i$ against x_i as the size discrepancy plot.
Assume the sizes of the suggested tests are correct, I can then construct a point-wise 95% confidence interval for a nominal size using a normal approximation to the binomial distribution. A point-wise 95% confidence interval centered on nominal size x_i is given by $x_i \pm 1.96\,[x_i(1-x_i)/\mathcal{R}]^{1/2}$. In the discrepancy plots, the interval will be represented by the black solid lines, so $\widehat{F}(x_i) - x_i$ should be within interval $0 \pm 1.96\,[x_i(1-x_i)/\mathcal{R}]^{1/2}$ with 95% chance.
To further illustrate the size property, I provide Figure 5 and Figure 6, the empirical rejection rate of the suggested tests against value of misspecified parameters with all 4 distributions of disturbance terms.
To save space I present the discrepancy plots for testing on λ and ρ from the 2WE model, and the empirical rejection rate for testing on ρ from misspecified SSPD model. Other results are summarized in tables. When the null model is one of the DPD, SSPD and SDPDW models, I 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 are listed as follows.

1. Figure 1 and Figure 2 present the size discrepancy plots when the null model is 2WE for high number of individuals. In general, size distortions for all tests lie inside the 95% point-wise confidence intervals. With $n = 100$, we can clearly see that $LM_\psi(DP) = LM_\psi(P)$ when the distribution is non-normal.

2. Figure 3 and Figure 4 present the size discrepancy plots when the null model is 2WE for low number of individuals. As expected, the tests based on direct approach begins to fail, since n/T now is close to 0. Most of transformation approach based tests are still generally inside the 95% interval except LM^t_ρ. For the non-normal distributions, the pattern is similar.

3. Comparing Figure 3 and Figure 4 with Figure 1 and Figure 2, I observe that when n is small, direct approach tests become inaccurate. As my simulation shows, when n is above 25, the direct approach based tests become acceptable, so direct approach is valid in general. The simulations with $n = 49$ and $T = 10$ also confirm such findings.

4. Figure 5 and Figure 6 present the empirical rejection rate of the LM^t_ρ, $LM^t_\rho(P)$ and $LM^t_\rho(DP)$ tests against value of misspecified λ at a moderate sample size $(n, T) = (49, 10)$. First, $LM^t_\rho(P)$ and $LM^t_\rho(DP)$ are the same regardless of distribution of error terms. Moreover, the rejection rates for $LM^t_\rho(P)$ and $LM^t_\rho(DP)$ remain at 5%, until the nuisance parameter exceeds 0.3, while LM_ρ clearly over rejects the null.

5. Table 1 to Table 4 presents the magnitude of size distortions as a function of the local misspecification in the alternative model, the DPD model. I expect to see robust versions of the one directional tests to outperform their non-robust counterparts, when the magnitude of misspecification is small. For values of γ_0 between 0.01 and 0.05, I do not observe significant difference between the robust tests and their non-robust counterparts. As the misspecification in γ_0 deteriorates, the non-robust tests become over-sized, especially LM_ρ. On the other hand,

the robust tests seem to over-correct for the misspecification substantially, resulting in under rejection of the null hypothesis. This pattern is seen for all the distribution specifications.

6. In Table 5 to Table 8, the null model is the SSPD model, and the source of the misspecification is the deviation of λ_0 from zero. In general, the robust tests are properly sized while their non-robust counterparts are severely over-sized.

7. Table 9 and Table 10 confirms previous findings. In general, $LM^{\star}_{\rho}$ outperforms LM_{ρ}, when λ_0 and γ_0 deviate locally from zero. For example, when $\lambda_0 = 0.2$ and $\gamma_0 = 0.1$, the actual sizes of robust $LM_{\rho}(DP)$ and $LM_{\rho}(\rho)$ are around 0.045 at the 5% level for all the distributions. The actual sizes of LM_{ρ} are around 0.75. The simulation shows that γ is a more sensitive parameter comparing to λ in the sense that when $\gamma > 0.2$, $LM_{\psi}(P)$ and $LM_{\psi}(DP)$ begin to perform poorly while they are acceptable even after $\lambda > 0.3$.

8. In the proof, I need the alternative hypothesis for parameter to be local. The robust tests use the residuals from the estimation of 2WE model, and implements a correction on the test statistics for a local misspecification of the alternative model, i.e., ignoring the spatial and temporal component(s). The bias in these residuals will be high when spatial or temporal dependence is strong or to say when the alternative is far from the null. Therefore, I expect poor performance for the robust tests as spatial parameters deviate from zero substantially in the alternative model. In the simulation, we observe such effect. When $\gamma > 0.2$ or $\lambda > 0.3$, non-negligible size distorsion are observed even for the robust tests.

9. Except for the extreme $n = 9$ case, we have empirical rates for $LM_{\psi}(DP)$ and $LM_{\psi}(P)$ being equal for testing with all distributions, and all parameters. The rejection rates some time differ at 0.001% level and is not recorded by the tables. Note that my simulation is done with 5000 repetitions, and it means there are only one or two samples have different test results for $LM_{\psi}(DP)$ $LM_{\psi}(P)$ tests. Then, as a conclusion, in general we can treat $LM_{\psi}(P)$ as a test that is also robust to distribution specifications. In an empirical study, it is typically valid to use $LM_{\psi}(P)$ tests regardless of the assumption on the type of error distribution. There are three cases when the equivalence can be violated: first, when number of individual is too small. Second, when the W_n matrix is dense. For example, a social network model with most of the individuals being influential, so everyone is connected to everyone. Third, the data contains some extreme outliers, they will push up the kurtosis and make the distributional misspecification worse. When such properties are observed in the sample, one might considering using $LM_{\psi}(DP)$ instead of $LM_{\psi}(P)$.

9.3 power Properties

The results on the power properties of tests are presented in Tables 11 to Table 22. For brevity, some of the simulation results are not presented in the paper. The results based on the dropped settings are similar and available upon request. The general observations on the power properties of my proposed tests are listed as follows.

1. In Table 11 Table 14, the null model is the 2WE model, and the alternative model is the DPD model. The results indicate that test statistics for γ_0 and the joint tests statistic have satisfactory power. The robust test $LM_\gamma(P)$ and $LM_\gamma(DP)$ have slightly lower power than LM_γ confirming the theoretical results. The test statistics for $H_0 : \rho_0 = 0$ and $H_0 : \lambda_0 = 0$ should lack of power when the alternative model is the DPD model, and this is confirmed in the size property. I also see that LM_γ over rejects the null, in Table 5 to Table 8 when there is deviation of λ.

2. In Table 15 to Table 18, the null model is the 2WE model and the alternative model is the SSPD model. The results indicate that all three test statistics for λ_0 and the joint tests statistic have satisfactory power. The robust test $LM_\lambda(P)$ and $LM_\lambda(DP)$ again have slightly lower power than LM_λ confirming again the theoretical results. Comparing them with Table 5 to Table 8, the robust test statistics for $H_0 : \gamma_0 = 0$ and $H_0 : \rho_0 = 0$ lack of power when the alternative model is the SSPD model as expected. For LM_γ, the size is also not too much distorted. However, LM_ρ grossly over rejects the null, confirming the (over) size problem in Table 5 to Table 8.

3. In Table 19 to Table 22, the null model is the 2WE model and the alternative model is the SDPDW model. The results indicate that both test statistics for γ_0, and λ_0 and the joint tests statistic have satisfactory power. The robust tests $LM_\psi(P)$ and $LM_\psi(DP)$ have slightly lower power than their non-robust counterparts. The test statistics for $H_0 : \rho_0 = 0$ should lack of power when the alternative model is the SSPDW model, and this is confirmed in Table 9 and Table 10, although again LM_ρ over rejects the null confirming the (over) size problem. Therefore, the robust tests are preferable comparing to the non-robust tests.

4. I also do the simulation when the null model is the 2WE model and the alternative model is the SDPD model. In this case, tests should have power since under the alternative the values of γ_0, ρ_0 and λ_0 are all not zero. This is observed for all robust and non-robust tests. The simulation is as expected. For brevity, I do not present the table for this simulation. The the table is available upon request.

5. Through all the simulation, I do not observe significant difference between $LM_\psi(DP)$ and $LM_\psi(P)$. Moreover, power for the same tests are basically the same with any of the distribution. Such facts confirm the corollary that $LM_\psi(DP) \to LM_\psi(P)$, so $LM_\psi(P)$ is robust to non-normal distributions.

9.4 Empirical Illustration

In this section, I illustrate how these tests can be implemented and interpreted in an empirical study. I will use the same data set as Bera et al. (2019) on capital productivity from Munnell

(1990), so I can compare the tests in this paper with it. The data set includes observations related to public capital productivity on 48 contiguous US States from 1970 to 1986 annually. Munnell specifies a Cobb-Douglas production function that relates the gross domestic product of a given state to public capital input, private capital input, labor input and state unemployment rate. I consider an extension of her specification to a spatial panel data model by including spatial, temporal and spatio-temporal dependence in gross domestic product, and apply the SDPD model. As such, in my model specification (2.1), Y_{it} will denote gross domestic product of state i at period t and X_{it} will include public capital input, private capital input, labor input and state unemployment rate for state i at time t. The spatial weights matrix W_n is generated from the contiguity information of the states, i.e., $w_{ij} = 1$ if states i and j share a common border, otherwise $w_{ij} = 0$. The weights matrix is row normalized.
The aim of such tests is to test the presence of spatial, temporal and spatio-temporal dependence in the gross domestic product. Table 23 presents the test statistics and corresponding p-value from Bera et al. (2019). Test results from this paper are presented in Table 24 and Table 25.

Table 23: LM tests from Bera et al. (2019)

LM_γ^d	$LM_\gamma^d(P)$	LM_ρ^d	$LM_\rho^d(P)$	LM_λ^d	$LM_\lambda^d(P)$	LM_κ^d
498.779	497.882	31.741	50.787	77.841	76.062	576.657
(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Table 24: Some LM tests for direct approach in this paper

$LM_\gamma^d(D)$	$LM_\gamma^d(DP)$	$LM_\rho^d(D)$	$LM_\rho^d(DP)$	$LM_\lambda^d(D)$	$LM_\lambda^d(DP)$	$LM_\kappa^d(D)$
498.779	497.882	31.741	50.787	77.841	76.062	576.657
(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Table 25: Some LM tests for transformation approach in this paper

LM_γ^t	$LM_\gamma^t(DP)$	LM_ρ^t	$LM_\rho^t(DP)$	LM_λ^t	$LM_\lambda^t(DP)$	LM_κ^t
509.391	498.762	32.416	52.684	79.314	78.403	578.797
(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

First, all tests taken together indicate strong evidence of temporal, spatial, and spatio-temporal dependence. These findings lead us to the conclusion that there is strong statistical evidence for the SDPD specification and thus the data should be analyzed with it. Second, the $LM_\psi^d(DP)$ and $LM_\psi^d(D)$ tests suggested in this paper give test statistics almost the same as $LM_\psi^d(P)$ and LM_ψ^d

tests. (I would like to emphasize that Table 24 is truly the testing result of tests in this paper, not copied from Table 23.) Though not observed on the table, there is indeed some negligible difference, for example $LM_\lambda^d(P) = 76.06186$ and $LM_\lambda^d(DP) = 76.06182$. Such result shows that $LM_\psi(P)$ and $LM_\psi(DP)$ tests are equivalent, so $LM_\psi(P)$ tests are also robust to different distributions, as theory suggests. Third, the transformation approach gives a slightly different test statistics. In this application, all statistics are slightly higher for transformation approach. However, this is not what expected from the theory. The simulation results also do not show that transformation approach based tests tend to reject more often. It should be a coincidence in this sample.

Though the adjusted and unadjusted tests lead to the same conclusion here. For some other data sets, it is competely possible that $LM_\rho(DP)$ or $LM_\rho(P)$ take high values while LM_ρ is small enough to accept the null. The later inference is invalid since it is probably due to the presence of temporal (γ) and spatial (λ) dependence. As a summary for this study, the distributional misspecification is generally not a problem for testing, while parametric misspecification can cause different inference. Both the $LM_\psi(DP)$ and $LM_\psi(P)$ tests are valid for both misspecifications.

10 Conclusion

In this paper, I introduce robust LM tests within the QML framework for a spatial dynamic panel data model. These tests are robust in the sense that they specifically account for the asymptotic bias in the limiting distribution of score functions of the QMLE, as well as the potential parametric and distributional misspecifications in the alternative model. I also show that the $LM_\psi(P)$ tests that are only robust to parametric misspecification, are actually asymptotically equivalent to the $LM_\psi(DP)$ tests that are robust to both misspecifications, so both $LM_\psi(P)$ and $LM_\psi(DP)$ can be applied as robust to both misspecifications tests.

The robust tests have the central chi-square distribution when the alternative model is misspecified, whereas the asymptotic null distributions of the standard LM tests deviate from the central chi-square distribution. Therefore, the robust tests obtain asymptotically the correct size. I derive the asymptotic distribution of my proposed tests under the null and the local alternative hypotheses. These tests can be used to test the presence of the contemporaneous dependence over space, dependence over time and spatial time dependence. Since these tests are robust to both misspecification of the alternative models, they are more suitable for the detection of the source of dependence in a spatial dynamic panel data model.

One attractive feature of my proposed tests is that the test statistics are easy to compute, and only require the estimates from a two-way error model. Therefore, these tests can be easily made available for practical applications using the standard statistical softwares. Calculation for the more complicated $LM_\psi(DP)$ test can also be saved by the equivalence of $LM_\psi(P)$ and $LM_\psi(DP)$. In a Monte Carlo study, I investigated the size and power properties of my proposed tests. The simulation results confirm that the robust tests have acceptable finite sample properties, and can be useful for the detection of the source of dependence. The results, hence, coincides with my analytical findings that the robust tests are valid, when the alternative models locally deviate from the true data generating process, and when the error is not normally distributed. Also, in empirical illustrations, I show that how these robust tests can be used. The simulation and empirical application also together confirm the equivalance between $LM_{\psi}(DP)$ and $LM_{\psi}(P)$. Importantly, all tests can be effortlessly implemented in a specification search by using the R software package `sdpdlm`.

In future studies, the testing approach can be extended to unstable SDPD models, where there are unit roots generated by temporal and spatial correlations in the ML or QML framework. The tests can also be formulated for the SDPD models that have time-varying or endogenous spatial weight matrices. Finally, possible heteroskedasticity can be introduced into the SDPD model. I leave these topics for future research.

Appendix

A The Score Functions, Hessian and Outer Product Matrices

In this section, I provide the parts of score functions, $\Sigma(\tilde{\theta})$ and $\Omega(\tilde{\theta})$ matrices that are required for the test statistics.

A.1 Direct Approach

Let $n^d = nT$ be the effective sample size for direct approach. The first-order conditions evaluated at $\tilde{\theta} = \left(0, 0, \tilde{\beta}', 0, \tilde{\sigma}^2\right)'$ are

$$\frac{1}{n^d}\frac{\partial \ln L^d_{nT}(\tilde{\theta})}{\partial \lambda} = \frac{1}{\tilde{\sigma}^2 n^d}\sum_{t=1}^{T}\tilde{Y}'_{nt}W'_n J_n \tilde{V}_{nt}(\tilde{\theta}), \tag{A.1}$$

$$\frac{1}{n^d}\frac{\partial \ln L^d_{nT}(\tilde{\theta})}{\partial \gamma} = \frac{1}{\tilde{\sigma}^2 n^d}\sum_{t=1}^{T}\tilde{Y}'_{n,t-1} J_n \tilde{V}_{nt}(\tilde{\theta}), \tag{A.2}$$

$$\frac{1}{n^d}\frac{\partial \ln L^d_{nT}(\tilde{\theta})}{\partial \rho} = \frac{1}{\tilde{\sigma}^2 n^d}\sum_{t=1}^{T}\left(W_n\tilde{Y}_{n,t-1}\right)' J_n \tilde{V}_{nt}(\tilde{\theta}). \tag{A.3}$$

where $\tilde{V}_{nt}(\tilde{\theta}) = \tilde{Y}_{nt} - \tilde{X}_{nt}\tilde{\beta}$. The elements of Hessian matrix $\Sigma(\tilde{\theta})$ evaluated at $\tilde{\theta}$ are listed in the following.

$$\begin{aligned}
&\Sigma_{\gamma}(\tilde{\theta}) = \frac{1}{\tilde{\sigma}^2 n^d}\sum_{t=1}^{T}\tilde{Y}'_{n,t-1}J_n\tilde{Y}_{n,t-1}, \quad \Sigma_{\gamma\rho}(\tilde{\theta}) = \frac{1}{\tilde{\sigma}^2 n^d}\sum_{t=1}^{T}\tilde{Y}'_{n,t-1}J_n W_n\tilde{Y}_{n,t-1},\\
&\Sigma_{\gamma\beta}(\tilde{\theta}) = \frac{1}{\tilde{\sigma}^2 n^d}\sum_{t=1}^{T}\tilde{Y}'_{n,t-1}J_n\tilde{X}_{nt}, \quad \Sigma_{\gamma\lambda}(\tilde{\theta}) = \frac{1}{\tilde{\sigma}^2 n^d}\sum_{t=1}^{T}\tilde{Y}'_{n,t-1}J_n W_n\tilde{X}_{nt}\tilde{\beta},\\
&\Sigma_{\gamma\sigma}(\tilde{\theta}) = 0_{1\times 1}.
\end{aligned} \tag{A.4}$$

$$\begin{aligned}
&\Sigma_{\rho}(\tilde{\theta}) = \frac{1}{\tilde{\sigma}^2 n^d}\sum_{t=1}^{T}\left(W_n\tilde{Y}_{n,t-1}\right)' J_n W_n\tilde{Y}_{n,t-1}, \quad \Sigma_{\rho\beta}(\tilde{\theta}) = \frac{1}{\tilde{\sigma}^2 n^d}\sum_{t=1}^{T}\left(W_n\tilde{Y}_{n,t-1}\right)' J_n\tilde{X}_{nt},\\
&\Sigma_{\rho\lambda}(\tilde{\theta}) = \frac{1}{\tilde{\sigma}^2 n^d}\sum_{t=1}^{T}\left(W_n\tilde{Y}_{n,t-1}\right)' J_n W_n\tilde{X}_{nt}\tilde{\beta}, \quad \Sigma_{\rho\sigma}(\tilde{\theta}) = 0_{1\times 1}.
\end{aligned} \tag{A.5}$$

$$\begin{aligned}
&\Sigma_{\beta}(\tilde{\theta}) = \frac{1}{\tilde{\sigma}^2 n^d}\sum_{t=1}^{T}\tilde{X}'_{nt}J_n\tilde{X}_{nt}, \quad \Sigma_{\beta\lambda}(\tilde{\theta}) = \frac{1}{\tilde{\sigma}^2 n^d}\sum_{t=1}^{T}\tilde{X}'_{nt}J_n W_n\tilde{X}_{nt}\tilde{\beta},\\
&\Sigma_{\beta\sigma}(\tilde{\theta}) = 0_{k_x\times 1}.
\end{aligned} \tag{A.6}$$

$$\Sigma_{\lambda}(\tilde{\theta}) = \frac{1}{\tilde{\sigma}^2 n^d} \sum_{t=1}^{T} \left(W_n \tilde{X}_{nt} \tilde{\beta}\right)' J_n W_n \tilde{X}_{nt} \tilde{\beta} + \frac{1}{n} \left[\mathrm{tr}\left(W_n' J_n W_n\right) + \mathrm{tr}\left(W_n^2\right)\right],$$
$$\Sigma_{\lambda\sigma}(\tilde{\theta}) = \frac{-1}{\tilde{\sigma}^2 n}. \tag{A.7}$$

$$\Sigma_{\sigma}(\tilde{\theta}) = \frac{1}{2\tilde{\sigma}^4}. \tag{A.8}$$

The elements of outer product matrix $\Omega(\tilde{\theta})$ evaluated at $\tilde{\theta}$ are listed in the following.

$$\Omega_{\gamma}(\tilde{\theta}) = \frac{1}{\tilde{\sigma}^2 n^d} \sum_{t=1}^{T} \tilde{Y}_{n,t-1}' J_n \tilde{Y}_{n,t-1}, \quad \Omega_{\gamma\rho}(\tilde{\theta}) = \frac{1}{\tilde{\sigma}^2 n^d} \sum_{t=1}^{T} \tilde{Y}_{n,t-1}' J_n W_n \tilde{Y}_{n,t-1},$$
$$\Omega_{\gamma\beta}(\tilde{\theta}) = \frac{1}{\tilde{\sigma}^2 n^d} \sum_{t=1}^{T} \tilde{Y}_{n,t-1}' J_n \tilde{X}_{nt}, \quad \Omega_{\gamma\lambda}(\tilde{\theta}) = \frac{1}{\tilde{\sigma}^2 n^d} \sum_{t=1}^{T} \tilde{Y}_{n,t-1}' J_n W_n \tilde{X}_{nt} \tilde{\beta},$$
$$\Omega_{\gamma\sigma}(\tilde{\theta}) = 0_{1\times 1}. \tag{A.9}$$

$$\Omega_{\rho}(\tilde{\theta}) = \frac{1}{\tilde{\sigma}^2 n^d} \sum_{t=1}^{T} \left(W_n \tilde{Y}_{n,t-1}\right)' J_n W_n \tilde{Y}_{n,t-1}, \quad \Omega_{\rho\beta}(\tilde{\theta}) = \frac{1}{\tilde{\sigma}^2 n^d} \sum_{t=1}^{T} \left(W_n \tilde{Y}_{n,t-1}\right)' J_n \tilde{X}_{nt},$$
$$\Omega_{\rho\lambda}(\tilde{\theta}) = \frac{1}{\tilde{\sigma}^2 n^d} \sum_{t=1}^{T} \left(W_n \tilde{Y}_{n,t-1}\right)' J_n W_n \tilde{X}_{nt} \tilde{\beta}, \quad \Omega_{\rho\sigma}(\tilde{\theta}) = 0_{1\times 1}. \tag{A.10}$$

$$\Omega_{\beta}(\tilde{\theta}) = \frac{1}{\tilde{\sigma}^2 n^d} \sum_{t=1}^{T} \tilde{X}_{nt}' J_n \tilde{X}_{nt}, \quad \Omega_{\beta\lambda}(\tilde{\theta}) = \frac{1}{\tilde{\sigma}^2 n^d} \sum_{t=1}^{T} \tilde{X}_{nt}' J_n W_n \tilde{X}_{nt} \tilde{\beta},$$
$$\Omega_{\beta\sigma}(\tilde{\theta}) = 0_{k_x \times 1}. \tag{A.11}$$

$$\Omega_{\lambda}(\tilde{\theta}) = \frac{1}{\tilde{\sigma}^2 n^d} \sum_{t=1}^{T} \left(W_n \tilde{X}_{nt} \tilde{\beta}\right)' J_n W_n \tilde{X}_{nt} \tilde{\beta} + \frac{1}{n} \left[\mathrm{tr}\left(W_n' J_n W_n\right) + \mathrm{tr}\left(W_n^2\right)\right],$$
$$\Omega_{\lambda\sigma}(\tilde{\theta}) = \frac{-1}{\tilde{\sigma}^2 n}. \tag{A.12}$$

$$\Omega_{\sigma}(\tilde{\theta}) = \frac{1}{2\tilde{\sigma}^4} + \frac{\tilde{\mu}_4 - 3\tilde{\sigma}^4}{\tilde{\sigma}^4} \frac{1}{4\tilde{\sigma}^2}. \tag{A.13}$$

where $W_{n,ii}^2$ is the (i,i) entry of W_n^2 and μ_4 denotes the fourth moment of v_{it}. Applying an approximation from Rahmatullah Imon (2003) can provide better results. I estimate $\tilde{\mu_4}$ by $(\frac{nT}{(n-1)(T-1)-k})^4\left[\sum_{i,t}\tilde{v_{it}}^4/nT-3(1-(\frac{nT}{(n-1)(T-1)-k})^{-2})(\sum_{i,t}\tilde{v_{it}}^2/nT)^2\right]$. $\tilde{v_{it}}$ denotes the residuals from the restricted model, then $\sum_{i,t}\tilde{v_{it}}^4/nT$ and $\sum_{i,t}\tilde{v_{it}}^2/nT$ can be viewed as the naive estimation of the fourth moment and second moment of the disturbance term.

A.2 Transformation Approach

Let $n^t=(n-1)T$ be the effective sample size. The first-order conditions evaluated at $\tilde{\theta}$ are

$$\frac{1}{n^t}\frac{\partial\ln L_{nT}(\tilde{\theta})}{\partial\lambda}=\frac{1}{\tilde{\sigma}^2n^t}\sum_{t=1}^{T}\tilde{Y}_{nt}'W_n'J_n\tilde{V}_{nt}(\tilde{\theta})+\frac{T}{n^t}, \tag{A.14}$$

$$\frac{1}{n^t}\frac{\partial\ln L_{nT}(\tilde{\theta})}{\partial\gamma}=\frac{1}{\tilde{\sigma}^2n^t}\sum_{t=1}^{T}\tilde{Y}_{n,t-1}'J_n\tilde{V}_{nt}(\tilde{\theta}), \tag{A.15}$$

$$\frac{1}{n^t}\frac{\partial\ln L_{nT}(\tilde{\theta})}{\partial\rho}=\frac{1}{\tilde{\sigma}^2n^t}\sum_{t=1}^{T}\left(W_n\tilde{Y}_{n,t-1}\right)'J_n\tilde{V}_{nt}(\tilde{\theta}). \tag{A.16}$$

where $\tilde{V}_{nt}(\tilde{\theta})=\tilde{Y}_{nt}-\tilde{X}_{nt}\tilde{\beta}$. The elements of Hessian matrix $\Sigma(\tilde{\theta})$ evaluated at $\tilde{\theta}$ are listed in the following.

$$\begin{aligned}&\Sigma_\gamma(\tilde{\theta})=\frac{1}{\tilde{\sigma}^2n^t}\sum_{t=1}^{T}\tilde{Y}_{n,t-1}'J_n\tilde{Y}_{n,t-1},\quad \Sigma_{\gamma\rho}(\tilde{\theta})=\frac{1}{\tilde{\sigma}^2n^t}\sum_{t=1}^{T}\tilde{Y}_{n,t-1}'J_nW_n\tilde{Y}_{n,t-1},\\ &\Sigma_{\gamma\beta}(\tilde{\theta})=\frac{1}{\tilde{\sigma}^2n^t}\sum_{t=1}^{T}\tilde{Y}_{n,t-1}'J_n\tilde{X}_{nt},\quad \Sigma_{\gamma\lambda}(\tilde{\theta})=\frac{1}{\tilde{\sigma}^2n^t}\sum_{t=1}^{T}\tilde{Y}_{n,t-1}'J_nW_n\tilde{X}_{nt}\tilde{\beta},\\ &\Sigma_{\gamma\sigma}(\tilde{\theta})=0_{1\times1}.\end{aligned} \tag{A.17}$$

$$\begin{aligned}&\Sigma_\rho(\tilde{\theta})=\frac{1}{\tilde{\sigma}^2n^t}\sum_{t=1}^{T}\left(W_n\tilde{Y}_{n,t-1}\right)'J_nW_n\tilde{Y}_{n,t-1},\quad \Sigma_{\rho\beta}(\tilde{\theta})=\frac{1}{\tilde{\sigma}^2n^t}\sum_{t=1}^{T}\left(W_n\tilde{Y}_{n,t-1}\right)'J_n\tilde{X}_{nt},\\ &\Sigma_{\rho\lambda}(\tilde{\theta})=\frac{1}{\tilde{\sigma}^2n^t}\sum_{t=1}^{T}\left(W_n\tilde{Y}_{n,t-1}\right)'J_nW_n\tilde{X}_{nt}\tilde{\beta},\quad \Sigma_{\rho\sigma}(\tilde{\theta})=0_{1\times1}.\end{aligned} \tag{A.18}$$

$$\begin{aligned}&\Sigma_\beta(\tilde{\theta})=\frac{1}{\tilde{\sigma}^2n^t}\sum_{t=1}^{T}\tilde{X}_{nt}'J_n\tilde{X}_{nt},\quad \Sigma_{\beta\lambda}(\tilde{\theta})=\frac{1}{\tilde{\sigma}^2n^t}\sum_{t=1}^{T}\tilde{X}_{nt}'J_nW_n\tilde{X}_{nt}\tilde{\beta},\\ &\Sigma_{\beta\sigma}(\tilde{\theta})=0_{k_x\times1}.\end{aligned} \tag{A.19}$$

$$\Sigma_{\lambda}(\tilde{\theta}) = \frac{1}{\tilde{\sigma}^2 n^t}\sum_{t=1}^{T}\left(W_n\tilde{X}_{nt}\tilde{\beta}\right)' J_nW_n\tilde{X}_{nt}\tilde{\beta} + \frac{1}{n-1}\left[\operatorname{tr}\left(W_n'J_nW_n\right) + \operatorname{tr}\left(W_n^2\right) - 1\right],$$

$$\Sigma_{\lambda\sigma}(\tilde{\theta}) = \frac{-1}{\tilde{\sigma}^2(n-1)}. \tag{A.20}$$

$$\Sigma_{\sigma}(\tilde{\theta}) = \frac{n+1}{2\tilde{\sigma}^4(n-1)}. \tag{A.21}$$

The elements of outer product matrix $\Omega(\tilde{\theta})$ evaluated at $\tilde{\theta}$ are listed in the following.

$$\Omega_{\gamma}(\tilde{\theta}) = \frac{1}{\tilde{\sigma}^2 n^t}\sum_{t=1}^{T}\tilde{Y}_{n,t-1}'J_n\tilde{Y}_{n,t-1}, \quad \Omega_{\gamma\rho}(\tilde{\theta}) = \frac{1}{\tilde{\sigma}^2 n^t}\sum_{t=1}^{T}\tilde{Y}_{n,t-1}'J_nW_n\tilde{Y}_{n,t-1},$$

$$\Omega_{\gamma\beta}(\tilde{\theta}) = \frac{1}{\tilde{\sigma}^2 n^t}\sum_{t=1}^{T}\tilde{Y}_{n,t-1}'J_n\tilde{X}_{nt}, \quad \Omega_{\gamma\lambda}(\tilde{\theta}) = \frac{1}{\tilde{\sigma}^2 n^t}\sum_{t=1}^{T}\tilde{Y}_{n,t-1}'J_nW_n\tilde{X}_{nt}\tilde{\beta},$$

$$\Omega_{\gamma\sigma}(\tilde{\theta}) = 0_{1\times 1}. \tag{A.22}$$

$$\Omega_{\rho}(\tilde{\theta}) = \frac{1}{\tilde{\sigma}^2 n^t}\sum_{t=1}^{T}\left(W_n\tilde{Y}_{n,t-1}\right)' J_nW_n\tilde{Y}_{n,t-1}, \quad \Omega_{\rho\beta}(\tilde{\theta}) = \frac{1}{\tilde{\sigma}^2 n^t}\sum_{t=1}^{T}\left(W_n\tilde{Y}_{n,t-1}\right)' J_n\tilde{X}_{nt},$$

$$\Omega_{\rho\lambda}(\tilde{\theta}) = \frac{1}{\tilde{\sigma}^2 n^t}\sum_{t=1}^{T}\left(W_n\tilde{Y}_{n,t-1}\right)' J_nW_n\tilde{X}_{nt}\tilde{\beta}, \quad \Omega_{\rho\sigma}(\tilde{\theta}) = 0_{1\times 1}. \tag{A.23}$$

$$\Omega_{\beta}(\tilde{\theta}) = \frac{1}{\tilde{\sigma}^2 n^t}\sum_{t=1}^{T}\tilde{X}_{nt}'J_n\tilde{X}_{nt}, \quad \Omega_{\beta\lambda}(\tilde{\theta}) = \frac{1}{\tilde{\sigma}^2 n^t}\sum_{t=1}^{T}\tilde{X}_{nt}'J_nW_n\tilde{X}_{nt}\tilde{\beta},$$

$$\Omega_{\beta\sigma}(\tilde{\theta}) = 0_{k_x\times 1}. \tag{A.24}$$

$$\Omega_{\lambda}(\tilde{\theta}) = \frac{1}{\tilde{\sigma}^2 n^t}\sum_{t=1}^{T}\left(W_n\tilde{X}_{nt}\tilde{\beta}\right)' J_nW_n\tilde{X}_{nt}\tilde{\beta} + \frac{1}{n-1}\left[\operatorname{tr}\left(W_n'J_nW_n\right) + \operatorname{tr}\left(W_n^2\right) - 1\right]$$

$$+ \frac{\tilde{\mu}_4 - 3\tilde{\sigma}^4}{\tilde{\sigma}^4(n-1)}\sum_{i=1}^{n}[(J_nW_n)_{ii}]^2,$$

$$\Omega_{\lambda\sigma}(\tilde{\theta}) = \frac{-1}{\tilde{\sigma}^2(n-1)} - \frac{\tilde{\mu}_4 - 3\tilde{\sigma}^4}{2\tilde{\sigma}^6(n-1)}. \tag{A.25}$$

$$\Omega_\sigma(\tilde{\theta}) = \frac{n+1}{2\tilde{\sigma}^4(n-1)} + \frac{\tilde{\mu}_4 - 3\tilde{\sigma}^4}{\tilde{\sigma}^4}\frac{1}{4\tilde{\sigma}^2}. \tag{A.26}$$

B Expressions for Adjusted Score Functions

In this section, I first provide the bias correction terms under the joint null:

$$\Delta_{nT,1}(\tilde{\theta}) = \frac{1}{\sqrt{(n-1)T}} \begin{bmatrix} n-1 \\ -1 \\ 0_{k_x \times 1} \\ -1 \\ \frac{n-1}{2\tilde{\sigma}^2} \end{bmatrix}, \tag{B.1}$$

$$\Delta_{nT,2}(\tilde{\theta}) = \frac{1}{\sqrt{nT}} \begin{bmatrix} n-1 \\ -1 \\ 0_{k_x \times 1} \\ -1 \\ \frac{n-1}{2\tilde{\sigma}^2} \end{bmatrix}, \tag{B.2}$$

$$\Delta_{nT,3}(\tilde{\theta}) = \sqrt{\frac{T}{n}} \begin{bmatrix} 0_{(k_x+2)\times 1} \\ 1 \\ \frac{1}{2\tilde{\sigma}^2} \end{bmatrix}. \tag{B.3}$$

$C_\psi(\tilde{\theta}) = D_\psi(\tilde{\theta}) - \frac{1}{\sqrt{nT}}\Sigma_{\psi\varphi}(\tilde{\theta})\Sigma_\varphi^{-1}(\tilde{\theta})\left(\Delta_{nT,2\varphi}(\tilde{\theta}) + \Delta_{nT,3\varphi}(\tilde{\theta})\right) + \frac{1}{\sqrt{nT}}\left(\Delta_{nT,2\psi}(\tilde{\theta}) + \Delta_{nT,3\psi}(\tilde{\theta})\right)$ for the direct approach, and $C_\psi(\tilde{\theta}) = D_\psi(\tilde{\theta}) - \frac{1}{\sqrt{(n-1)T}}\Sigma_{\psi\varphi}(\tilde{\theta})\Sigma_\varphi^{-1}(\tilde{\theta})\left(\Delta_{nT,1\varphi}(\tilde{\theta})\right) + \frac{1}{\sqrt{(n-1)T}}\left(\Delta_{nT,1\psi}(\tilde{\theta})\right)$ for the transformation approach to calculate the corresponding adjusted score functions.

B.1 Direct Approach

$$C_\gamma(\tilde{\theta}) = \frac{1}{\tilde{\sigma}^2 n^d}\sum_{t=1}^{T} \tilde{Y}'_{n,t-1} J_n \tilde{V}_{nt}(\tilde{\theta}) + \frac{n-1}{n^d},$$

$$C_\rho(\tilde{\theta}) = \frac{1}{\tilde{\sigma}^2 n^d}\sum_{t=1}^{T} \left(W_n \tilde{Y}_{n,t-1}\right)' J_n \tilde{V}_{nt}(\tilde{\theta}) - \frac{1}{n^d},$$

$$C_\lambda(\tilde{\theta}) = \frac{1}{\tilde{\sigma}^2 n^d}\sum_{t=1}^{T}\tilde{Y}_{nt}' W_n' J_n \tilde{V}_{nt}(\tilde{\theta}) - \frac{1}{\sqrt{n^d}}\left(\frac{1}{\tilde{\sigma}^2 n^d}\sum_{t=1}^{T}\left(W_n\tilde{X}_{nt}\tilde{\beta}\right)' J_n\tilde{X}_{nt},\, \frac{-1}{\tilde{\sigma}^2 n}\right)$$

$$\times\begin{bmatrix}\frac{1}{\tilde{\sigma}^2 n^d}\sum_{t=1}^{T}\tilde{X}_{nt}' J_n\tilde{X}_{nt} & 0_{k_x\times 1}\\ 0_{1\times k_x} & \frac{1}{2\tilde{\sigma}^4}\end{bmatrix}^{-1}\begin{bmatrix}0_{k_x\times 1}\\ \frac{1}{\sqrt{n^d}}\frac{n-1}{2\tilde{\sigma}^2}+\sqrt{\frac{T}{n}}\frac{1}{2\tilde{\sigma}^2}\end{bmatrix} + \frac{-1}{n^d} + \frac{1}{n}.$$

B.2 Transformation Approach

$$C_\gamma(\tilde{\theta}) = \frac{1}{\tilde{\sigma}^2 n^t}\sum_{t=1}^{T}\tilde{Y}_{n,t-1}' J_n\tilde{V}_{nt}(\tilde{\theta}) + \frac{1}{T},$$

$$C_\rho(\tilde{\theta}) = \frac{1}{\tilde{\sigma}^2 n^t}\sum_{t=1}^{T}\left(W_n\tilde{Y}_{n,t-1}\right)' J_n\tilde{V}_{nt}(\tilde{\theta}) + \frac{-1}{n^t},$$

$$C_\lambda(\tilde{\theta}) = \frac{1}{\tilde{\sigma}^2 n^t}\sum_{t=1}^{T}\tilde{Y}_{nt}' W_n' J_n\tilde{V}_{nt}(\tilde{\theta}) + \frac{T}{n^t} - \frac{1}{\sqrt{n^t}}\left(\frac{1}{\tilde{\sigma}^2 n^t}\sum_{t=1}^{T}\left(W_n\tilde{X}_{nt}\tilde{\beta}\right)' J_n\tilde{X}_{nt},\, \frac{-1}{\tilde{\sigma}^2(n-1)}\right)$$

$$\times\begin{bmatrix}\frac{1}{\tilde{\sigma}^2 n^t}\sum_{t=1}^{T}\tilde{X}_{nt}' J_n\tilde{X}_{nt} & 0_{k_x\times 1}\\ 0_{1\times k_x} & \frac{1}{2\tilde{\sigma}^4}\end{bmatrix}^{-1}\begin{bmatrix}0_{k_x\times 1}\\ \frac{1}{\sqrt{n^t}}\frac{n-1}{2\tilde{\sigma}^2}\end{bmatrix} + \frac{-1}{n^t}.$$

C Proofs of Propositions

C.1 Proof of Proposition 1

The proof of Proposition 1 is provided in Lee and Yu (2010).

C.2 Proof of Proposition 2

The result is directly from (4.8) and (4.9).

C.3 Proof of Proposition 3

The result is directly from (5.8) and (5.9).

C.4 Proof of Proposition 4:

The first result of the proposition directly follows from $\sqrt{n^\star}C_\psi(\tilde{\theta}) \xrightarrow{d} N\left(\Sigma_{\psi\cdot\varphi}\zeta + \Sigma_{\psi\phi\cdot\varphi}\delta,\, \Sigma_{\psi\cdot\varphi}\right)$. Here, I will only prove the last two results. To this end, I need to determine the joint distribution of $\mathbf{D}(\tilde{\theta}) = \left(D_\psi'(\tilde{\theta}),\, D_\phi'(\tilde{\theta})\right)'$ under H_0^ψ and H_A^ϕ. Let $\kappa = \left(\psi', \phi'\right)'$. The first-order Taylor expansions of scores $\mathbf{D}(\tilde{\theta})$ and $D_{\varphi}(\tilde{\theta})$ around $\theta_0 = \left(\varphi_0^{'},\, \psi_0^{'}\, \phi_0^{'}\right)^{'}$ under H_A^{ψ} and H_A^{ϕ} can be stated as

$$\sqrt{n^{\star}}\mathbf{D}(\tilde{\theta}) = \sqrt{n^{\star}}\mathbf{D}(\theta_0) - \mathbf{D}_{\kappa}(\theta_0)\left(\zeta^{'},\, \delta^{'}\right)^{'} + \sqrt{n^{\star}}\mathbf{D}_{\varphi}(\theta_0)(\tilde{\varphi} - \varphi_0) + o_p(1), \tag{C.1}$$

$$\sqrt{n^{\star}}D_{\varphi}(\tilde{\theta}) = \sqrt{n^{\star}}D_{\varphi}(\theta_0) - \mathbf{D}_{\varphi\kappa}(\theta_0)\left(\zeta^{'},\, \delta^{'}\right)^{'} + \sqrt{n^{\star}}D_{\varphi\varphi}(\theta_0)(\tilde{\varphi} - \varphi_0) + o_p(1), \tag{C.2}$$

where $\mathbf{D}_{\kappa} = \begin{bmatrix} D_{\psi\psi} & D_{\psi\phi} \\ D_{\phi\psi} & D_{\phi\phi} \end{bmatrix}$, $\mathbf{D}_{\varphi} = \begin{bmatrix} D_{\psi\varphi} \\ D_{\phi\varphi} \end{bmatrix}$ and $\mathbf{D}_{\varphi\kappa} = \begin{bmatrix} D_{\psi\varphi} & D_{\phi\varphi} \end{bmatrix}$. Under Proposition 1, the results in (C.1) and (C.2) imply that

$$\sqrt{n^{\star}}\mathbf{D}(\tilde{\theta}) = \left[-\mathbf{\Sigma}_{\psi\phi,\varphi}\Sigma_{\varphi}^{-1},\, I_{r+s}\right]\begin{bmatrix} \sqrt{n^{\star}}D_{\varphi}(\theta_0) \\ \sqrt{n^{\star}}\mathbf{D}(\theta_0) \end{bmatrix} + \begin{bmatrix} \Sigma_{\psi\cdot\varphi} & \Sigma_{\psi\phi\cdot\varphi} \\ \Sigma_{\psi\phi\cdot\varphi}^{'} & \Sigma_{\phi\cdot\varphi} \end{bmatrix}\begin{bmatrix} \zeta \\ \delta \end{bmatrix} + o_p(1), \tag{C.3}$$

where $\mathbf{\Sigma}_{\psi\phi,\varphi} = \left(\Sigma_{\psi\varphi}^{'},\, \Sigma_{\phi\varphi}^{'}\right)^{'}$. By Proposition 1, we have

$$\begin{bmatrix} \sqrt{n^{\star}}D_{\varphi}(\theta_0) \\ \sqrt{n^{\star}}\mathbf{D}(\theta_0) \end{bmatrix} + \begin{bmatrix} \Delta_{nT,\varphi} \\ \Delta_{nT,\psi} \\ \Delta_{nT,\phi} \end{bmatrix} \xrightarrow{d} N\left(0,\, \Sigma\right). \tag{C.4}$$

Therefore, under H_0^{ψ} and H_A^{ϕ}, the results in (C.3) and (C.4) imply that

$$\sqrt{n^{\star}}\mathbf{D}(\tilde{\theta}) - \begin{bmatrix} \Sigma_{\psi\varphi}\Sigma_{\varphi}^{-1}\Delta_{nT,\varphi} \\ \Sigma_{\phi\varphi}\Sigma_{\varphi}^{-1}\Delta_{nT,\varphi} \end{bmatrix} + \begin{bmatrix} \Delta_{nT,\psi} \\ \Delta_{nT,\phi} \end{bmatrix} \xrightarrow{d} N\left(\begin{bmatrix} \Sigma_{\psi\phi\cdot\varphi}\delta \\ \Sigma_{\phi\cdot\varphi}\delta \end{bmatrix}, \begin{bmatrix} \Sigma_{\psi\cdot\varphi} & \Sigma_{\psi\phi\cdot\varphi} \\ \Sigma_{\psi\phi\cdot\varphi}^{'} & \Sigma_{\phi\cdot\varphi} \end{bmatrix}\right). \tag{C.5}$$

The result in (C.5) can be used to determine the distribution of the adjusted score. Let $\mathbf{C}(\tilde{\theta}) = \left(C_{\psi}(\tilde{\theta}), C_{\phi}(\tilde{\theta})\right)^{'}$ the adjusted score function can be written as

$$\sqrt{n^{\star}}C_{\psi}^{*}(\tilde{\theta}) = \left[I_r,\, -\Sigma_{\psi\phi\cdot\varphi}\Sigma_{\phi\cdot\varphi}^{-1}\right]\sqrt{n^{\star}}\mathbf{C}(\tilde{\theta}) + o_p(1), \tag{C.6}$$

where

$$\begin{aligned} \sqrt{n^{\star}}\mathbf{C}(\tilde{\theta}) = \begin{bmatrix} \sqrt{n^{\star}}C_{\psi}(\tilde{\theta}) \\ \sqrt{n^{\star}}C_{\phi}(\tilde{\theta}) \end{bmatrix} &= \begin{bmatrix} D_{\psi}(\tilde{\theta}) \\ D_{\phi}(\tilde{\theta}) \end{bmatrix} - \begin{bmatrix} \Sigma_{\psi\varphi}(\tilde{\theta})\Sigma_{\varphi}^{-1}(\tilde{\theta})\Delta_{nT,\varphi}(\tilde{\theta}) - \Delta_{nT,\psi}(\tilde{\theta}) \\ \Sigma_{\phi\varphi}(\tilde{\theta})\Sigma_{\varphi}^{-1}(\tilde{\theta})\Delta_{nT,\varphi}(\tilde{\theta}) - \Delta_{nT,\phi}(\tilde{\theta}) \end{bmatrix} \\ &= \begin{bmatrix} D_{\psi}(\tilde{\theta}) \\ D_{\phi}(\tilde{\theta}) \end{bmatrix} - \begin{bmatrix} \Sigma_{\psi\varphi}\Sigma_{\varphi}^{-1}\Delta_{nT,\varphi} - \Delta_{nT,\psi} \\ \Sigma_{\phi\varphi}\Sigma_{\varphi}^{-1}\Delta_{nT,\varphi} - \Delta_{nT,\phi} \end{bmatrix} + o_p(1). \end{aligned} \tag{C.7}$$

The results in (C.5) and (C.7) imply that

$$\sqrt{n^{\star}}\mathbf{C}(\tilde{\theta}) = \begin{bmatrix} \sqrt{n^{\star}}C_{\psi}(\tilde{\theta}) \\ \sqrt{n^{\star}}C_{\phi}(\tilde{\theta}) \end{bmatrix} \xrightarrow{d} N\left(\begin{bmatrix} \Sigma_{\psi\phi\cdot\varphi}\delta \\ \Sigma_{\phi\cdot\varphi}\delta \end{bmatrix}, \begin{bmatrix} \Sigma_{\psi\cdot\varphi} & \Sigma_{\psi\phi\cdot\varphi} \\ \Sigma_{\psi\phi\cdot\varphi}^{'} & \Sigma_{\phi\cdot\varphi} \end{bmatrix}\right). \tag{C.8}$$

Using (C.8) in (C.6), I get

$$\sqrt{n^\star}C_\psi^*(\tilde\theta) \xrightarrow{d} N\left(0,\, \Sigma_{\psi\cdot\varphi} - \Sigma_{\psi\phi\cdot\varphi}\Sigma_{\phi\cdot\varphi}^{-1}\Sigma_{\phi\phi\cdot\varphi}\right). \tag{C.9}$$

This last result implies that $LM_\psi^* \xrightarrow{d} \chi_r^2$.

For the asymptotic power comparisons, I need to consider the distributions of LM_ψ^{d*} and LM_ψ under H_A^ψ and H_0^ϕ. By using (C.3), I can get

$$\sqrt{n^\star}\mathbf{C}(\tilde\theta) = \begin{bmatrix}\sqrt{n^\star}C_\psi(\tilde\theta)\\ \sqrt{n^\star}C_\phi(\tilde\theta)\end{bmatrix} \xrightarrow{d} N\left(\begin{bmatrix}\Sigma_{\psi\cdot\varphi}\zeta\\ \Sigma'_{\psi\phi\cdot\varphi}\zeta\end{bmatrix}, \begin{bmatrix}\Sigma_{\psi\cdot\varphi} & \Sigma_{\psi\phi\cdot\varphi}\\ \Sigma'_{\phi\psi\cdot\varphi} & \Sigma_{\phi\cdot\varphi}\end{bmatrix}\right). \tag{C.10}$$

Using (C.10) in (C.6), I get

$$\sqrt{n^\star}C_\psi^*(\tilde\theta) \xrightarrow{d} N\left(\left(\Sigma_{\psi\cdot\varphi} - \Sigma_{\psi\phi\cdot\varphi}\Sigma_{\phi\cdot\varphi}^{-1}\Sigma_{\phi\phi\cdot\varphi}\right)\zeta,\, \Sigma_{\psi\cdot\varphi} - \Sigma_{\psi\phi\cdot\varphi}\Sigma_{\phi\cdot\varphi}^{-1}\Sigma_{\phi\phi\cdot\varphi}\right). \tag{C.11}$$

Therefore, $LM_\psi^* \xrightarrow{d} \chi_r^2(\xi_4)$, where $\xi_4 = \zeta'\left(\Sigma_{\psi\cdot\varphi} - \Sigma_{\psi\phi\cdot\varphi}\Sigma_{\phi\cdot\varphi}^{-1}\Sigma_{\phi\phi\cdot\varphi}\right)\zeta$, which proves the last part of proposition.

C.5 Proof of Proposition 5:

The proof requires the asymptotic distribution of the adjusted score function $C_\psi^\star$. Distributional misspecification changes the aysmptotic distribution of the bias corrected score function. (C.4) is no longer valid and should be:

$$\begin{bmatrix}\sqrt{n^\star}D_\varphi(\theta_0)\\ \sqrt{n^\star}\mathbf{D}(\theta_0)\end{bmatrix} + \begin{bmatrix}\Delta_{nT,\varphi}\\ \Delta_{nT,\psi}\\ \Delta_{nT,\phi}\end{bmatrix} \xrightarrow{d} N(0,\, \Omega). \tag{C.12}$$

Distributional misspecification does not change the derivatives of score functions, so (C.3), (C.6) and (C.7) are still valid. For (C.3),

$$\sqrt{n^\star}\mathbf{D}(\tilde\theta) = \left[-\boldsymbol{\Sigma}_{\psi\phi,\varphi}\Sigma_\varphi^{-1},\, I_{r+s}\right]\begin{bmatrix}\sqrt{n^\star}D_\varphi(\theta_0)\\ \sqrt{n^\star}\mathbf{D}(\theta_0)\end{bmatrix} + \begin{bmatrix}\Sigma_{\psi\cdot\varphi} & \Sigma_{\psi\phi\cdot\varphi}\\ \Sigma'_{\psi\phi\cdot\varphi} & \Sigma_{\phi\cdot\varphi}\end{bmatrix}\begin{bmatrix}\zeta\\ \delta\end{bmatrix} + o_p(1), \tag{C.13}$$

(C.12) and (C.13) imply,

$$\begin{bmatrix}D_\psi(\tilde\theta)\\ D_\phi(\tilde\theta)\end{bmatrix} - \begin{bmatrix}\Sigma_{\psi\varphi}(\tilde\theta)\Sigma_\varphi^{-1}(\tilde\theta)\Delta_{nT,\varphi}(\tilde\theta) - \Delta_{nT,\psi}(\tilde\theta)\\ \Sigma_{\phi\varphi}(\tilde\theta)\Sigma_\varphi^{-1}(\tilde\theta)\Delta_{nT,\varphi}(\tilde\theta) - \Delta_{nT,\phi}(\tilde\theta)\end{bmatrix} \to$$
$$N\left(\begin{bmatrix}\Sigma_{\psi\cdot\varphi}\zeta + \Sigma_{\psi\phi\cdot\varphi}\delta\\ \Sigma_{\phi\cdot\varphi}\delta + \Sigma_{\phi\psi\cdot\varphi}\zeta\end{bmatrix}, \begin{bmatrix}\mathcal{B}_{\psi\cdot\varphi} & \mathcal{B}_{\psi\phi\cdot\varphi}\\ \mathcal{B}_{\phi\psi\cdot\varphi} & \mathcal{B}_{\phi\cdot\varphi}\end{bmatrix}\right) \tag{C.14}$$

We have (C.6) and (C.7),

$$\sqrt{n^\star} C^*_\psi(\tilde{\theta}) = \left[I_r,\ -\Sigma_{\psi\phi\cdot\varphi}\Sigma^{-1}_{\phi\cdot\varphi} \right] \sqrt{n^\star}\mathbf{C}(\tilde{\theta}) + o_p(1), \tag{C.15}$$

$$\begin{aligned} \sqrt{n^\star}\mathbf{C}(\tilde{\theta}) &= \begin{bmatrix} \sqrt{n^\star}C_\psi(\tilde{\theta}) \\ \sqrt{n^\star}C_\phi(\tilde{\theta}) \end{bmatrix} = \begin{bmatrix} D_\psi(\tilde{\theta}) \\ D_\phi(\tilde{\theta}) \end{bmatrix} - \begin{bmatrix} \Sigma_{\psi\varphi}(\tilde{\theta})\Sigma^{-1}_{\varphi}(\tilde{\theta})\Delta_{nT,\varphi}(\tilde{\theta}) - \Delta_{nT,\psi}(\tilde{\theta}) \\ \Sigma_{\phi\varphi}(\tilde{\theta})\Sigma^{-1}_{\varphi}(\tilde{\theta})\Delta_{nT,\varphi}(\tilde{\theta}) - \Delta_{nT,\phi}(\tilde{\theta}) \end{bmatrix} \\ &= \begin{bmatrix} D_\psi(\tilde{\theta}) \\ D_\phi(\tilde{\theta}) \end{bmatrix} - \begin{bmatrix} \Sigma_{\psi\varphi}\Sigma^{-1}_{\varphi}\Delta_{nT,\varphi} - \Delta_{nT,\psi} \\ \Sigma_{\phi\varphi}\Sigma^{-1}_{\varphi}\Delta_{nT,\varphi} - \Delta_{nT,\phi} \end{bmatrix} + o_p(1). \end{aligned} \tag{C.16}$$

together with (C.14),

$$\begin{aligned} \sqrt{n^\star} C^*_\psi(\tilde{\theta}) &= \left[I_r,\ -\Sigma_{\psi\phi\cdot\varphi}\Sigma^{-1}_{\phi\cdot\varphi} \right] \left(\begin{bmatrix} D_\psi(\tilde{\theta}) \\ D_\phi(\tilde{\theta}) \end{bmatrix} - \begin{bmatrix} \Sigma_{\psi\varphi}(\tilde{\theta})\Sigma^{-1}_{\varphi}(\tilde{\theta})\Delta_{nT,\varphi}(\tilde{\theta}) - \Delta_{nT,\psi}(\tilde{\theta}) \\ \Sigma_{\phi\varphi}(\tilde{\theta})\Sigma^{-1}_{\varphi}(\tilde{\theta})\Delta_{nT,\varphi}(\tilde{\theta}) - \Delta_{nT,\phi}(\tilde{\theta}) \end{bmatrix} \right) \\ &\to N\left(\left(\Sigma_{\psi\cdot\varphi}(\theta_0) - \Sigma_{\psi\phi\cdot\varphi}(\theta_0)\Sigma^{-1}_{\phi\cdot\varphi}(\theta_0)\Sigma_{\phi\psi\cdot\varphi}(\theta_0) \right)\zeta, \mathcal{D}_{\psi\cdot\varphi}(\theta_0) \right) \end{aligned} \tag{C.17}$$

The second part of Proposition 5 directly follows from above. Consider under H_0^ψ, $\zeta = 0$, the above result then implies the first part of proposition.

C.6 Proof of Corollary 1:

I begin with the proof of the first part of corollary, where ψ is a subset of (γ, ρ). Using either the direct or transformation approach,

$$LM_\psi = n^\star C'_\psi(\tilde{\theta})\Sigma^{-1}_{\psi\cdot\varphi}(\tilde{\theta})C_\psi(\tilde{\theta}), \tag{C.18}$$

$$LM_\psi(D) = n^\star C'_\psi(\tilde{\theta})\mathcal{B}^{-1}_{\psi\cdot\varphi}(\tilde{\theta})C_\psi(\tilde{\theta}), \tag{C.19}$$

As long as $\Sigma_{\psi\cdot\varphi}(\tilde{\theta}) = \mathcal{B}_{\psi\cdot\varphi}(\tilde{\theta})$, we have $LM_\psi = LM_\psi(D)$.

$$\Sigma_{\psi\cdot\varphi}(\tilde{\theta}) = \left[\Sigma_\psi(\tilde{\theta}) - \Sigma_{\psi\varphi}(\tilde{\theta})\Sigma^{-1}_\varphi(\tilde{\theta})\Sigma_{\varphi\psi}(\tilde{\theta}) \right] \tag{C.20}$$

$$\begin{aligned} \mathcal{B}_{\psi\cdot\varphi}(\tilde{\theta}) &= \Omega_\psi(\tilde{\theta}) - \Sigma_{\psi\varphi}(\tilde{\theta})\Sigma^{-1}_\varphi(\tilde{\theta})\Omega_{\varphi\psi}(\tilde{\theta}) - \Omega_{\psi\varphi}(\tilde{\theta})\Sigma^{-1}_\varphi(\tilde{\theta})\Sigma_{\varphi\psi}(\tilde{\theta}) \\ &+ \Sigma_{\psi\varphi}(\tilde{\theta})\Sigma^{-1}_\varphi(\tilde{\theta})\Omega_\varphi(\tilde{\theta})\Sigma^{-1}_\varphi(\tilde{\theta})\Sigma_{\varphi\psi}(\tilde{\theta}) \end{aligned} \tag{C.21}$$

By Appendix A,

$$\Omega_\psi(\tilde{\theta}) = \Sigma_\psi(\tilde{\theta}), \tag{C.22}$$

and

$$\Omega_{\varphi,\psi}(\tilde{\theta}) = \Sigma_{\varphi,\psi}(\tilde{\theta}), \tag{C.23}$$

$$\Omega_{\psi,\varphi}(\tilde{\theta}) = \Sigma_{\psi,\varphi}(\tilde{\theta}), \tag{C.24}$$

Then,

$$\Sigma_{\psi\varphi}(\tilde{\theta})\Sigma_{\varphi}^{-1}(\tilde{\theta})\Sigma_{\varphi\psi}(\tilde{\theta}) = \Sigma_{\psi\varphi}(\tilde{\theta})\Sigma_{\varphi}^{-1}(\tilde{\theta})\Omega_{\varphi\psi}(\tilde{\theta}) = \Omega_{\psi\varphi}(\tilde{\theta})\Sigma_{\varphi}^{-1}(\tilde{\theta})\Sigma_{\varphi\psi}(\tilde{\theta}). \tag{C.25}$$

For the last term of $\mathcal{B}_{\psi\cdot\varphi}(\tilde{\theta})$,

$$\Omega_{\varphi}(\tilde{\theta}) = \begin{bmatrix} \Sigma_{\beta}(\tilde{\theta}) & 0 \\ 0 & \Omega_{\sigma^2}(\tilde{\theta}) \end{bmatrix}, \Sigma_{\varphi}(\tilde{\theta}) = \begin{bmatrix} \Sigma_{\beta}(\tilde{\theta}) & 0 \\ 0 & \Sigma_{\sigma^2}(\tilde{\theta}) \end{bmatrix} \tag{C.26}$$

so

$$\Sigma_{\varphi}^{-1}(\tilde{\theta})\Omega_{\varphi}(\tilde{\theta})\Sigma_{\varphi}^{-1}(\tilde{\theta}) = \begin{bmatrix} \Sigma_{\beta}(\tilde{\theta})^{-1} & 0 \\ 0 & \Sigma_{\sigma^2}(\tilde{\theta})^{-1}\Omega_{\sigma^2}(\tilde{\theta})^{-1}\Sigma_{\sigma^2}(\tilde{\theta})^{-1} \end{bmatrix} \tag{C.27}$$

Since ψ is a subset of (γ, ρ), $\Sigma_{\psi\sigma^2} = 0$,

$$\Sigma_{\psi\varphi}(\tilde{\theta}) = \Sigma_{\varphi\psi}^{'}(\tilde{\theta}) = \begin{bmatrix}\Sigma_{\psi\beta}(\tilde{\theta}) & \Sigma_{\psi\sigma^2}(\tilde{\theta})\end{bmatrix} = \begin{bmatrix}\Sigma_{\psi\beta}(\tilde{\theta}) & 0\end{bmatrix} \tag{C.28}$$

Plug it into $\Sigma_{\psi\varphi}(\tilde{\theta})\Sigma_{\varphi}^{-1}(\tilde{\theta})\Omega_{\varphi}(\tilde{\theta})\Sigma_{\varphi}^{-1}(\tilde{\theta})\Sigma_{\varphi\psi}(\tilde{\theta})$ and $\Sigma_{\psi\varphi}(\tilde{\theta})\Sigma_{\varphi}^{-1}(\tilde{\theta})\Sigma_{\varphi\psi}(\tilde{\theta})$,

$$\Sigma_{\psi\varphi}(\tilde{\theta})\Sigma_{\varphi}^{-1}(\tilde{\theta})\Omega_{\varphi}(\tilde{\theta})\Sigma_{\varphi}^{-1}(\tilde{\theta})\Sigma_{\varphi\psi}(\tilde{\theta}) = \Sigma_{\psi\beta}(\tilde{\theta})\Sigma_{\beta}^{-1}(\tilde{\theta})\Sigma_{\beta\psi}(\tilde{\theta}) = \Sigma_{\psi\varphi}(\tilde{\theta})\Sigma_{\varphi}^{-1}(\tilde{\theta})\Sigma_{\varphi\psi}(\tilde{\theta}) \tag{C.29}$$

(C.22), (C.25) and (C.29) imply:

$$\mathcal{B}_{\psi\cdot\varphi}(\tilde{\theta}) = \Sigma_{\psi\cdot\varphi}(\tilde{\theta}), \tag{C.30}$$

so,

$$LM_{\psi} = LM_{\psi}(D). \tag{C.31}$$

For the second part of the corollary, let's first consider the transformation approach case, let $\eta = (\gamma, \rho, \beta)'$.

$$\Omega(\tilde{\theta}) = \Sigma(\tilde{\theta}) + \mathcal{S} =$$

$$\begin{bmatrix} \Sigma_{\eta}(\tilde{\theta}) & \Sigma_{\eta\lambda}(\tilde{\theta}) & 0 \\ \Sigma_{\lambda\eta}(\tilde{\theta}) & \Sigma_{\lambda}(\tilde{\theta}) & \frac{-1}{\tilde{\sigma}^2(n-1)} \\ 0 & \frac{-1}{\tilde{\sigma}^2(n-1)} & \Sigma_{\sigma^2}(\tilde{\theta}) \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{\tilde{\mu}_4 - 3\tilde{\sigma}^4}{\tilde{\sigma}^4(n-1)} \sum_{i=1}^{n}[(J_nW_n)_{ii}]^2 & -\frac{\tilde{\mu}_4 - 3\tilde{\sigma}^4}{2\tilde{\sigma}^6(n-1)} \\ 0 & -\frac{\tilde{\mu}_4 - 3\tilde{\sigma}^4}{2\tilde{\sigma}^6(n-1)} & \frac{\tilde{\mu}_4 - 3\tilde{\sigma}^4}{\tilde{\sigma}^4}\frac{1}{4\tilde{\sigma}^2} \end{bmatrix} \tag{C.32}$$

Consider that each individual has a number of neighbors (n_b) that grows slower than $\sqrt{n}$ or is bounded. W_n in this case is sparse.

$$[(J_nW_n)_{ii}]^2 = O(1)\frac{n_b^2}{n^2} = o(\frac{1}{n}) \tag{C.33}$$

so,

$$\frac{\tilde{\mu}_4 - 3\tilde{\sigma}^4}{\tilde{\sigma}^4} \frac{1}{(n-1)} \sum_{i=1}^{n} [(J_n W_n)_{ii}]^2 = o(\frac{1}{n}) \tag{C.34}$$

Then,

$$\mathcal{S} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \frac{\tilde{\mu}_4 - 3\tilde{\sigma}^4}{\tilde{\sigma}^4} \frac{1}{4\tilde{\sigma}^2} \end{bmatrix} + O(\frac{1}{n}) \tag{C.35}$$

Note that $\Sigma_{\lambda\sigma^2}(\tilde{\theta}) \to 0$, which means λ and σ^2 are asymptotically uncorrelated. Given what we have now, (C.22) through (C.29) are valid even when ψ contains λ, so $LM_\psi(D) \to LM_\psi$.

Next I show $LM_\psi(DP) \to LM_\psi(P)$.

$$LM_\psi(DP) = n^\star C_\psi^{*'}(\tilde{\theta}) \mathcal{D}_{\psi\cdot\varphi}^{-1}(\tilde{\theta}) C_\psi^*(\tilde{\theta}), \tag{C.36}$$

$$LM_\psi(P) = n^\star C_\psi^{*'}(\tilde{\theta}) \left[\Sigma_{\psi\cdot\varphi}(\tilde{\theta}) - \Sigma_{\psi\phi\cdot\varphi}(\tilde{\theta}) \Sigma_{\phi\cdot\varphi}^{-1}(\tilde{\theta}) \Sigma_{\psi\phi\cdot\varphi}^{'}(\tilde{\theta})\right]^{-1} C_\psi^*(\tilde{\theta}) \tag{C.37}$$

It suffices to show $\mathcal{D}_{\psi\cdot\varphi}(\tilde{\theta}) \to \Sigma_{\psi\cdot\varphi}(\tilde{\theta}) - \Sigma_{\psi\phi\cdot\varphi}(\tilde{\theta}) \Sigma_{\phi\cdot\varphi}^{-1}(\tilde{\theta}) \Sigma_{\psi\phi\cdot\varphi}^{'}(\tilde{\theta})$. We have,

$$\begin{aligned} \mathcal{D}_{\psi\cdot\varphi}(\tilde{\theta}) &= \mathcal{B}_{\psi\cdot\varphi}(\tilde{\theta}) - \Sigma_{\psi\phi\cdot\varphi}(\tilde{\theta}) \Sigma_{\phi\cdot\varphi}^{-1}(\tilde{\theta}) \mathcal{B}_{\phi\psi\cdot\varphi}(\tilde{\theta}) - \mathcal{B}_{\psi\phi\cdot\varphi}(\tilde{\theta}) \Sigma_{\phi\cdot\varphi}^{-1}(\tilde{\theta}) \Sigma_{\phi\psi\cdot\varphi}(\tilde{\theta}) \\ &+ \Sigma_{\psi\phi\cdot\varphi}(\tilde{\theta}) \Sigma_{\phi\cdot\varphi}^{-1}(\tilde{\theta}) \mathcal{B}_{\phi\cdot\varphi}(\tilde{\theta}) \Sigma_{\phi\cdot\varphi}^{-1}(\tilde{\theta}) \Sigma_{\phi\psi\cdot\varphi}(\tilde{\theta}) \end{aligned} \tag{C.38}$$

$$\begin{aligned} \mathcal{B}_{\psi\phi\cdot\varphi}(\tilde{\theta}) &= \Omega_{\psi\phi}(\tilde{\theta}) - \Sigma_{\psi\varphi}(\tilde{\theta}) \Sigma_\varphi^{-1}(\tilde{\theta}) \Omega_{\varphi\phi}(\tilde{\theta}) - \Omega_{\psi\varphi}(\tilde{\theta}) \Sigma_\varphi^{-1}(\tilde{\theta}) \Sigma_{\varphi\phi}(\tilde{\theta}) \\ &+ \Sigma_{\psi\varphi}(\tilde{\theta}) \Sigma_\varphi^{-1}(\tilde{\theta}) \Omega_\varphi(\tilde{\theta}) \Sigma_\varphi^{-1}(\tilde{\theta}) \Sigma_{\varphi\phi}(\tilde{\theta}) \end{aligned} \tag{C.39}$$

. Compare (C.39) with

$$\Sigma_{\psi\phi\cdot\varphi}(\tilde{\theta}) = \Sigma_{\psi\phi}(\tilde{\theta}) - \Sigma_{\psi\varphi}(\tilde{\theta}) \Sigma_\varphi^{-1}(\tilde{\theta}) \Sigma_{\varphi\phi}(\tilde{\theta}), \tag{C.40}$$

Again, by checking Appendix A, now ϕ contains λ, $\Sigma_{\phi\sigma^2}(\tilde{\theta}) = O(1/n)$. The following equations hold,

$$\Sigma_{\psi\phi}(\tilde{\theta}) = \Omega_{\psi\phi}(\tilde{\theta}), \tag{C.41}$$

$$\Sigma_{\psi\varphi}(\tilde{\theta})\Sigma_{\varphi}^{-1}(\tilde{\theta})\Sigma_{\varphi\phi}(\tilde{\theta}) = \begin{bmatrix}\Sigma_{\psi\beta}(\tilde{\theta}) & 0\end{bmatrix}\begin{bmatrix}\Sigma_{\beta}(\tilde{\theta})^{-1} & 0 \\ 0 & \Sigma_{\sigma^2}(\tilde{\theta})^{-1}\end{bmatrix}\begin{bmatrix}\Sigma_{\beta\phi}(\tilde{\theta}) \\ \Sigma_{\sigma^2\phi}(\tilde{\theta})\end{bmatrix} + O(\frac{1}{n}) \tag{C.42}$$

$$= \begin{bmatrix}\Sigma_{\psi\beta}(\tilde{\theta})\Sigma_{\beta}^{-1}(\tilde{\theta})\Sigma_{\beta\phi}(\tilde{\theta})\end{bmatrix} + O(\frac{1}{n}), \tag{C.43}$$

$$\Sigma_{\psi\varphi}(\tilde{\theta})\Sigma_{\varphi}^{-1}(\tilde{\theta})\Omega_{\varphi\phi}(\tilde{\theta}) = \begin{bmatrix}\Sigma_{\psi\beta}(\tilde{\theta}) & 0\end{bmatrix}\begin{bmatrix}\Sigma_{\beta}(\tilde{\theta})^{-1} & 0 \\ 0 & \Sigma_{\sigma^2}(\tilde{\theta})^{-1}\end{bmatrix}\begin{bmatrix}\Sigma_{\beta\phi}(\tilde{\theta}) \\ \Omega_{\sigma^2\phi}(\tilde{\theta})\end{bmatrix} + O(\frac{1}{n}) \tag{C.44}$$

$$= \begin{bmatrix}\Sigma_{\psi\beta}(\tilde{\theta})\Sigma_{\beta}^{-1}(\tilde{\theta})\Sigma_{\beta\phi}(\tilde{\theta})\end{bmatrix} + O(\frac{1}{n}), \tag{C.45}$$

$$\Omega_{\psi\varphi}(\tilde{\theta}) = \Sigma_{\psi\varphi}(\tilde{\theta}) + O(\frac{1}{n}), \tag{C.46}$$

$$\Omega_{\psi\varphi}(\tilde{\theta})\Sigma_{\varphi}^{-1}(\tilde{\theta})\Sigma_{\varphi\phi}(\tilde{\theta}) = \Sigma_{\psi\varphi}(\tilde{\theta})\Sigma_{\varphi}^{-1}(\tilde{\theta})\Sigma_{\varphi\phi}(\tilde{\theta}) + O(\frac{1}{n}), \tag{C.47}$$

$$\Sigma_{\psi\varphi}(\tilde{\theta})\Sigma_{\varphi}^{-1}(\tilde{\theta})\Omega_{\varphi}(\tilde{\theta})\Sigma_{\varphi}^{-1}(\tilde{\theta})\Sigma_{\varphi\phi}(\tilde{\theta}) = \begin{bmatrix}\Sigma_{\psi\beta}(\tilde{\theta}) & 0\end{bmatrix} \tag{C.48}$$

$$\begin{bmatrix}\Sigma_{\beta}(\tilde{\theta})^{-1} & 0 \\ 0 & \Sigma_{\sigma^2}(\tilde{\theta})^{-1}\Omega_{\sigma^2}(\tilde{\theta})^{-1}\Sigma_{\sigma^2}(\tilde{\theta})^{-1}\end{bmatrix}\begin{bmatrix}\Sigma_{\beta\phi}(\tilde{\theta}) \\ \Sigma_{\sigma^2\phi}(\tilde{\theta})\end{bmatrix} + O(\frac{1}{n})$$

$$= \begin{bmatrix}\Sigma_{\psi\beta}(\tilde{\theta})\Sigma_{\beta}^{-1}(\tilde{\theta})\Sigma_{\beta\phi}(\tilde{\theta})\end{bmatrix} + O(\frac{1}{n}) \tag{C.49}$$

(C.41) through (C.49) then imply

$$\mathcal{B}_{\psi\phi\cdot\varphi}(\tilde{\theta}) = \Sigma_{\psi\phi\cdot\varphi}(\tilde{\theta}) + O(\frac{1}{n}) \tag{C.50}$$

Take transpose, $\mathcal{B}_{\phi\psi\cdot\varphi}(\tilde{\theta}) = \Sigma_{\phi\psi\cdot\varphi}(\tilde{\theta}) + O(\frac{1}{n})$. (C.50) further implies,

$$\mathcal{B}_{\psi\cdot\varphi}(\tilde{\theta}) = \Sigma_{\psi\cdot\varphi}(\tilde{\theta}) + O(\frac{1}{n}) \tag{C.51}$$

$$\mathcal{B}_{\phi\cdot\varphi}(\tilde{\theta}) = \Sigma_{\phi\cdot\varphi}(\tilde{\theta}) + O(\frac{1}{n}) \tag{C.52}$$

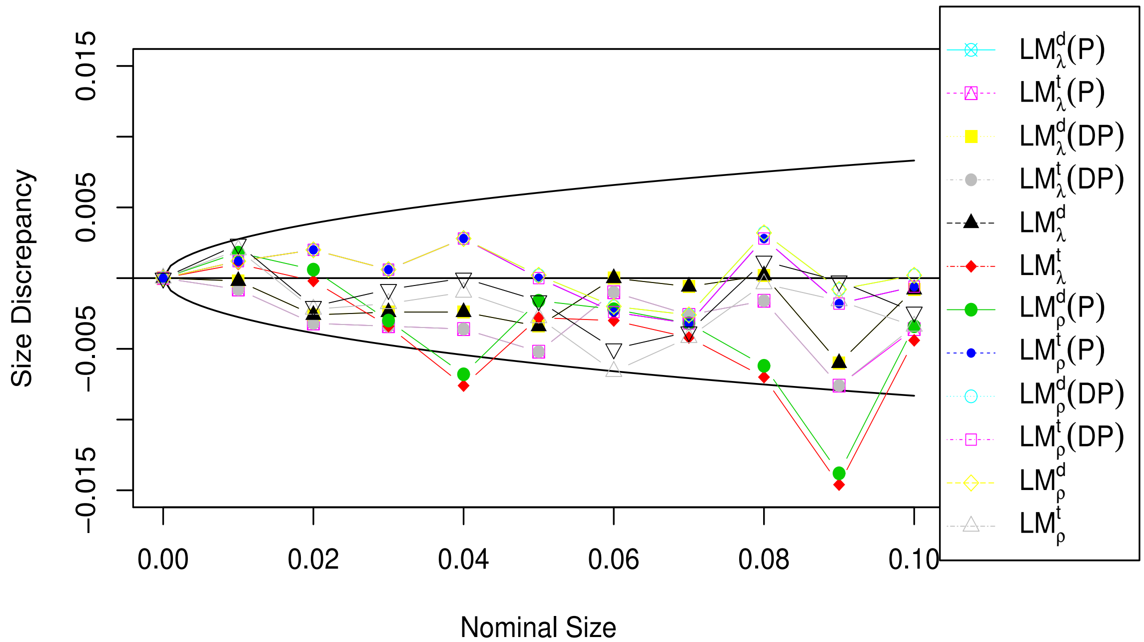
Then,

$$\mathcal{D}_{\psi\cdot\varphi}(\tilde{\theta}) = \begin{bmatrix}\Sigma_{\psi}(\tilde{\theta}) - \Sigma_{\psi\varphi}(\tilde{\theta})\Sigma_{\varphi}^{-1}(\tilde{\theta})\Sigma_{\varphi\psi}(\tilde{\theta})\end{bmatrix} + O(\frac{1}{n}) = \Sigma_{\psi\cdot\varphi}(\tilde{\theta}) + O(\frac{1}{n}) \tag{C.53}$$

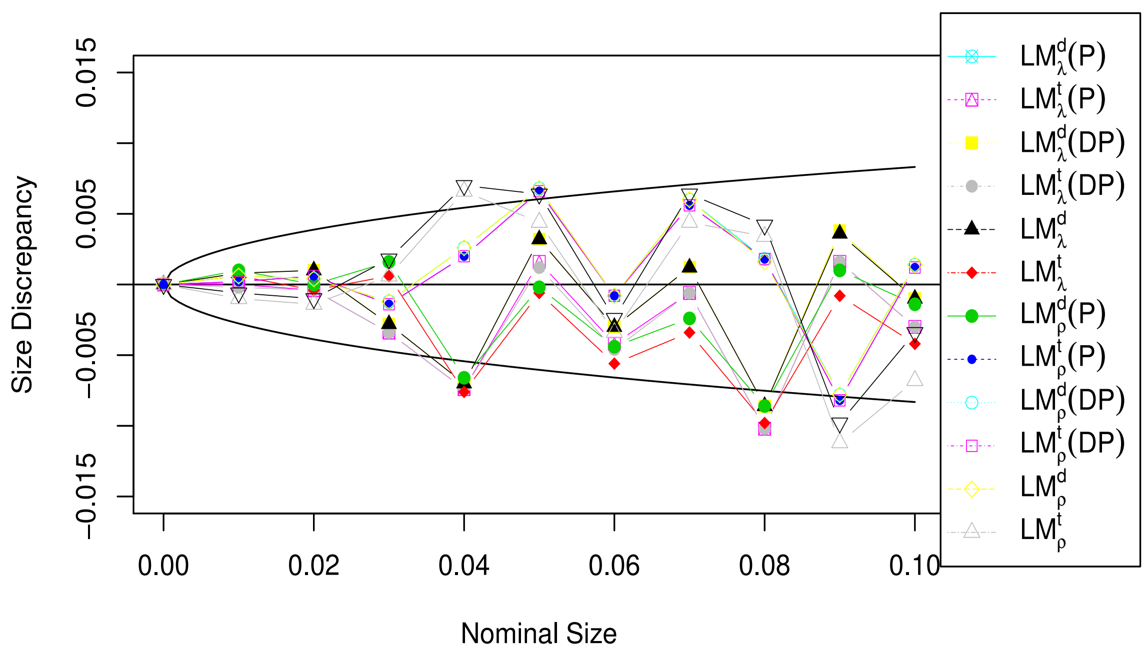
Therefore, $LM_{\psi}(DP) \to LM_{\psi}(P)$. For direct approach, (C.35) is still true. In fact, the equation holds without adding the remainder term. The rest of the proof remains valid, thus we still have $LM_{\psi}(DP) \to LM_{\psi}(P)$.

D Simulation Results

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

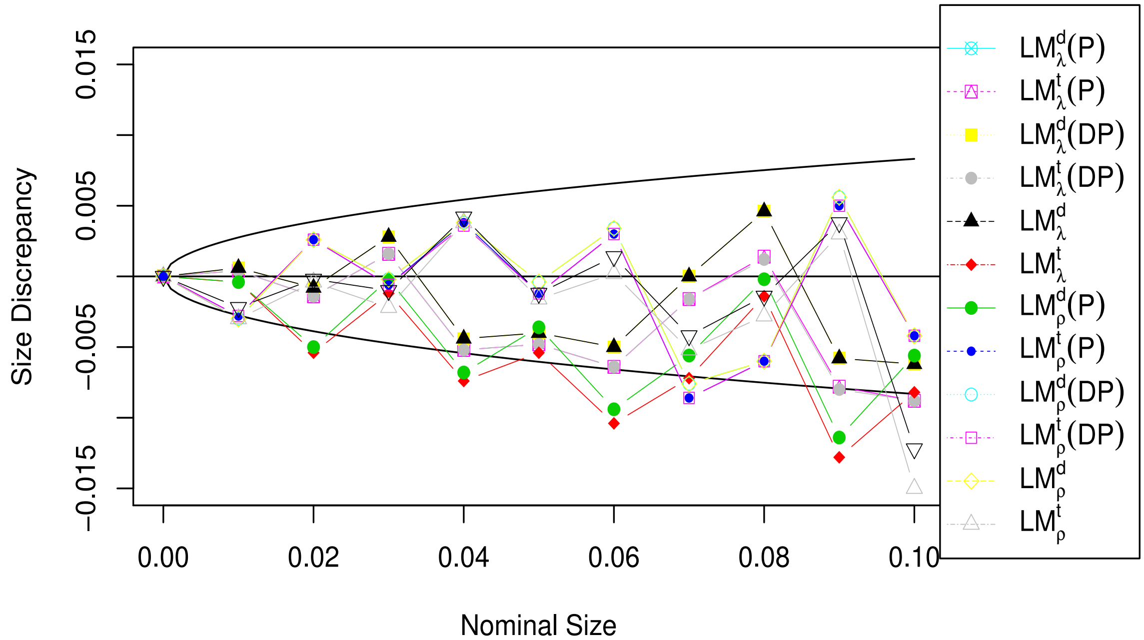


(a) Normal errors

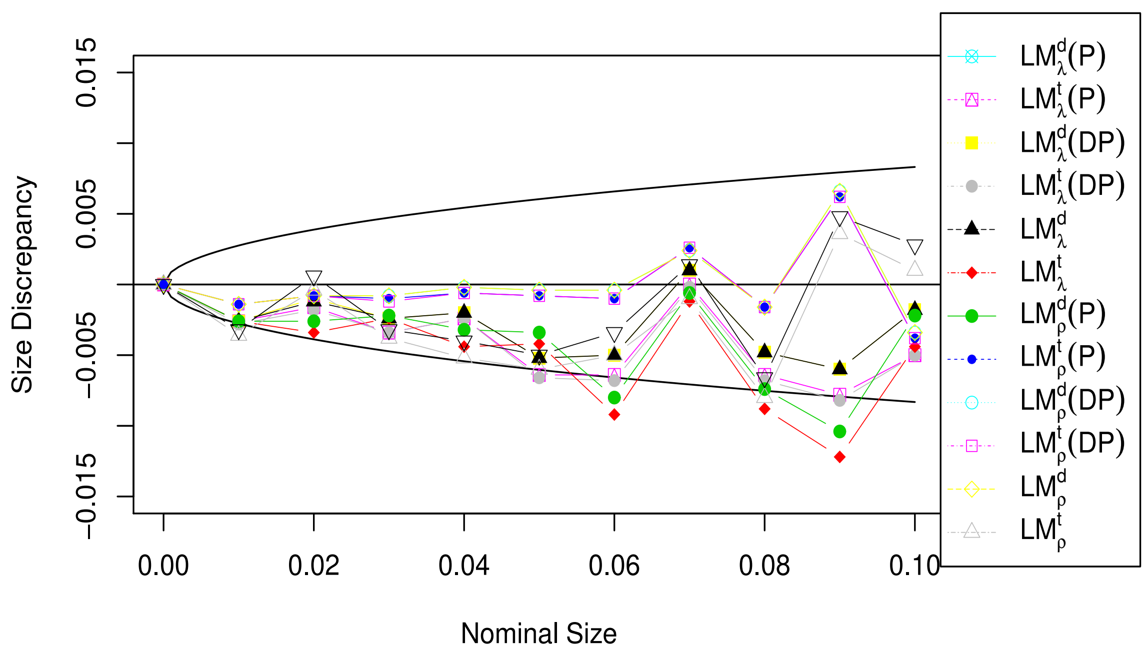


(b) T errors

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

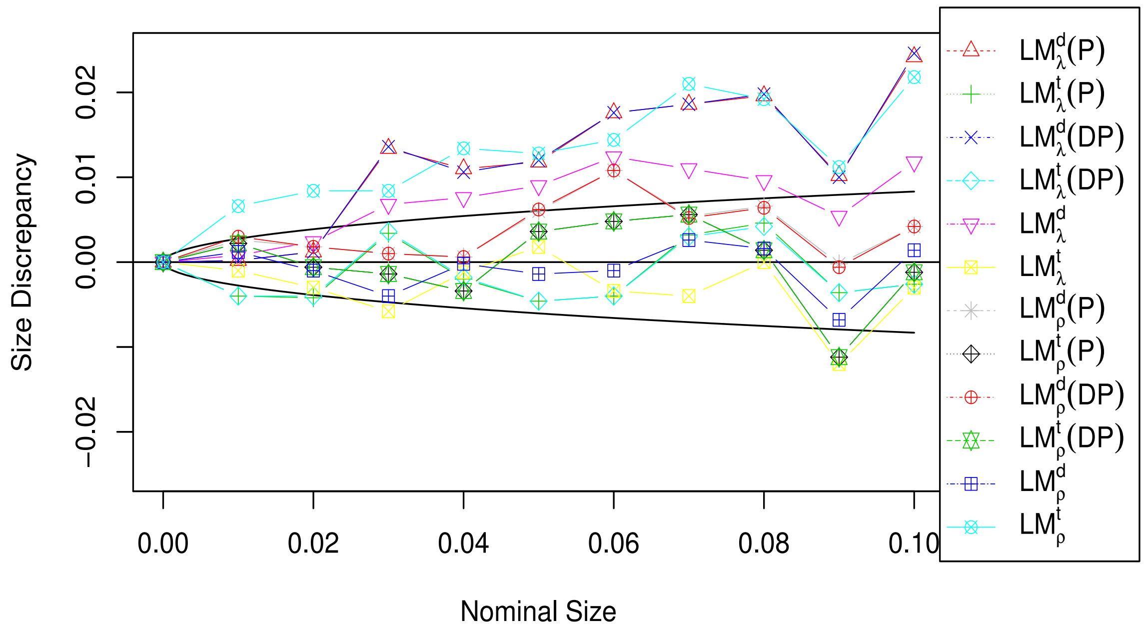


(a) Chi-square errors

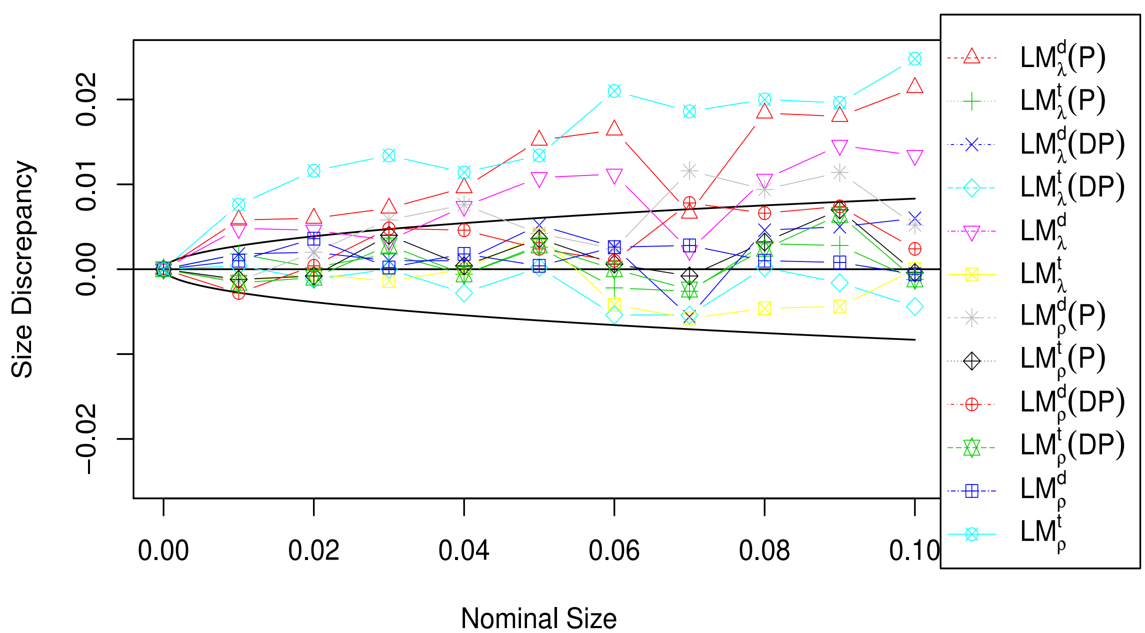


(b) Mixture normal errors

Figure 3: Size discrepancy plots when $(n, T) = (9, 111)$.

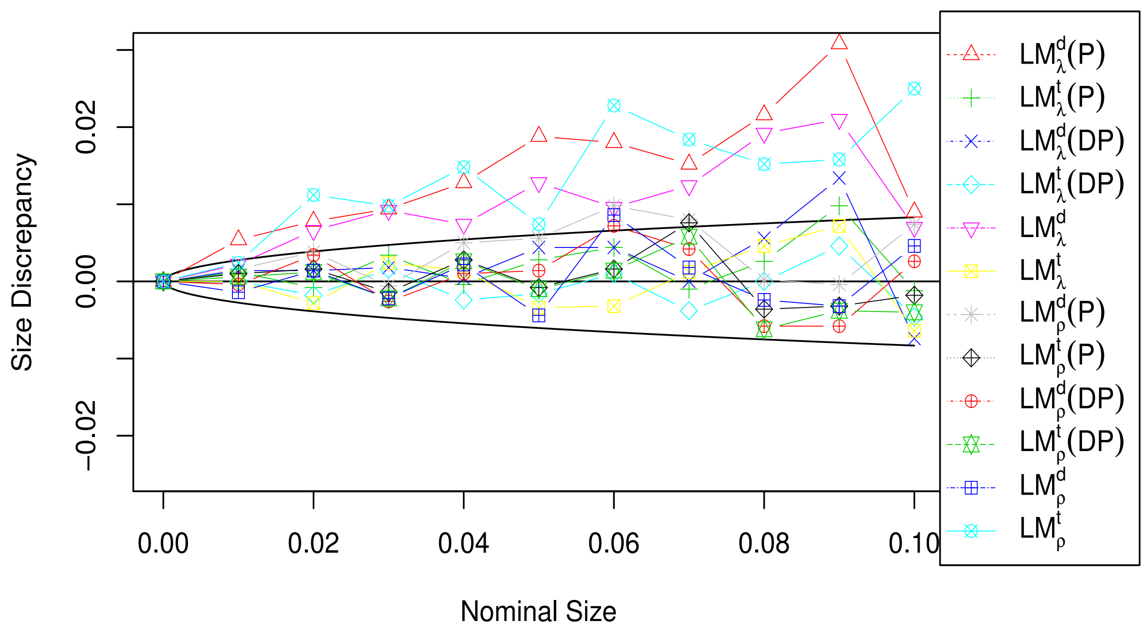


(a) Normal errors

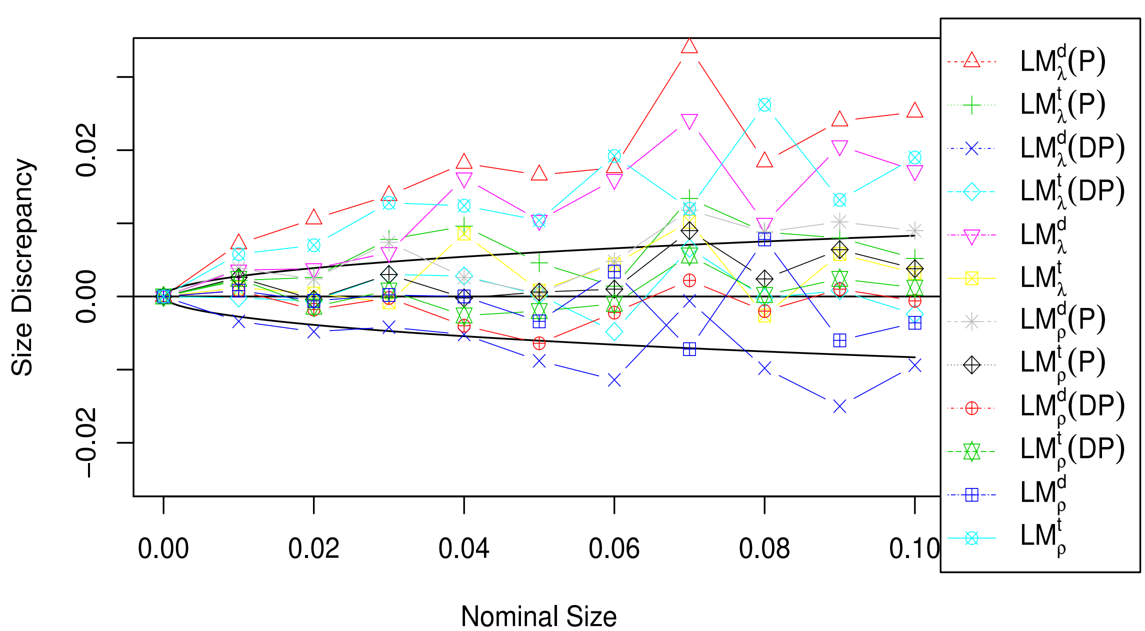


(b) T errors

Figure 4: Size discrepancy plots when $(n, T) = (9, 111)$.

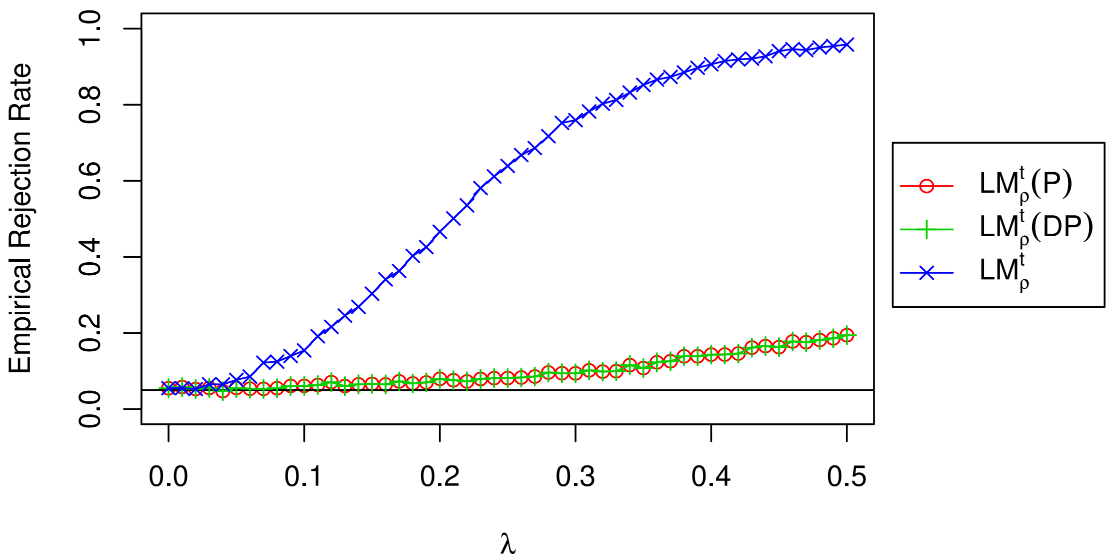


(a) Chi-square errors

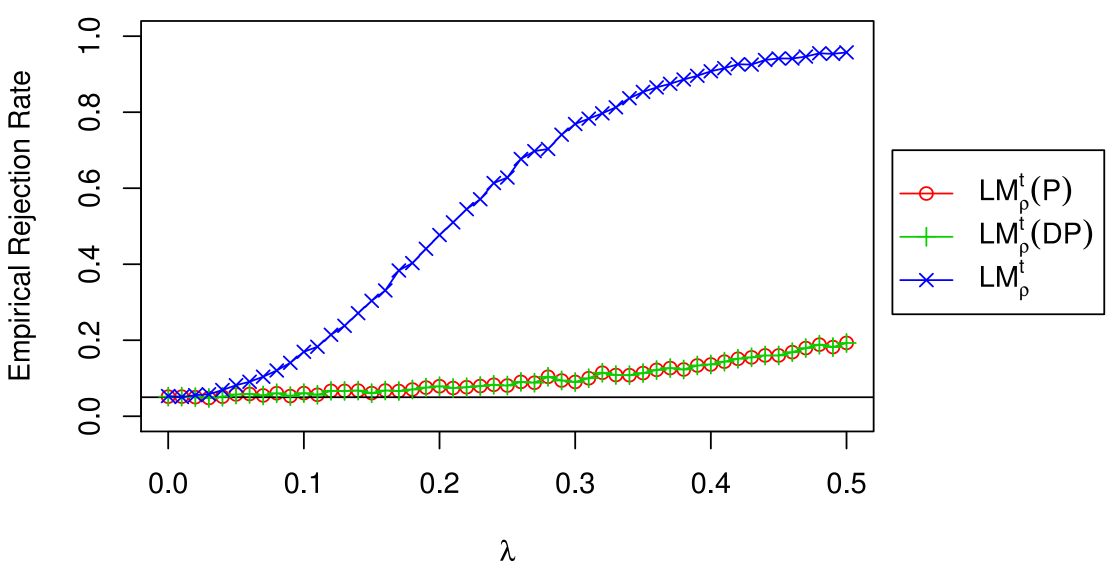


(b) Mixture normal errors

Figure 5: Empirical rejection rate for misspecified model.

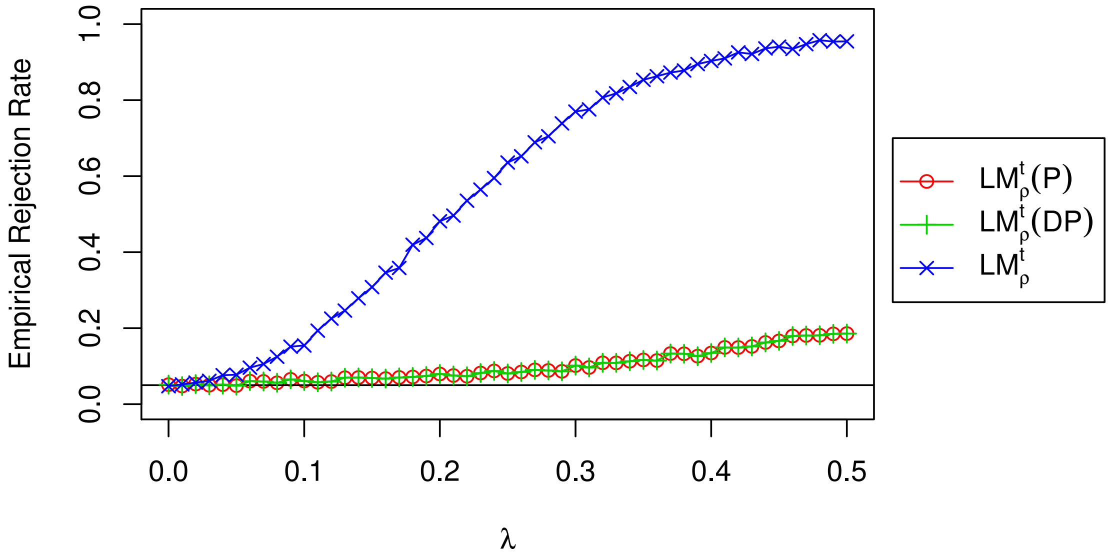


(a) Normal errors, $(n, T) = (49, 10)$

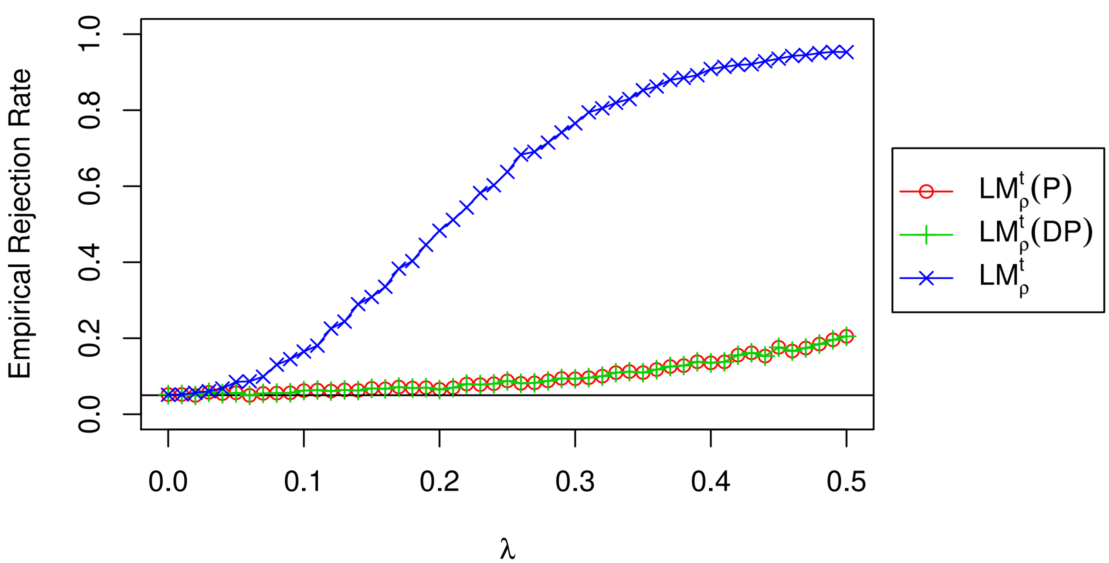


(b) T errors, $(n, T) = (49, 10)$

Figure 6: Empirical rejection rate for misspecified model.



(a) Chi-square errors, $(n, T) = (49, 10)$



(b) Mixture normal errors, $(n, T) = (49, 10)$

Table 1: Empirical sizes when : $\gamma \neq 0$ I

γ_0	$LM_\rho^t(P)$	$LM_\rho^t(DP)$	LM_ρ^t	$LM_\lambda^t(P)$	$LM_\lambda^t(DP)$	LM_λ^t
0.010	0.048	0.048	0.050	0.049	0.049	0.055
0.020	0.047	0.047	0.053	0.042	0.042	0.048
0.030	0.046	0.046	0.059	0.044	0.044	0.051
0.040	0.051	0.051	0.065	0.041	0.041	0.050
0.050	0.043	0.043	0.062	0.038	0.038	0.048
0.100	0.033	0.033	0.089	0.035	0.035	0.058
0.200	0.013	0.013	0.152	0.025	0.025	0.071
0.300	0.003	0.003	0.243	0.019	0.019	0.095

$(n, T) = (49, 10)$, normal distribution.

Table 2: Empirical sizes when : $\gamma \neq 0$ II

γ_0	$LM_\rho^t(P)$	$LM_\rho^t(DP)$	LM_ρ^t	$LM_\lambda^t(P)$	$LM_\lambda^t(DP)$	LM_λ^t
0.010	0.046	0.046	0.048	0.042	0.042	0.046
0.020	0.050	0.050	0.056	0.044	0.044	0.048
0.030	0.044	0.044	0.059	0.037	0.037	0.044
0.040	0.043	0.043	0.056	0.042	0.042	0.043
0.050	0.041	0.041	0.058	0.039	0.039	0.046
0.100	0.035	0.035	0.086	0.031	0.031	0.055
0.200	0.015	0.015	0.159	0.023	0.022	0.077
0.300	0.003	0.003	0.233	0.023	0.023	0.100

$(n, T) = (49, 10)$, t distribution.

Table 3: Empirical sizes when : $\gamma \neq 0$ III

γ_0	$LM_\rho^t(P)$	$LM_\rho^t(DP)$	LM_ρ^t	$LM_\lambda^t(P)$	$LM_\lambda^t(DP)$	LM_λ^t
0.010	0.046	0.046	0.053	0.044	0.044	0.048
0.020	0.050	0.050	0.054	0.049	0.049	0.051
0.030	0.049	0.049	0.059	0.048	0.048	0.048
0.040	0.048	0.048	0.062	0.043	0.043	0.054
0.050	0.043	0.043	0.065	0.045	0.045	0.057
0.100	0.036	0.036	0.083	0.036	0.036	0.061
0.200	0.015	0.015	0.155	0.023	0.023	0.071
0.300	0.004	0.004	0.238	0.022	0.022	0.096

$(n, T) = (49, 10)$, chi-square distribution.

Table 4: Empirical sizes when : $\gamma \neq 0$ IV

γ_0	$LM^t_\rho(P)$	$LM^t_\rho(DP)$	LM^t_ρ	$LM^t_\lambda(P)$	$LM^t_\lambda(DP)$	LM^t_λ
0.010	0.041	0.041	0.043	0.045	0.045	0.041
0.020	0.047	0.047	0.053	0.047	0.047	0.046
0.030	0.044	0.044	0.060	0.042	0.042	0.052
0.040	0.043	0.043	0.058	0.041	0.041	0.047
0.050	0.047	0.047	0.064	0.038	0.038	0.048
0.100	0.030	0.030	0.080	0.033	0.033	0.056
0.200	0.014	0.014	0.151	0.024	0.024	0.073
0.300	0.004	0.004	0.243	0.025	0.025	0.104

$(n, T) = (49, 10)$, mixture normal distribution.

Table 5: Empirical sizes when : $\lambda \neq 0$ I

λ_0	$LM^t_\gamma(P)$	$LM^t_\gamma(DP)$	LM^t_γ	$LM^t_\rho(P)$	$LM^t_\rho(DP)$	LM^t_ρ
0.010	0.043	0.043	0.046	0.054	0.054	0.053
0.020	0.053	0.053	0.057	0.055	0.055	0.055
0.030	0.050	0.050	0.049	0.050	0.050	0.060
0.040	0.045	0.045	0.048	0.053	0.053	0.071
0.050	0.048	0.048	0.049	0.054	0.054	0.080
0.100	0.052	0.052	0.055	0.059	0.059	0.160
0.200	0.041	0.041	0.060	0.082	0.082	0.468
0.300	0.042	0.042	0.112	0.103	0.103	0.758

$(n, T) = (49, 10)$, normal distribution.

Table 6: Empirical sizes when : $\lambda \neq 0$ II

λ_0	$LM^t_\gamma(P)$	$LM^t_\gamma(DP)$	LM^t_γ	$LM^t_\rho(P)$	$LM^t_\rho(DP)$	LM^t_ρ
0.010	0.046	0.046	0.048	0.051	0.051	0.053
0.020	0.049	0.049	0.050	0.056	0.056	0.056
0.030	0.048	0.048	0.049	0.053	0.053	0.063
0.040	0.049	0.049	0.050	0.051	0.051	0.064
0.050	0.051	0.051	0.051	0.052	0.052	0.084
0.100	0.045	0.045	0.050	0.064	0.064	0.159
0.200	0.049	0.049	0.072	0.073	0.073	0.468
0.300	0.043	0.043	0.117	0.093	0.092	0.767

$(n, T) = (49, 10)$, t distribution.

Table 7: Empirical sizes when : $\lambda \neq 0$ III

λ_0	$LM^t_\gamma(P)$	$LM^t_\gamma(DP)$	LM^t_γ	$LM^t_\rho(P)$	$LM^t_\rho(DP)$	LM^t_ρ
0.010	0.047	0.047	0.050	0.048	0.048	0.053
0.020	0.051	0.051	0.053	0.052	0.052	0.058
0.030	0.050	0.050	0.054	0.053	0.053	0.064
0.040	0.051	0.051	0.054	0.051	0.051	0.068
0.050	0.045	0.045	0.049	0.057	0.057	0.076
0.100	0.041	0.041	0.048	0.053	0.053	0.169
0.200	0.046	0.046	0.075	0.068	0.068	0.479
0.300	0.043	0.043	0.129	0.096	0.096	0.763

$(n, T) = (49, 10)$, chi-square distribution.

Table 8: Empirical sizes when : $\lambda \neq 0$ IV

λ_0	$LM^t_\gamma(P)$	$LM^t_\gamma(DP)$	LM^t_γ	$LM^t_\rho(P)$	$LM^t_\rho(DP)$	LM^t_ρ
0.010	0.042	0.042	0.047	0.056	0.056	0.053
0.020	0.046	0.046	0.049	0.058	0.058	0.062
0.030	0.050	0.050	0.055	0.053	0.053	0.060
0.040	0.048	0.048	0.052	0.049	0.049	0.063
0.050	0.049	0.049	0.052	0.053	0.053	0.076
0.100	0.045	0.045	0.048	0.067	0.067	0.170
0.200	0.045	0.045	0.064	0.075	0.075	0.478
0.300	0.038	0.038	0.120	0.100	0.100	0.758

$(n, T) = (49, 10)$, mixture normal distribution.

Table 9: Empirical sizes when: $\gamma \neq 0$, $\lambda \neq 0$ I

		Normal Distribution			T Distribution		
γ_0	λ_0	$LM^t_\rho(P)$	$LM^t_\rho(DP)$	LM^t_ρ	$LM^t_\rho(P)$	$LM^t_\rho(DP)$	LM^t_ρ
0.010	0.010	0.054	0.054	0.060	0.056	0.056	0.055
0.030	0.010	0.050	0.050	0.058	0.042	0.042	0.061
0.050	0.010	0.043	0.043	0.065	0.044	0.044	0.068
0.100	0.010	0.035	0.035	0.077	0.033	0.033	0.086
0.200	0.010	0.015	0.015	0.148	0.014	0.014	0.145
0.300	0.010	0.003	0.003	0.224	0.005	0.005	0.216
0.010	0.030	0.053	0.053	0.063	0.049	0.049	0.059
0.030	0.030	0.046	0.046	0.073	0.049	0.049	0.064
0.050	0.030	0.045	0.045	0.080	0.045	0.045	0.075
0.100	0.030	0.036	0.036	0.086	0.030	0.030	0.086
0.200	0.030	0.011	0.011	0.145	0.014	0.014	0.137
0.300	0.030	0.003	0.003	0.210	0.004	0.004	0.202
0.010	0.050	0.057	0.057	0.080	0.052	0.052	0.087
0.030	0.050	0.050	0.050	0.093	0.048	0.048	0.091
0.050	0.050	0.047	0.047	0.099	0.049	0.049	0.096
0.100	0.050	0.037	0.037	0.125	0.034	0.034	0.120
0.200	0.050	0.014	0.014	0.186	0.017	0.017	0.175
0.300	0.050	0.003	0.003	0.252	0.004	0.004	0.247
0.010	0.100	0.062	0.062	0.180	0.059	0.059	0.173
0.030	0.100	0.053	0.053	0.188	0.053	0.053	0.198
0.050	0.100	0.053	0.053	0.217	0.052	0.052	0.233
0.100	0.100	0.042	0.042	0.288	0.037	0.037	0.292
0.200	0.100	0.019	0.019	0.414	0.017	0.017	0.401
0.300	0.100	0.006	0.006	0.486	0.005	0.005	0.490
0.010	0.200	0.081	0.081	0.506	0.075	0.075	0.504
0.030	0.200	0.060	0.060	0.576	0.061	0.061	0.561
0.050	0.200	0.058	0.058	0.634	0.057	0.057	0.630
0.100	0.200	0.040	0.040	0.738	0.045	0.045	0.742
0.200	0.200	0.030	0.030	0.885	0.030	0.030	0.882
0.300	0.200	0.021	0.021	0.923	0.020	0.020	0.926
0.010	0.300	0.089	0.089	0.791	0.092	0.092	0.789
0.030	0.300	0.077	0.077	0.849	0.083	0.083	0.850
0.050	0.300	0.064	0.064	0.890	0.062	0.062	0.893
0.100	0.300	0.042	0.042	0.958	0.047	0.047	0.954
0.200	0.300	0.045	0.045	0.992	0.046	0.046	0.993
0.300	0.300	0.058	0.058	0.997	0.056	0.056	0.997

$(n, T) = (49, 10)$.

Table 10: Empirical sizes when: $\gamma \neq 0$, $\lambda \neq 0$ II

		Chi-square Distribution			Mixture Normal Distribution		
γ_0	λ_0	$LM_\rho^t(P)$	$LM_\rho^t(DP)$	LM_ρ^t	$LM_\rho^t(P)$	$LM_\rho^t(DP)$	LM_ρ^t
0.010	0.010	0.045	0.045	0.050	0.052	0.052	0.055
0.030	0.010	0.047	0.047	0.060	0.044	0.044	0.060
0.050	0.010	0.044	0.044	0.069	0.047	0.047	0.064
0.100	0.010	0.034	0.034	0.082	0.033	0.033	0.076
0.200	0.010	0.013	0.013	0.142	0.016	0.016	0.148
0.300	0.010	0.004	0.004	0.218	0.003	0.003	0.219
0.010	0.030	0.048	0.047	0.065	0.049	0.049	0.061
0.030	0.030	0.052	0.052	0.069	0.048	0.048	0.073
0.050	0.030	0.046	0.046	0.077	0.049	0.049	0.079
0.100	0.030	0.034	0.034	0.094	0.035	0.035	0.090
0.200	0.030	0.014	0.014	0.142	0.016	0.016	0.143
0.300	0.030	0.004	0.004	0.211	0.003	0.003	0.216
0.010	0.050	0.050	0.050	0.082	0.056	0.056	0.087
0.030	0.050	0.051	0.051	0.093	0.047	0.047	0.090
0.050	0.050	0.046	0.046	0.101	0.046	0.046	0.096
0.100	0.050	0.038	0.038	0.120	0.036	0.036	0.124
0.200	0.050	0.016	0.016	0.182	0.015	0.015	0.186
0.300	0.050	0.004	0.004	0.246	0.005	0.005	0.245
0.010	0.100	0.055	0.055	0.175	0.054	0.054	0.172
0.030	0.100	0.053	0.053	0.196	0.051	0.051	0.194
0.050	0.100	0.044	0.044	0.221	0.051	0.051	0.230
0.100	0.100	0.037	0.037	0.287	0.041	0.041	0.286
0.200	0.100	0.018	0.018	0.398	0.017	0.017	0.403
0.300	0.100	0.007	0.007	0.487	0.005	0.005	0.502
0.010	0.200	0.068	0.068	0.523	0.070	0.070	0.513
0.030	0.200	0.061	0.061	0.578	0.066	0.066	0.557
0.050	0.200	0.058	0.058	0.621	0.051	0.051	0.631
0.100	0.200	0.044	0.044	0.749	0.038	0.038	0.756
0.200	0.200	0.030	0.030	0.882	0.030	0.030	0.880
0.300	0.200	0.022	0.022	0.921	0.021	0.021	0.930
0.010	0.300	0.074	0.074	0.812	0.095	0.095	0.800
0.030	0.300	0.072	0.072	0.856	0.075	0.075	0.853
0.050	0.300	0.056	0.056	0.904	0.070	0.070	0.896
0.100	0.300	0.043	0.043	0.961	0.051	0.051	0.959
0.200	0.300	0.040	0.040	0.991	0.046	0.046	0.992
0.300	0.300	0.062	0.062	0.997	0.065	0.065	0.996

$(n, T) = (49, 10)$.

Table 11: power of tests when: $\gamma \neq 0$ I

γ_0	$LM_\gamma^t(P)$	$LM_\gamma^t(DP)$	LM_γ^t	LM_κ^t	$LM_\kappa^t(D)$
0.010	0.082	0.082	0.089	0.325	0.318
0.020	0.151	0.151	0.160	0.413	0.407
0.030	0.284	0.284	0.296	0.515	0.509
0.040	0.444	0.444	0.462	0.674	0.666
0.050	0.623	0.623	0.643	0.800	0.795
0.100	0.995	0.995	0.995	0.998	0.998

(n,T)=(49,10), normal distribution.

Table 12: power of tests when: $\gamma \neq 0$ II

γ_0	$LM_\gamma^t(P)$	$LM_\gamma^t(DP)$	LM_γ^t	LM_κ^t	$LM_\kappa^t(D)$
0.010	0.076	0.076	0.079	0.321	0.314
0.020	0.158	0.158	0.166	0.412	0.405
0.030	0.288	0.288	0.303	0.525	0.522
0.040	0.437	0.437	0.454	0.661	0.654
0.050	0.615	0.615	0.632	0.792	0.787
0.100	0.995	0.995	0.996	0.998	0.998

(n,T)=(49,10), t distribution.

Table 13: power of tests when: $\gamma \neq 0$ III

γ_0	$LM_\gamma^t(P)$	$LM_\gamma^t(DP)$	LM_γ^t	LM_κ^t	$LM_\kappa^t(D)$
0.010	0.077	0.077	0.084	0.316	0.309
0.020	0.156	0.156	0.163	0.406	0.401
0.030	0.290	0.290	0.299	0.527	0.522
0.040	0.438	0.438	0.458	0.663	0.658
0.050	0.631	0.631	0.645	0.803	0.798
0.100	0.997	0.997	0.998	0.999	0.999

(n,T)=(49,10), chi-square distribution.

Table 14: power of tests when: $\gamma \neq 0$ IV

γ_0	$LM^t_\gamma(P)$	$LM^t_\gamma(DP)$	LM^t_γ	LM^t_κ	$LM^t_\kappa(D)$
0.010	0.075	0.075	0.081	0.318	0.311
0.020	0.158	0.158	0.164	0.404	0.397
0.030	0.279	0.279	0.293	0.527	0.520
0.040	0.445	0.445	0.462	0.677	0.672
0.050	0.613	0.613	0.634	0.788	0.783
0.100	0.993	0.993	0.994	0.998	0.998

(n,T)=(49,10), mixture normal distribution.

Table 15: power of tests when: $\lambda \neq 0$ I

λ_0	$LM^t_\lambda(P)$	$LM^t_\lambda(DP)$	LM^t_λ	LM^t_κ	$LM^t_\kappa(D)$
0.010	0.082	0.082	0.089	0.325	0.318
0.020	0.151	0.151	0.160	0.413	0.407
0.030	0.284	0.284	0.296	0.515	0.509
0.040	0.444	0.444	0.462	0.674	0.666
0.050	0.623	0.623	0.643	0.800	0.795
0.100	0.995	0.995	0.995	0.998	0.998

(n,T)=(49,10), normal distribution.

Table 16: power of tests when: $\lambda \neq 0$ II

λ_0	$LM^t_\lambda(P)$	$LM^t_\lambda(DP)$	LM^t_λ	LM^t_κ	$LM^t_\kappa(D)$
0.010	0.076	0.076	0.079	0.321	0.314
0.020	0.158	0.158	0.166	0.412	0.405
0.030	0.288	0.288	0.303	0.525	0.522
0.040	0.437	0.437	0.454	0.661	0.654
0.050	0.615	0.615	0.632	0.792	0.787
0.100	0.995	0.995	0.996	0.998	0.998

(n,T)=(49,10), t distribution.

Table 17: power of tests when: $\lambda \neq 0$ III

λ_0	$LM_\lambda^t(P)$	$LM_\lambda^t(DP)$	LM_λ^t	LM_κ^t	$LM_\kappa^t(D)$
0.010	0.077	0.077	0.084	0.316	0.309
0.020	0.156	0.156	0.163	0.406	0.401
0.030	0.290	0.290	0.299	0.527	0.522
0.040	0.438	0.438	0.458	0.663	0.658
0.050	0.631	0.631	0.645	0.803	0.798
0.100	0.997	0.997	0.998	0.999	0.999

(n,T)=(49,10), chi-square distribution.

Table 18: power of tests when: $\lambda \neq 0$ IV

λ_0	$LM_\lambda^t(P)$	$LM_\lambda^t(DP)$	LM_λ^t	LM_κ^t	$LM_\kappa^t(D)$
0.010	0.075	0.075	0.081	0.318	0.311
0.020	0.158	0.158	0.164	0.404	0.397
0.030	0.279	0.279	0.293	0.527	0.520
0.040	0.445	0.445	0.462	0.677	0.672
0.050	0.613	0.613	0.634	0.788	0.783
0.100	0.993	0.993	0.994	0.998	0.998

(n,T)=(49,10), mixture normal distribution.

Table 19: power of tests when: $\gamma \neq 0$, $\lambda \neq 0$ with normal errors I

γ_0	λ_0	$LM^t_\lambda(P)$	$LM^t_\lambda(DP)$	LM^t_λ	$LM^t_\gamma(P)$	$LM^t_\gamma(DP)$	LM^t_γ	LM^t_κ	$LM^t_\kappa(D)$
0.010	0.010	0.059	0.059	0.060	0.076	0.076	0.081	0.329	0.322
0.020	0.010	0.057	0.057	0.064	0.143	0.143	0.153	0.408	0.401
0.030	0.010	0.052	0.052	0.054	0.279	0.279	0.294	0.530	0.525
0.040	0.010	0.054	0.054	0.063	0.446	0.446	0.465	0.677	0.671
0.050	0.010	0.050	0.050	0.065	0.615	0.615	0.630	0.796	0.791
0.100	0.010	0.046	0.046	0.070	0.995	0.995	0.997	0.998	0.998
0.200	0.010	0.031	0.031	0.086	1.000	1.000	1.000	1.000	1.000
0.010	0.020	0.071	0.071	0.078	0.085	0.085	0.092	0.362	0.357
0.020	0.020	0.078	0.078	0.077	0.154	0.154	0.164	0.452	0.444
0.030	0.020	0.067	0.067	0.081	0.274	0.274	0.288	0.557	0.550
0.040	0.020	0.068	0.068	0.078	0.435	0.435	0.457	0.687	0.684
0.050	0.020	0.071	0.071	0.077	0.618	0.618	0.635	0.801	0.796
0.100	0.020	0.064	0.064	0.094	0.994	0.994	0.995	0.998	0.998
0.200	0.020	0.051	0.051	0.119	1.000	1.000	1.000	1.000	1.000
0.010	0.030	0.098	0.098	0.104	0.074	0.074	0.084	0.387	0.382
0.020	0.030	0.095	0.095	0.103	0.153	0.153	0.164	0.469	0.463
0.030	0.030	0.098	0.098	0.113	0.286	0.286	0.300	0.593	0.587
0.040	0.030	0.091	0.091	0.110	0.443	0.443	0.460	0.707	0.701
0.050	0.030	0.092	0.092	0.119	0.616	0.616	0.634	0.821	0.817
0.100	0.030	0.086	0.086	0.127	0.995	0.995	0.996	0.998	0.998
0.200	0.030	0.078	0.078	0.154	1.000	1.000	1.000	1.000	1.000
0.010	0.040	0.141	0.141	0.158	0.084	0.084	0.092	0.444	0.438
0.020	0.040	0.137	0.137	0.161	0.153	0.153	0.164	0.501	0.497
0.030	0.040	0.139	0.139	0.165	0.279	0.279	0.292	0.632	0.628
0.040	0.040	0.141	0.141	0.159	0.439	0.439	0.459	0.740	0.735
0.050	0.040	0.144	0.144	0.178	0.626	0.626	0.645	0.849	0.844
0.100	0.040	0.126	0.126	0.178	0.995	0.995	0.996	1.000	1.000
0.200	0.040	0.122	0.122	0.212	1.000	1.000	1.000	1.000	1.000
0.010	0.050	0.188	0.188	0.211	0.077	0.077	0.087	0.496	0.492
0.020	0.050	0.191	0.191	0.224	0.158	0.158	0.173	0.564	0.559
0.030	0.050	0.187	0.187	0.218	0.279	0.279	0.293	0.653	0.649
0.040	0.050	0.194	0.194	0.233	0.455	0.455	0.475	0.777	0.774
0.050	0.050	0.189	0.189	0.234	0.608	0.608	0.626	0.855	0.852
0.100	0.050	0.182	0.182	0.257	0.995	0.995	0.996	0.999	0.999
0.200	0.050	0.159	0.159	0.276	1.000	1.000	1.000	1.000	1.000

$(n, T) = (49, 10)$, normal distribution.

Table 20: power of tests when: $\gamma \neq 0$, $\lambda \neq 0$ with normal errors II

γ_0	λ_0	$LM^t_\lambda(P)$	$LM^t_\lambda(DP)$	LM^t_λ	$LM^t_\gamma(P)$	$LM^t_\gamma(DP)$	LM^t_γ	LM^t_κ	$LM^t_\kappa(D)$
0.010	0.100	0.551	0.551	0.622	0.082	0.082	0.097	0.797	0.796
0.020	0.100	0.536	0.536	0.617	0.149	0.149	0.170	0.834	0.830
0.030	0.100	0.550	0.550	0.627	0.275	0.275	0.304	0.877	0.874
0.040	0.100	0.540	0.540	0.627	0.446	0.446	0.483	0.912	0.911
0.050	0.100	0.554	0.554	0.651	0.625	0.625	0.665	0.953	0.951
0.100	0.100	0.568	0.568	0.678	0.994	0.994	0.995	0.999	0.999
0.200	0.100	0.578	0.578	0.697	1.000	1.000	1.000	1.000	1.000
0.010	0.200	0.981	0.981	0.990	0.075	0.075	0.121	0.996	0.996
0.020	0.200	0.986	0.986	0.994	0.159	0.159	0.231	0.998	0.998
0.030	0.200	0.986	0.986	0.994	0.291	0.291	0.384	0.999	0.999
0.040	0.200	0.985	0.985	0.993	0.448	0.448	0.546	0.999	0.999
0.050	0.200	0.986	0.986	0.994	0.619	0.619	0.704	0.999	0.999
0.100	0.200	0.991	0.991	0.995	0.995	0.995	0.998	1.000	1.000
0.200	0.200	0.994	0.994	0.996	1.000	1.000	1.000	1.000	1.000

$(n, T) = (49, 10)$, normal distribution.

Table 21: power of tests when: $\gamma \neq 0$, $\lambda \neq 0$ with mixture normal errors I

γ_0	λ_0	$LM^t_\lambda(P)$	$LM^t_\lambda(DP)$	LM^t_λ	$LM^t_\gamma(P)$	$LM^t_\gamma(DP)$	LM^t_γ	LM^t_κ	$LM^t_\kappa(D)$
0.010	0.010	0.056	0.056	0.053	0.082	0.082	0.087	0.320	0.314
0.020	0.010	0.050	0.050	0.051	0.149	0.149	0.160	0.407	0.401
0.030	0.010	0.055	0.055	0.062	0.276	0.276	0.291	0.542	0.538
0.040	0.010	0.054	0.053	0.057	0.449	0.449	0.464	0.680	0.674
0.050	0.010	0.049	0.049	0.059	0.620	0.620	0.642	0.804	0.800
0.100	0.010	0.045	0.044	0.071	0.995	0.995	0.995	0.999	0.999
0.200	0.010	0.036	0.036	0.087	1.000	1.000	1.000	1.000	1.000
0.010	0.020	0.067	0.067	0.075	0.081	0.081	0.086	0.352	0.344
0.020	0.020	0.070	0.070	0.081	0.157	0.157	0.167	0.442	0.436
0.030	0.020	0.067	0.066	0.072	0.281	0.281	0.294	0.551	0.545
0.040	0.020	0.073	0.072	0.089	0.445	0.445	0.463	0.694	0.688
0.050	0.020	0.068	0.068	0.082	0.635	0.635	0.652	0.812	0.807
0.100	0.020	0.056	0.056	0.091	0.992	0.992	0.993	0.998	0.998
0.200	0.020	0.048	0.048	0.118	1.000	1.000	1.000	1.000	1.000

$(n, T) = (49, 10)$, mixture normal distribution.

Table 22: power of tests when: $\gamma \neq 0$, $\lambda \neq 0$ with mixture normal errors II

γ_0	λ_0	$LM^t_\lambda(P)$	$LM^t_\lambda(DP)$	LM^t_λ	$LM^t_\gamma(P)$	$LM^t_\gamma(DP)$	LM^t_γ	LM^t_κ	$LM^t_\kappa(D)$
0.010	0.030	0.093	0.093	0.101	0.082	0.082	0.089	0.379	0.374
0.020	0.030	0.101	0.101	0.111	0.155	0.155	0.169	0.476	0.472
0.030	0.030	0.100	0.100	0.117	0.294	0.294	0.305	0.590	0.585
0.040	0.030	0.100	0.100	0.117	0.446	0.446	0.464	0.708	0.703
0.050	0.030	0.091	0.091	0.115	0.631	0.631	0.646	0.830	0.825
0.100	0.030	0.087	0.087	0.131	0.993	0.993	0.994	0.998	0.998
0.200	0.030	0.079	0.079	0.160	1.000	1.000	1.000	1.000	1.000
0.010	0.040	0.139	0.139	0.156	0.079	0.079	0.085	0.437	0.433
0.020	0.040	0.131	0.131	0.152	0.155	0.155	0.168	0.516	0.509
0.030	0.040	0.141	0.141	0.167	0.277	0.277	0.292	0.617	0.611
0.040	0.040	0.127	0.127	0.160	0.457	0.457	0.478	0.750	0.746
0.050	0.040	0.133	0.133	0.176	0.636	0.636	0.652	0.838	0.835
0.100	0.040	0.127	0.127	0.181	0.993	0.993	0.994	0.998	0.998
0.200	0.040	0.116	0.116	0.215	1.000	1.000	1.000	1.000	1.000
0.010	0.050	0.188	0.187	0.211	0.082	0.082	0.090	0.492	0.488
0.020	0.050	0.185	0.185	0.221	0.159	0.159	0.170	0.568	0.563
0.030	0.050	0.176	0.176	0.212	0.288	0.288	0.304	0.670	0.663
0.040	0.050	0.180	0.180	0.224	0.440	0.440	0.463	0.759	0.754
0.050	0.050	0.174	0.173	0.222	0.626	0.626	0.641	0.857	0.855
0.100	0.050	0.174	0.173	0.240	0.994	0.994	0.995	0.999	0.999
0.200	0.050	0.174	0.173	0.279	1.000	1.000	1.000	1.000	1.000
0.010	0.100	0.568	0.567	0.628	0.076	0.076	0.092	0.807	0.805
0.020	0.100	0.552	0.551	0.629	0.153	0.153	0.175	0.834	0.832
0.030	0.100	0.557	0.557	0.653	0.287	0.287	0.319	0.883	0.881
0.040	0.100	0.571	0.571	0.648	0.451	0.451	0.490	0.920	0.918
0.050	0.100	0.566	0.566	0.651	0.620	0.620	0.655	0.952	0.951
0.100	0.100	0.564	0.564	0.671	0.994	0.994	0.996	0.999	0.999
0.200	0.100	0.591	0.591	0.708	1.000	1.000	1.000	1.000	1.000
0.010	0.200	0.982	0.982	0.991	0.081	0.081	0.124	0.998	0.998
0.020	0.200	0.984	0.984	0.991	0.172	0.172	0.246	0.998	0.998
0.030	0.200	0.983	0.983	0.990	0.298	0.298	0.395	0.997	0.997
0.040	0.200	0.985	0.985	0.993	0.445	0.445	0.551	0.999	0.999
0.050	0.200	0.989	0.989	0.993	0.634	0.634	0.721	1.000	1.000
0.100	0.200	0.990	0.990	0.995	0.997	0.997	0.998	1.000	1.000
0.200	0.200	0.993	0.993	0.995	1.000	1.000	1.000	1.000	1.000

$(n, T) = (49, 10)$, mixture normal distribution.

References

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

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

Anselin, Luc and Rosina Moreno (2003). "Properties of tests for spatial error components". In: *Regional Science and Urban Economics* 33.5, pp. 595 –618.

Anselin, Luc and Serge Rey (1991). "Properties of Tests for Spatial Dependence in Linear Regression Models". In: *Geographical Analysis* 23.2, pp. 112–131.

Anselin, Luc et al. (1996). "Simple diagnostic tests for spatial dependence". In: *Regional Science and Urban Economics* 26.1, pp. 77 –104.

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

Baltagi, Badi H. and Zhenlin Yang (2013). "Standardized LM tests for spatial error dependence in linear or panel regressions". In: *The Econometrics Journal* 16.1, pp. 103–134.

Baltagi, Badi H., Seuck Heun Song, and Won Koh (2003). "Testing panel data regression models with spatial error correlation". In: *Journal of Econometrics* 117.1, pp. 123 –150.

Baltagi, Badi H. et al. (2007). "Testing for serial correlation, spatial autocorrelation and random effects using panel data". In: *Journal of Econometrics* 140.1, pp. 5 –51.

Bera, Anil K and Yannis Bilias (2001). "Rao's score, Neyman's C (α) and Silvey's LM tests: an essay on historical developments and some new results". In: *Journal of Statistical Planning and Inference* 97.1, pp. 9–44.

Bera, Anil K. and Mann J. Yoon (1993). "Specification Testing with Locally Misspecified Alternatives". In: *Econometric Theory* 9.4.

Bera, Anil K et al. (2017). "Adjustments of Rao's score test for distributional and local parametric misspecifications". In: *Available at SSRN 3074781.*

Bera, Anil K., Osman Doğan, and Süleyman Taşpınar (2018a). "Simple tests for endogeneity of spatial weights matrices". In: *Regional Science and Urban Economics* 69, pp. 130 –142.

— (2018b). "Testing Spatial Dependence in Spatial Models with Endogenous Weights Matrices". In: *Journal of Econometric Methods.* (forthcoming).

Bera, Anil K et al. (2019). "Robust LM tests for spatial dynamic panel data models". In: *Regional Science and Urban Economics* 76, pp. 47–66.

Born, Benjamin and Jorg Breitung (2011). "Simple regression-based tests for spatial dependence". In: *The Econometrics Journal* 14.2, pp. 330–342.

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

Doğan, Osman, Süleyman Taşpınar, and Anil K Bera (2018). "Simple tests for social interaction models with network structures". In: *Spatial Economic Analysis* 13.2, pp. 212–246.

Elhorst, J Paul (2003). "Specification and estimation of spatial panel data models". In: *International regional science review* 26.3, pp. 244–268.

Fang, Ying, Sung Y Park, and Jinfeng Zhang (2014). “A simple spatial dependence test robust to local and distributional misspecifications”. In: *Economics Letters* 124.2, pp. 203–206.

Kapoor, Mudit, Harry H. Kelejian, and Ingmar R. Prucha (2007). “Panel data models with spatially correlated error components”. In: *Journal of Econometrics* 140.1, pp. 97 –130.

Kelejian, Harry H. and Ingmar R. Prucha (2001). “On the Asymptotic Distribution of the Moran I Test Statistic with Applications”. In: *Journal of Econometrics* 104.2, pp. 219–257.

— (2010). “Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances”. In: *Journal of Econometrics* 157, pp. 53–67.

Kelejian, Harry H. and Dennis P. Robinson (1992). “Spatial autocorrelation: A new computationally simple test with an application to per-capita county police expenditures”. In: *Regional Science and Urban Economics* 22.3. Special Issue Space and Applied Econometrics, pp. 317 –331.

Korniotis, George M. (2010). “Estimating Panel Models With Internal and External Habit Formation”. In: *Journal of Business & Economic Statistics* 28.1, pp. 145–158.

Kullback, Solomon and Richard A Leibler (1951). “On information and sufficiency”. In: *The annals of mathematical statistics* 22.1, pp. 79–86.

Lee, Lung-fei and Jihai Yu (Apr. 2010). “A Spatial Dynamic Panel Data Model with both Time and Individual Fixed Effects”. In: *Econometric Theory* 26 (02), pp. 564–597. ISSN: 1469-4360.

— (2012). “QML Estimation of Spatial Dynamic Panel Data Models with Time Varying Spatial Weights Matrices”. In: *Spatial Economic Analysis* 7.1, pp. 31–74.

— (2014). “Efficient GMM estimation of spatial dynamic panel data models with fixed effects”. In: *Journal of Econometrics* 180.2.

Mahalanobis, Prasanta Chandra (1944). “On large-scale sample surveys”. In: *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 231.584, pp. 329–451.

Moran, Patrick AP (1950). “Notes on continuous stochastic phenomena”. In: *Biometrika* 37.1/2, pp. 17–23.

Munnell, Alicia H. (1990). “Why has productivity growth declined? Productivity and public investment”. In: *New England Economic Review* Jan, pp. 3–22.

Neyman, Jerzy (1979). “C (α) tests and their use”. In:

Parent, Olivier and P. James LeSage (2010). “A spatial dynamic panel model with random effects applied to commuting times”. In: *Transportation Research Part B: Methodological* 44.5, pp. 633 –645.

— (2011). “A space-time filter for panel data models containing random effects”. In: *Computational Statistics & Data Analysis* 55.1, pp. 475 –490.

— (2012). “Spatial dynamic panel data models with random effects”. In: *Regional Science and Urban Economics* 42.4, pp. 727 –738.

Pinkse, Joris (2004). “Moran-Flavored Tests with Nuisance Parameters: Examples”. In: *Advances in Spatial Econometrics: Methodology, Tools and Applications*. Ed. by Luc Anselin, Raymond J. G. M. Florax, and Sergio J. Rey. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 67–77.

Rahmatullah Imon, AHM (2003). “Regression residuals, moments, and their use in tests for normality”. In: *Communications in Statistics-Theory and Methods* 32.5, pp. 1021–1034.

Rao, C Radhakrishna (1948). “Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation”. In: *Mathematical Proceedings of the Cambridge Philosophical Society*. Vol. 44. 1. Cambridge University Press, pp. 50–57.

Robinson, P.M. (2008). “Correlation testing in time series, spatial and cross-sectional data”. In: *Journal of Econometrics* 147.1, pp. 5 –16.

Su, Liangjun and Zhenlin Yang (2015). “QML estimation of dynamic panel data models with spatial errors”. In: *Journal of Econometrics* 185.1, pp. 230 –258.

Taşpınar, Süleyman, Osman Doğan, and Anil K Bera (2017). “GMM Gradient Tests for Spatial Dynamic Panel Data Models”. In: *Regional Science and Urban Economics* 65, pp. 65–88.

Taşpınar, Süleyman, Osman Doğan, and Wim P. M. Vijverberg (2018). “GMM inference in spatial autoregressive models”. In: *Econometric Reviews* 37.9, pp. 931–954.

White, Halbert (1982). “Maximum likelihood estimation of misspecified models”. In: *Econometrica: Journal of the Econometric Society*, pp. 1–25.

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

— (2016). “Joint Tests for Dynamic and Spatial Effects in Short Panel Data Models with Fixed Effects”. Working paper, Singapore Management University.

— (2018). “Unified M-estimation of fixed-effects spatial dynamic models with short panels”. In: *Journal of Econometrics* 205.2, pp. 423 –447.

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