***Accepted in Econometric Reviews, Forthcoming.***

# Testing the Endogeneity of a Spatial Weight Matrix in the Weak-Tied Spatial Dynamic Panel Data Model

Jieun Lee*

November 10, 2024

**Abstract**

*Weak ties* in large-scale networks facilitate the diffusion of information and resources, generating spillover effects. While geographic proximity may generate straightforward networks, a non-predetermined spillover framework better accommodates socioeconomic interactions. In such cases, it is important to test the endogeneity of a spatial weight matrix before full estimation to ensure valid and efficient model selection. I propose a robust score test to assess this endogeneity in the weak-tied spatial dynamic panel data model, emphasizing its asymptotic optimality and computational efficiency. First, I correct score function biases to address the incidental parameters problem. Second, I adjust the score functions to enhance robustness against local parametric misspecifications. Monte Carlo simulations demonstrate favorable finite sample properties. In my application, I illustrate how this test helps model selection for the Solow-Swan growth model, considering spillovers from geographic and knowledge proximities. The estimation results reveal endogenous spillover effects with weakly positive impacts on real income per worker, while savings show positive internal effects but negative spillover effects possibly due to competition for investment. Notably, the labor growth rate positively influences knowledge accumulation and innovation, leading to beneficial effects on real income per worker through both internal and spillover channels.

JEL codes: C01, C12, C31, C33, C55
Keywords: Spatial dynamic panel data model, weak ties, endogenous spillover effects, score test (or Lagrange multiplier test), model selection.

*First author, Corresponding author. Hubert Department of Global Health, Rollins School of Public Health, Emory University. *jieun.lee@emory.edu.*

# 1 Introduction

*Weak ties* in large-scale networks, characterized by infrequent interaction or low intensity, are renowned for their indispensable role in individuals' opportunities and community integration (Granovetter 1973; Brown and Konrad 2001; Patacchini and Zenou 2008). They facilitate the diffusion of information and resources across diverse social circles, enabling individuals to access novel information and job opportunities. Subsequently, *spillover effects* arise as a complex interplay of such interactions and feedback mechanisms inherent in networked systems.

Spillover effects have been of interest across various fields in social sciences. In economics, for example, the economic interdependence of nations is generally presupposed and discussed in economic growth literature (Lucas 1990; De Long and Summers 1991; Moreno and Trehan 1997) and the knowledge spillover[1] has been actively studied in the endogenous growth literature (Romer 1990; Acs, Braunerhjelm, Audretsch, and Carlsson 2009; Aghion, Bergeaud, Gigout, and Melitz 2019). Also, peer- or network effects have been identified as significant factors in social interaction models, with wide applications in areas such as education, health behavior, the labor market, criminal activities, and consumption patterns (Gaviria and Raphael 2001; Falk and Ichino 2006; Fowler and Christakis 2008; Ali and Dwyer 2009; Lin 2010, 2014; Moretti 2011; Zimmerman and Messner 2011; Card and Giuliano 2013; Angrist 2014; Bursztyn, Ederer, Ferman, and Yuchtman 2014; Cornelissen, Dustmann, and Schonberg 2017).

Notably, a spatial dynamic panel data (SDPD) model offers an advantageous extension of the dynamic panel data model by incorporating spatial dependencies (Lee and Yu 2012; Qu, Lee, and Yu 2017). It finds applications in various topics, including the endogenous growth model (Bonnal 2010; Ho, Wang, and Yu 2013; Amidi, Fagheh, and Javaheri 2020), habit formation model (Brida and Risso 2009; Seetaram and Dwyer 2009; Korniotis 2010), and the impacts of climate change and weather (Tokunaga, Okiyama, and Ikegawa 2017; Bille and Rogna 2022).

Meanwhile, an intuitive factor that naturally fosters interconnectedness and encourages spillovers is physical contiguity that represents adjacency with neighboring areas or distance measured by geographic metrics. In such networks, interactions are expected to decay at a specific rate as the distance between units increases. Consequently, a spatial weight matrix ($W$) has been constructed to represent the weighted ties or connections among spatial units, primarily based on predetermined or non-stochastic geography (Moran 1950; Cliff and Ord 1972; Anselin 1988, 2001; Ertur and Koch 2007; Kelejian and Prucha 2010; Elhorst 2010, 2014; Lee and Yu 2010, 2012).

[1] *Knowledge* is defined as a set of technological opportunities and is multiplied by knowledge sharing through economic interactions such as trade.

However, the literature emphasizes the pivotal role of *non-predetermined* distances, particularly economic ones, in facilitating interactions. In the literature on endogenous network formation, networks are depicted as outcomes of choices among cross-sectional units based on non-predetermined socioeconomic features, rather than being given exogenously (Hoff, Raftery, and Handcock 2002; Handcock, Raftery, and Tantrum 2007; Graham 2017; Jochmans 2018). In empirical studies, the economic distance dictated by transportation costs significantly influences trade flow magnitudes (Roemer 1977) and establishes interdependencies among countries' random variables (Conley and Ligon 2002). Moreover, an analysis of unemployment in Chicago from 1980 to 1990 reveals a pronounced spatial dependence concerning the economic distance of occupation proportions across tracts (Conley and Topa 2002) and agglomeration spillovers are observed to occur more prominently between firms sharing labor pools and similar technologies (Greenstone, Hornbeck, and Moretti 2010). Furthermore, current actions of forward-looking agents can impact their future network links through time-varying economic indicators, as exemplified by empirical evidence that a state's welfare expenditure is reinforced by its neighbors' decisions on welfare expenditure (Jeong and Lee 2021).

In such cases, spillover effects could be *endogenous*, subject to the dependency between stochastic features of the outcome and those of the non-predetermined triggers (Pinkse and Slade 2010). This is because economic variables, as an example of non-predetermined triggers, are expected to be broadly and closely connected, and thus the random shock that affects the outcome might be correlated with that of the stochastic triggers.[2] If $W$ is endogenous, this violates the assumption of exogenous $W$ in the conventional SDPD model (Lee and Yu 2012); thus, inferences drawn from this model are invalid.

For valid inference, the endogeneity of $W$ must be appropriately addressed. Accordingly, we consider the SDPD model with an endogenous spatial weight matrix (Qu, Lee, and Yu 2017). The endogeneity of $WY$ in the spatial autoregressive model can be resolved by several estimation methods, such as maximum likelihood (ML) estimation, instrumental variables estimation, or the generalized method of moments (Qu and Lee 2015; Qu, Lee, and Yu 2017; Qu, Lee, and Yang 2021). This paper primarily employs ML estimation to develop a score test equivalent to the $C(\alpha)$ test (Neyman 1959), achieving asymptotic optimality.

Specifically, a control function approach with a joint normal specification on the error terms controls the endogeneity of $W$ (Qu, Lee, and Yu 2017). Notably, this estimator remains consistent regardless of whether $W$ is exogenous or endogenous. However, it is parsimoniously recommended

[2]For example, consider GDP, which is affected by the aggregate supply effect through the international supply chain, and the aggregate demand effect such as consumption. When the COVID shock occurred, it hit both the demand and supply sides, and thus, the consumption and trade might be correlated due to the common COVID shock. If this is the case, the spillover triggered by trade is endogenous.

only when $W$ is endogenous due to its considerable computational complexity. This complexity arises because this model entails an auxiliary equation, with the number of additional parameters being proportional to the *square* of the dimensionality of the stochastic triggers ($Z$) for $W$. Furthermore, it involves the number of unique values in the variance-covariance matrix of the error terms in the vector autoregressive (VAR) process, which is expected to be substantial in high-dimensional data.[3] Given the growing necessity to use high-dimensional data in the era of big data, this may result in notably heightened computational costs, a factor that cannot be disregarded.

Therefore, researchers are strongly encouraged to determine *before* their full analysis whether the conventional SDPD model, which requires minimal computational costs for analyzing spatial dynamic panel data, is sufficient or if they should use the more robust SDPD model that accounts for an endogenous spatial weight matrix but may need additional computational resources. Table 1 summarizes a comprehensive guide to spatial models, organized according to the type of $W$. Given that the selection of a valid and efficient spatial model hinges on the endogeneity of $W$, conducting tests to assess its endogeneity is strongly encouraged.

**Table 1 is here**

Therefore, the development of a valid test to determine the endogeneity of $W$ in the SDPD model is important. Moreover, considering spatial models are known to be more computationally intensive than non-spatial models, it is crucial to minimize the computational cost of the model selection process to promote the use of spatial models. Otherwise, a computationally costly model selection process can lead to prolonged periods for selecting valid spatial models, followed by additional extended periods for model estimation. Thus, I propose a robust score test (or Lagrange multiplier test), which is asymptotically optimal and computationally efficient, to statistically determine the endogeneity of $W$ in the weak-tied SDPD model, where weak-tiedness is referred to as small spatial dependence.

While there are various options for testing the endogeneity of $W$, such as the Wald test developed directly from the estimation, the key advantage of my proposed test is that it is evaluated using estimates under the null hypothesis and from *non-spatial* models, leveraging the asymptotic interpretation of weak ties in large-scale networks. Notably, my proposed score test is even more computationally efficient than the conventional Rao's score test. The conventional Rao's score test

[3]Specifically, equation (3) shows that the number of additional parameters in Qu, Lee, and Yu (2017) is $p \times (p + k_2 + 1) + J$, where $p$ represents the dimension of $Z$ following a VAR process, $k_2$ is the dimension of explanatory variable for $Z$, and $J$ denotes the dimension of distinct elements in the variance-covariance matrix of $Z$.

framework evaluated under the null inherently offers computational efficiency, especially when $Z$ is high-dimensional as the alternative involves the dimensionality of $Z$. However, while the conventional score test requires evaluation using spatial models, my proposed score test achieves *further* computational efficiency by reducing spatial models to non-spatial ones. Considering spatial models pose challenges due to higher computational costs associated with including spatially autoregressive terms in $Y$ (i.e., $WY$), where these costs increase significantly with the number of cross-sectional units, my proposed test offers a valuable solution. Still, my proposed test provides valid inference and achieves asymptotic optimality due to its equivalence to the $C(\alpha)$ test, despite its simplicity.

Importantly, weak ties can be asymptotically interpreted by leveraging the properties of large-scale networks (i.e., large $n$). In such networks, weak ties are expected to form through *dilution effects*, which refer to the diminishing influence or impact of individual nodes or connections as the network size increases. This phenomenon occurs for several reasons: as the network grows, the average path length between nodes typically increases, making direct connections less significant; with more nodes, the degree (number of connections per node) can increase, leading to individual connections carrying less weight; larger networks often have more redundant paths, diluting the impact of any single path; and larger networks are usually more complex and noisy, which can obscure the influence of individual nodes or connections. Asymptotically speaking, we can interpret these effects as spatial dependencies drifting towards zero with increasing network size. In this context, local alternative frameworks (Neyman 1937) can be utilized to model the structure of spatial dependence parameters in the presence of local parametric misspecifications.

Accordingly, given a known form of alternatives, we can derive the analytical form of the score test under the joint null hypotheses, which pertain to testing parameters related to the endogeneity of $W$ as well as spatial dependence parameters. Consequently, the score test for the weak-tied SDPD model requires estimation of fewer parameters, similar to those in the standard dynamic panel data model with two-way fixed effects, thereby enhancing computational efficiency while still yielding asymptotically optimal inference as valid as the $C(\alpha)$ test (Neyman 1959).

To derive the valid score test in the weak-tied SDPD model, I make two major adjustments: (i) analytical bias correction to resolve the incidental parameters problem; and (ii) adjustment for local parametric misspecifications in spatial dependence parameters. First, I analytically correct the biases of the score functions so that they are centered around zero. This is because, unlike in the cross-section, the incidental parameters problem arises in panel data due to the two-way fixed effects (Neyman and Scott 1948).[4] Then the score functions are not centered around zero since the biases

[4]The incidental parameters such as individual- and time fixed effects weaken the ability of MLE to consistently estimate the structural parameter, which is in particular acute in the dynamic panel data model as the incidental parameters are involved with the structural parameter.

composed of the two-way fixed effects do not vanish even with large $n$ and $T$ under the assumption that $\frac{n}{T}$ converges to some finite positive real number (Yu, de Jong, and Lee 2008; Qu, Lee, and Yu 2017).[5] It should be emphasized that the bias correction in this paper quite differs from others since the bias corrections in Qu, Lee, and Yu (2017) or any forms of $C(\alpha)$ test are evaluated at the estimates in a SDPD model, while my testing is evaluated under the joint null and thus requires much simpler estimation of a standard two-way fixed effects model. Moreover, through the categorization of parameters into three types—testing parameters, spatial dependence parameters susceptible to local parametric misspecification, and all other pure nuisance parameters—, my bias correction method incorporates an additional statistical process, resulting in biases of distinct forms. This process primarily involves removing the influence of nuisance parameters on the other parameters of interest. This is achieved by conditioning the score functions of the parameters of interest to those of pure nuisance parameters, thereby revealing the true relationships among the parameters.

Second, I address potential local parametric misspecifications by applying the framework of Bera and Yoon (1993). To remove the effects of the presence of local misspecifications in spatial dependence parameters on the score test for the testing parameter, I orthogonalize the score functions of testing parameters and spatial dependence parameters so that their score functions become independent regardless of the presence of local misspecifications. After orthogolized, the distribution of the score function of the testing parameter is unaffected by the presence of local misspecifications in spatial dependence parameters and thus my testing attains valid inference in terms of size and power. Computational efficiency is also achieved because it is sufficient for my testing to estimate a non-spatial model for valid inference, while the $C(\alpha)$ test requires the estimation of a spatial model. Furthermore, asymptotic optimality is achieved because my testing is asymptotically equivalent to the $C(\alpha)$ test, which is asymptotically optimal.

In summary, my testing offers *robust* inferences regarding biases stemming from incidental parameters and local misspecifications in spatial dependence parameters. It consistently exhibits asymptotic central chi-squared distribution under the null hypothesis. Furthermore, Monte Carlo simulations demonstrate its favorable finite sample properties in terms of size and power. Notably, it achieves asymptotic optimality and computational efficiency, outperforming any $C(\alpha)$-type test, as it relies solely on estimating a standard two-way fixed effects model.

Literature on testing and model specifications has been mostly considered for cross-sectional spatial models (Anselin 1988, 2001; Kelejian and Robinson 1992; Baltagi and Li 2001; Yang 2010; Bera, Dogan, and Taspinar 2018a,b; Dogan, Taspinar, and Bera 2018), whereas only a few studies

[5]While there are three cases for the relative converging speed of $\frac{n}{T}$, either goes to some finite positive real number $k < \infty$, 0, or $\infty$, only the first case is parsimoniously employed since not only the last two cases are very uncommon in practice but also the score function degenerates so that it is impossible to develop the score test.

are found for the spatial static panel models (Baltagi, Song, and Koh 2003; Baltagi, Song, Jung, and Koh 2007; Baltagi and Liu 2008; Baltagi, Song, and Kwon 2009; Debarsy and Ertur 2010; Baltagi and Yang 2013) and the SDPD models (Yang, Yu, and Liu 2016; Taspinar, Dogan, and Bera 2017; Bera, Dogan, Taspinar, and Leiluo 2019). However, most of these studies target spatial dependence, and the endogeneity tests for $W$ have been developed primarily on the cross-sectional spatial models (Qu and Lee 2015; Cheng and Lee 2017; Bera, Dogan, and Taspinar 2018a).

Various hypothesis tests to determine the endogeneity of $W$ in the SDPD model can indeed be developed. For example, the Wald test, which utilizes estimates under both the null and alternative hypotheses from spatial models, can be directly derived from Qu, Lee, and Yu (2017). However, to the best of my knowledge, an asymptotically optimal endogeneity test for $W$ that uses only estimates under the null and from non-spatial models, thereby attaining a computational advantage, has not yet been developed. This paper is the first to propose such a test.

The rest of the paper is organized as follows: Section 2 provides a summary of the model specification outlined in Qu, Lee, and Yu (2017). In Section 3, I discuss two significant adjustments made to derive the valid score test for determining the endogeneity of $W$. Section 4 presents the results of a Monte Carlo simulation aimed at investigating the finite sample properties, size, and power characteristics of my testing method, particularly focusing on network size and time duration where it performs optimally in practice. In Section 5, a research episode is illustrated where my proposed test is applied in the early stages of research for valid and efficient spatial model selection, using data from the Penn World Table version 7.1 within the context of the SDPD neoclassical growth model. Finally, Section 6 concludes the paper.

# 2 Model Specification and Score Test for Endogeneity of $W$

Consider a random field of error terms $\{(\epsilon_{l(i,t)}, v_{l(i,t)}) : l(i,t) \in D_{nT},\ i \in \mathcal{N},\ t \in \mathcal{T}\}$ defined on a probability space $(\Omega, \mathcal{F}, P)$, where $\ell$ represents a location, $D_{nT} \subset D$ is a finite set and $D$ satisfies Assumption A.1 in Appendix A.1. To simplify the notation, let $(\epsilon_{l(i,t)}, v_{l(i,t)})$ be denoted by $(\epsilon_{it}, v_{it})$, where $\epsilon_{it}$ is the $p \times 1$ column vector for $t = 1, \ldots, T$ formulated from the $i^{th}$ row of $n \times p$ matrix $e_{nt}$, and $v_{it}$ is the $i^{th}$ element of $n \times 1$ vector $V_{nt}$, for all $t = 1, \cdots, T$.

**Assumption 1.** The error terms $v_{it}$ and $\epsilon_{it}$ are *i.i.d.* over $i$ and $t$ and $(v_{it}, \epsilon_{it}')' \overset{iid}{\sim} N(0, \Sigma_{v\epsilon 0})$, where $\Sigma_{v\epsilon 0} = \begin{pmatrix} \sigma_{v0}^2 & \sigma_{v\epsilon 0}' \\ \sigma_{v\epsilon 0} & \Sigma_{\epsilon 0} \end{pmatrix}$, $\sigma_{v0}^2$ is a scalar variance of $v_{it}$, $\Sigma_{\epsilon 0}$ is a $p \times p$ covariance matrix of $\epsilon_{it} = (\epsilon_{it,1}, \cdots, \epsilon_{it,p})'$, and $\sigma_{v\epsilon 0}$ is a $p \times 1$ covariance matrix between $v_{it}$ and $\epsilon_{it}$.

Consider a SDPD model with individual and time fixed effects (Qu, Lee, and Yu 2017)[6]

$$Y_{nt} = \lambda_0 W_{nt} Y_{nt} + R_{nt}\phi_{10} + c_{n10} + \tau_{t10} 1_n + V_{nt}, \tag{1}$$

for $t = 1, \ldots, T$, where $Y_{nt} = (y_{1t}, \ldots, y_{nt})'$ is an $n \times 1$ vector of observations on the dependent variable, $W_{nt} = \{(w_{ij})_{i,j=1,\ldots,n}\}_t$ is an $n \times n$ spatial weight matrix with zero diagonal elements, and $\lambda_0$ is the spatially autoregressive or contemporaneous spatial dependence parameter. $R_{nt} = (Y_{n,t-1}, W_{n,t-1}Y_{n,t-1}, X_{1nt})$ is a matrix of exogenous variables, consisting of a time dynamic dependent, spatial dynamic dependent, and an $n \times k_1$ matrix of nonstochastic explanatory variables. Accordingly, $\phi_{1,0}$ is a vector of $(\gamma_0, \rho_0, \beta_0')'$ where $\rho_0$ is the dynamic spatial dependence parameter, $\gamma_0$ is the dynamic dependence parameter, and $\beta_0$ is a $k_1 \times 1$ vector of parameters. $c_{n10}$ is an $n \times 1$ column vector of individual fixed effects and $\tau_{t10}$ is the $t^{th}$ element of $T \times 1$ time fixed effects vector $\tau_{T10}$, and $V_{nt} = (v_{1t}, \cdots, v_{nt})'$ is an $n \times 1$ vector of error terms with zero mean and variance $\sigma_0^2$.

Now suppose $W_{nt}$ is constructed as a function of $Z_{nt}$ such that

$$\begin{aligned} (W_{nt})_{ij} &= w_{ij,nt} = w_{ij}^d \cdot h(z_{i,nt}, z_{j,nt}), \\ Z_{nt} &= K_{nt}\Phi_2 + c_{n20} + 1_n \tau_{t20}' + e_{nt}, \end{aligned} \tag{2}$$

for $t = 1, \ldots, T$, where $w_{ij}^d$ represents the *predetermined* part in $(W_{nt})_{ij}$ (e.g., geographic proximity) and $h(z_{i,nt}, z_{j,nt})$ models the *non-predetermined* part in $(W_{nt})_{ij}$ (e.g., socioeconomic proximity), where $h(\cdot,\cdot)$ is a nonnegative and uniformly bounded function and $Z_{nt} = (z_{1t}, \cdots, z_{nt})'$ is an $n \times p$ matrix following vector autoregressive (VAR) process with a $p \times 1$ vector $z_{it} = (z_{it,1}, \cdots, z_{it,p})'$. $K_{nt} = (Z_{n,t-1}, X_{2nt})$ with $\Phi_2 = (\kappa_0', \Gamma_0')'$, where $Z_{n,t-1}$ is an $n \times p$ matrix with its associated $p \times p$ parameter matrix $\kappa_0$ and $X_{2nt}$ is an $n \times k_2$ matrix of deterministic explanatory variables with its associated $k_2 \times p$ parameter matrix $\Gamma_0$. $e_{nt} = (\epsilon_{1t}, \cdots, \epsilon_{nt})'$ is an $n \times p$ matrix of errors, where $\epsilon_{it} = (\epsilon_{it,1}, \cdots, \epsilon_{it,p})'$ is a $p \times 1$ vector. $c_{n20}$ is an $n \times 1$ column vector of individual fixed effects and $\tau_{t20}$ is the $t^{\text{th}}$ element of $T \times 1$ time fixed effects vector $\tau_{T20}$. $1_n$ is an $n \times 1$ vector of ones. The initial values $Y_{n0}$ and $Z_{n0}$ are assumed to be observable[7].

A control function approach may apply, using the joint normal specification such that $\mathrm{E}(v_{it}|\epsilon_{it}) = \epsilon_{it}'\delta_0$ where $\delta_0 = \Sigma_{\epsilon 0}^{-1}\sigma_{v\epsilon 0}$ and $\mathrm{Var}(v_{it}|\epsilon_{it}) = \sigma_{\xi 0}^2$ where $\sigma_{\xi 0}^2 = \sigma_{v0}^2 - \sigma_{v\epsilon 0}'\Sigma_{\epsilon 0}^{-1}\sigma_{v\epsilon 0}$. Let $\xi_{nt} := V_{nt} - e_{nt}\delta_0$. Since the expectation of $\xi_{nt}$ conditional on $e_{nt}$ is zero, $\xi_{nt}$ is uncorrelated with $e_{nt}$. Hence (1) and (2)

[6] Additional assumptions in Qu, Lee, and Yu (2017) on the spatial setting are provided in Appendix A.1.

[7] Caveats regarding the unknown initial conditions problem are discussed in Section 3.4.2.

can be combined as

$$\begin{aligned} Y_{nt} = {} & \lambda_0 W_{nt} Y_{nt} + R_{nt}\phi_{10} + c_{n10} + \tau_{t10} 1_n \\ & + (Z_{nt} - K_{nt}\Phi_2 - c_{n20} - 1_n \tau'_{t20})\delta_0 + \xi_{nt}, \end{aligned} \tag{3}$$

where $\xi_{nt}|e_{nt} \sim N(0, \sigma^2_{\xi 0} I_n)$ and $\xi_{nt}$ are *i.i.d.* across $i$ and $t$. This makes $\mathrm{Cov}(\xi_{nt}, e_{nt}) = 0$. Note that $(Z_{nt} - K_{nt}\Phi_2 - c_{n20} - 1_n \tau'_{t20})$ are the control variables that address the endogeneity of $W_{nt}$. In a matrix form,

$$\begin{aligned} \mathbf{Y}_{nT} = {} & \lambda_0 \mathbf{W}_{nT}\mathbf{Y}_{nT} + \mathbf{R}_{nT}\phi_{10} + \ell_0(\gamma_0, \rho_0) + \overline{c_{1nT,0}} + \overline{\tau_{1nT,0}} \\ & + (\mathbf{Z}_{nT} - \mathbf{K}_{nT}\Phi_2)\delta_0 + \Xi_{nT}, \end{aligned}$$

where $\mathbf{Y}_{nT} = (Y_{n1}, Y_{n2}, \ldots, Y_{nT})'$ and other variables with a subscript $nT$ are defined similarly[8]. $\ell_0 = (\gamma_0 Y_{n0} + \rho_0 W_{n0} Y_{n0}, 0, \ldots, 0)'$ and $\overline{c_{1nT,0}} = c_{1nT,0} - c_{2nT,0}\delta_0$; $\overline{\tau_{1nT,0}} = \tau_{1nT,0} - \tau_{2nT,0}\delta_0$.

The concentrated log-likelihood function is given as

$$\begin{aligned} \ln L^c_{nT}(\theta) = {} & -\frac{nT}{2}\ln 2\pi + \ln|\mathbf{S}_{nT}(\lambda,\gamma,\rho)| - \frac{nT}{2}\ln\sigma^2_\xi - \frac{nT}{2}\ln|\Sigma_\epsilon| \\ & - \frac{1}{2}\mathrm{vec}'(\mathbf{Z}_{nT} - \mathbf{K}_{nT}\Phi_2)(\Sigma_\epsilon^{-1} \otimes \mathbf{J}_{nT})\mathrm{vec}(\mathbf{Z}_{nT} - \mathbf{K}_{nT}\Phi_2) \\ & - \frac{1}{2\sigma^2_\xi}\big((\mathbf{I}_{nT} - \lambda_0\mathbf{W}_{nT})\mathbf{Y}_{nT} - \mathbf{R}_{nT}\phi_{10} - (\mathbf{Z}_{nT} - \mathbf{K}_{nT}\Phi_2)\delta - \ell_0(\gamma,\rho)\big)' \\ & \times \mathbf{J}_{nT} \times \big((\mathbf{I}_{nT} - \lambda_0\mathbf{W}_{nT})\mathbf{Y}_{nT} - \mathbf{R}_{nT}\phi_{10} - (\mathbf{Z}_{nT} - \mathbf{K}_{nT}\Phi_2)\delta - \ell_0(\gamma,\rho)\big), \end{aligned}$$

where $\theta = (\lambda, \phi'_1, \delta', \phi'_2, \sigma^2_\xi, \alpha')'$, $\phi_1 = (\gamma, \rho, \beta')'$, $\phi_2 = \mathrm{vec}(\Phi_2)$ with $\Phi_2 = (\kappa', \Gamma')'$, and $\alpha$ is the $J \times 1$ column vector of distinct elements in $\Sigma_\epsilon$. $J_n = I_n - \frac{1}{n}1_n 1'_n$ and $J_T = I_T - \frac{1}{T}1_T 1'_T$ are the two orthogonal projectors for taking time deviation and cross-sectional deviation from their means, respectively (Lee and Yu 2010, 2012), $\mathbf{J}_{nT} = J_T \otimes J_n$, and $\mathbf{S}_{nT}(\lambda,\gamma,\rho) = \mathbf{I}_{nT} - \lambda\mathbf{W}_{nT} - \gamma\mathbf{I}_{nT,-1} - \rho\mathbf{W}_{n,T-1}$ with $\mathbf{S}_{nT} = \mathbf{S}_{nT}(\lambda_0, \gamma_0, \rho_0)$.

[8]Explicit matrix forms are provided in Appendix A.2.

The first-derivative of the unbiased concentrated log-likelihood function is given as

$$\frac{\partial \ln L^c_{1,nT}(\theta_0)}{\partial\theta} = \begin{pmatrix} \frac{1}{\sigma^2_{\xi 0}}(\mathbf{W}_{nT}\mathbf{Y}_{nT})'\mathbf{J}_{nT}\Xi_{nT} - \mathrm{tr}(\mathbf{G}_{1,nT}\mathbf{J}_{nT}) \\ \frac{1}{\sigma^2_{\xi 0}}\mathbf{R}'_{nT}\mathbf{J}_{nT}\Xi_{nT} - \big(\mathrm{tr}(\mathbf{G}_{2,nT}\mathbf{J}_{nT}), \mathrm{tr}(\mathbf{G}_{3,nT}\mathbf{J}_{nT}), 0\big)' \\ \frac{1}{\sigma^2_{\xi 0}}\varepsilon'_{nT}\mathbf{J}_{nT}\Xi_{nT} \\ \Sigma^{-1}_{\epsilon 0}\otimes \mathbf{K}'_{nT}\mathbf{J}_{nT}\mathrm{vec}(\varepsilon_{nT}) + \mathrm{vec}\left(\frac{n-1}{T}\left(\sum_{t=1}^{T-1}\sum_{h=1}^{T-t}\kappa'^{(h-1)}\right)\Sigma^{-1}_{\epsilon 0}\right) - \frac{1}{\sigma^2_{\xi 0}}\delta_0\otimes \mathbf{K}'_{nT}\mathbf{J}_{nT}\Xi_{nT} \\ -\frac{(n-1)(T-1)}{2\sigma^2_{\xi 0}} + \frac{1}{2\sigma^2_{\xi 0}}\Xi'_{nT}\mathbf{J}_{nT}\Xi_{nT} \\ -\frac{(n-1)(T-1)}{2}\frac{\partial \ln|\Sigma_{\epsilon 0}|}{\partial \alpha} - \frac{1}{2}\frac{\partial}{\partial\alpha}\left(\mathrm{tr}(\Sigma^{-1}_{\epsilon 0}\varepsilon'_{nT}\mathbf{J}_{nT}\varepsilon_{nT})\right) \end{pmatrix},$$

where $\mathbf{G}_{1,nT}(\lambda,\gamma,\rho) = \mathbf{W}_{nT}\mathbf{S}^{-1}_{nT}(\lambda,\gamma,\rho)$, $\mathbf{G}_{2,nT}(\lambda,\gamma,\rho) = \mathbf{I}_{nT,-1}\mathbf{S}^{-1}_{nT}(\lambda,\gamma,\rho)$, $\mathbf{G}_{3,nT}(\lambda,\gamma,\rho) = \mathbf{W}_{n,T-1}\mathbf{S}^{-1}_{nT}(\lambda,\gamma,\rho)$, and $\mathbf{G}_{j,nT} = \mathbf{G}_{j,nT}(\lambda_0,\gamma_0,\rho_0)$ for $j = 1,2,3$, where $\mathbf{I}_{nT}$ is the $nT\times nT$ identity matrix and $\mathbf{I}_{nT,-1}$ is a one-time lagged identity matrix, i.e.,

$$\mathbf{I}_{nT,-1} = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ I_n & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & \dots & I_n & 0 \end{pmatrix}.$$

The negative Hessian matrix of the concentrated log-likelihood is given as

$$\mathrm{E}\left(-\frac{\partial^2 \ln L^c_{nT}(\theta_0)}{\partial\theta\partial\theta'}\right) = \frac{1}{\sigma^2_{\xi 0}}\times$$

$$\begin{pmatrix} I_{\lambda\lambda} & * & * & * & * & \mathbf{0}_{1\times J} \\ \mathrm{E}\mathbf{R}'_{nT}\mathbf{J}_{nT}\mathbf{Q}_{1,nT} & \mathrm{E}\mathbf{R}'_{nT}\mathbf{J}_{nT}\mathbf{R}_{nT} & \mathbf{0}_{(k_1+2)\times p} & * & \mathbf{0}_{(k_1+2)\times 1} & \mathbf{0}_{(k_1+2)\times J} \\ \mathrm{E}[\varepsilon'_{nT}\mathbf{J}_{nT}\mathbf{Q}_{1,nT}] & \mathbf{0}_{p\times(k_1+2)} & \mathrm{E}\varepsilon'_{nT}\mathbf{J}_{nT}\varepsilon_{nT} & \mathbf{0}_{p\times k_{\phi_2}} & \mathbf{0}_{p\times 1} & \mathbf{0}_{p\times J} \\ +\mathrm{E}[\varepsilon'_{nT}\mathbf{J}_{nT}(\mathbf{G}_{1,nT}\varepsilon_{nT}\delta_0)] & & & & & \\ -\delta_0\otimes\mathrm{E}\mathbf{K}'_{nT}\mathbf{J}_{nT}textbfQ_{1,nT} & -\delta_0\otimes\mathrm{E}\mathbf{K}'_{nT}\mathbf{J}_{nT}\mathbf{R}_{nT} & \mathbf{0}_{k_{\phi_2}\times p} & I_{\Phi_2\Phi_2} & \mathbf{0}_{k_{\phi_2}\times 1} & \mathbf{0}_{k_{\phi_2}\times J} \\ \mathrm{Etr}(\mathbf{G}_{1,nT}) & \mathbf{0}_{1\times(k_1+2)} & \mathbf{0}_{1\times p} & \mathbf{0}_{1\times k_{\phi_2}} & \frac{1}{\sigma^2_{\xi 0}}\left(\frac{nT}{2} - T - n + 1\right) & \mathbf{0}_{1\times J} \\ \mathbf{0}_{J\times 1} & \mathbf{0}_{J\times(k_1+2)} & \mathbf{0}_{J\times p} & \mathbf{0}_{J\times k_{\phi_2}} & \mathbf{0}_{J\times 1} & I_{\alpha\alpha} \end{pmatrix} + O(n), \tag{4}$$

where $\mathbf{Q}_{1,nT} = \mathbf{G}_{1,nT}(\mathbf{X}_{1,nT}\beta_0 + \ell_0(\gamma_0,\rho_0) + \varepsilon_{nT}\delta_0 + c_{1,nT,0} + \tau_{1,nT,0})$, $I_{\lambda\lambda} = \mathrm{E}\big(\mathbf{Q}'_{1,nT}\mathbf{J}_{nT}\mathbf{Q}_{1,nT} + \sigma^2_{\xi 0}\mathrm{tr}(\mathbf{G}^2_{1,nT} + \mathbf{G}'_{1,nT}\mathbf{J}_{nT}\mathbf{G}_{1,nT})\big)$, $I_{\phi_2\phi_2} = (\sigma^2_{\xi 0}\Sigma^{-1}_{\epsilon 0} + \delta_0\delta_0')\otimes\mathrm{E}\mathbf{K}'_{nT}\mathbf{J}_{nT}\mathbf{K}_{nT}$, and $I_{\alpha\alpha}$ is a $J\times J$ matrix with its $(k,j)$ element being $\frac{nT}{2}\sigma^2_{\xi 0}\mathrm{tr}\left(\Sigma^{-1}_{\epsilon 0}\frac{\partial\Sigma_{\epsilon 0}}{\partial\alpha_k}\Sigma^{-1}_{\epsilon 0}\frac{\partial\Sigma_{\epsilon 0}}{\partial\alpha_j}\right)$.

Now similarly to Qu and Lee (2015), one may set a hypothesis for testing the endogeneity of $W_{nT}$ as $H_0^{QL}$: $\Sigma_\epsilon^{-1/2}\sigma_{v\epsilon}/\sigma_v = 0$. Accordingly, a score test for $H_0^{QL}$ can be constructed from

$$g_{nT}(\hat{\vartheta}) = \frac{1}{\hat{\sigma}_v}\hat{\Sigma}_\epsilon^{-1/2}\big(\mathbf{Z}_{nT} - \mathbf{K}_{nT}\hat{\Phi}_2\big)'\big((\mathbf{I}_{nT} - \hat{\lambda}\mathbf{W}_{nT})\mathbf{Y}_{nT} - \mathbf{R}_{nT}\hat{\phi}_1 - \ell(\hat{\gamma},\hat{\rho})\big),$$

where $\hat{\vartheta} = (\hat{\lambda}, \hat{\phi}_1', \hat{\phi}_2', \hat{\sigma}_v^2, \hat{\alpha}')'$ with $\phi_2 = \text{vec}(\Phi_2)$ and $\alpha := \alpha(\Sigma_\epsilon)$ representing a column vector consisting of distinct elements in $\Sigma_\epsilon$, is the ML estimates under the null.

Denoting $\vartheta_0$ as the true value, $g_{nT}(\vartheta_0) = \frac{1}{\sigma_{v0}}\Sigma_\epsilon^{-1/2}\varepsilon_{nT}'\mathbf{V}_{nT}$. By the mean value theorem,

$$g_{nT}(\hat{\vartheta}) = g_{nT}(\vartheta_0) - \frac{1}{nT}\frac{\partial g_{nT}(\bar{\vartheta})}{\partial \vartheta'}\left(\frac{\partial^2 \ln L_{1,nT}^{c,H_0^{QL}}(\dot{\vartheta})}{\partial\vartheta\partial\vartheta'}\right)^{-1}\frac{\partial \ln L_{1,nT}^{c,H_0^{QL}}(\vartheta 0)}{\partial\vartheta},$$

where $\bar{\vartheta}, \dot{\vartheta} \in (\hat{\vartheta}, \vartheta_0)$ and $\ln L_{1,nT}^{c,H_0^{QL}}(\vartheta)$ is the unbiased concentrated log-likelihood function under $H_0^{QL}$. Under the null of independence between $\varepsilon_{nT}$ and $\mathbf{V}_{nT}$, $\text{E}\Big(\frac{1}{nT}\frac{\partial g_{nT}(\vartheta_0)}{\partial\vartheta'}\Big) = 0$. Hence,

$$\frac{1}{\sqrt{nT}}g_{nT}(\hat{\vartheta}) = \frac{1}{\sqrt{nT}}g_{nT}(\vartheta_0) + o_p(1) \xrightarrow{d} N(0, I_p).$$

Hence, the score test statistic is given as

$$\begin{aligned} LM^{QL}(\hat{\vartheta}) = &\frac{1}{\hat{\sigma}_v^2}\big((\mathbf{I}_{nT} - \hat{\lambda}\mathbf{W}_{nT})\mathbf{Y}_{nT} - \mathbf{R}_{nT}\hat{\phi}_1 - \ell(\hat{\gamma},\hat{\rho})\big)'\big(\mathbf{Z}_{nT} - \mathbf{K}_{nT}\hat{\phi}_2\big) \\ &\times\ \hat{\Sigma}_\epsilon^{-1}\big(\mathbf{Z}_{nT} - \mathbf{K}_{nT}\hat{\phi}_2\big)'\big((\mathbf{I}_{nT} - \hat{\lambda}\mathbf{W}_{nT})\mathbf{Y}_{nT} - \mathbf{R}_{nT}\hat{\phi}_1 - \ell(\hat{\gamma},\hat{\rho})\big) \xrightarrow{d} \chi_p^2. \end{aligned}$$

# 3 Score Test for Endogeneity of $W$ with Local Misspecifications

In this section, the concept of local parametric misspecification is introduced while developing a score test for testing the endogeneity of $W_{nt}$. This approach leverages asymptotic interpretations of the properties of large-scale networks, emphasizing weak ties through dilution effects.

Let $\theta = (\lambda, \phi_1', \delta', \phi_2', \sigma_\xi^2, \alpha')'$ be a set of parameters with its true parameter $\theta_0$, where $\phi_1 = (\gamma, \rho, \beta')'$, $\phi_2 = \text{vec}(\Phi_2)$ with $\Phi_2 = (\kappa', \Gamma')'$, and $\alpha$ is the $J \times 1$ column vector of distinct elements in $\Sigma_\epsilon$. Suppose the set of parameters $\theta$ is partitioned by the *three* types of parameters, $(\delta', \eta', \omega')'$, where $\delta$ is our *testing* parameter, $\eta$ consists of spatial dependence parameters susceptible to *local misspecifications*, and $\omega$ is the *nuisance* parameters containing the rest of the parameters. In this partition, $(\delta', \eta')'$ represent the parameters of our interest. Let $L_\psi(\theta) = \frac{1}{nT}\frac{\partial \ln L_{nT}^c(\theta)}{\partial\psi}$ be a score function with respect to $\psi$, where $\psi \in \{\delta, \eta, \omega\}$. Let $\mathcal{I}_{\psi\psi} := \mathrm{E}\frac{1}{nT}\Big( - \frac{\partial^2 \ln L^c_{nT}(\theta)}{\partial\psi\partial\psi'}\Big)$ with $I_{\psi\psi}(\theta) = -\frac{1}{nT}\frac{\partial^2 \ln L^c_{nT}(\theta)}{\partial\psi\partial\psi'}$ such that $I_{\psi\psi}(\theta) = \mathcal{I}_{\psi\psi}(\theta) + o_p(1)$ uniformly on $\Theta$.

From (3), a test to determine the endogeneity of $W_{nt}$ can be formulated by examining whether $\delta_0$ equals zero. This is because, given that $W_{nt}$ is a function of the auxiliary error term ($e_{nt}$), a value of $\delta_0 = 0$ implies that the outcome error is purely idiosyncratic, indicating a lack of evidence for the endogeneity of $W_{nt}$. Conversely, if $\delta_0 \neq 0$, it suggests that the outcome error is influenced by $e_{nt}$, indicating the endogeneity of $W_{nt}$. Therefore, the null hypothesis for assessing the endogeneity of $W_{nt}$ can be formulated as $H_0^\delta : \delta_0 = 0$ (indicating no evidence for the endogeneity of $W_{nt}$), while the alternative hypothesis is $H_a^\delta : \delta_0 \neq 0$ (suggesting the endogeneity of $W_{nt}$).

## 3.1 Challenges

Our testing primarily encounters two challenges: (i) the incidental parameters problem in panel data; and (ii) misspecifications in spatial dependence parameters or distributional misspecification. For misspecifications in spatial dependence parameters, this involves implications on inference when the score test is evaluated at estimates from a model where spatial dependence parameters are assumed to be zero, but are in fact non-zero but locally misspecified.

### 3.1.1 Biases due to the Incidental Parameters Problem

As is well known, the two-way fixed effects model yields the incidental parameters problem (Neyman and Scott 1948). Consequently, the score function of $\ln L^c_{nT}(\theta)$ is biased. According to Qu, Lee, and Yu (2017), this bias can be decomposed into an unbiased part and a biased part as

$$\frac{\partial \ln L^c_{nT}(\theta_0)}{\partial\theta} = \frac{\partial \ln L^c_{1,nT}(\theta_0)}{\partial\theta} + \Delta_{nT}, \tag{5}$$

where

$$\Delta_{nT} = \mathrm{E}\left(\frac{\partial \ln L^c_{nT}(\theta_0)}{\partial\theta}\right) = \begin{pmatrix} \mathrm{tr}(\mathbf{W}_{nT}\mathbf{S}^{-1}_{nT}\mathbf{J}_{nT}) - \mathrm{tr}(\mathbf{G}_{1,nT}) \\ \big(\mathrm{tr}(\mathbf{G}_{2,nT}\mathbf{J}_{nT}), \mathrm{tr}(\mathbf{G}_{3,nT}\mathbf{J}_{nT}), 0\big)' \\ 0_{p\times 1} \\ -\mathrm{vec}\left(\frac{n-1}{T}\left(\sum_{t=1}^{T-1}\sum_{h=1}^{T-t}\kappa'^{(h-1)}\right)\Sigma_{\epsilon 0}^{-1}\right) \\ -\frac{n+T-1}{2\sigma^2_{\xi 0}} \\ -\frac{n+T-1}{2}\frac{\partial \ln|\Sigma_{\epsilon 0}|}{\partial\alpha} \end{pmatrix}.$$

Thus, the unbiased score function is asymptotically normally distributed as

$$\frac{1}{\sqrt{nT}}\frac{\partial \ln L^c_{1,nT}(\theta_0)}{\partial\theta} \xrightarrow{d} N\Big(0,\ \mathcal{I}_{\theta\theta}\Big),$$

where $\mathcal{I}_{\theta\theta} := \lim\limits_{n,T\to\infty} \ell_{nT,\theta_0} = -\frac{1}{nT}\mathrm{E}\left(\frac{\partial^2 \ln L^c_{nT}(\theta_0)}{\partial\theta\partial\theta'}\right)$. Since the biases are from the two-way fixed effects, $\Delta_{nT}$ is decomposed into the bias from the individual- and that from the time fixed effects, respectively, i.e.,

$$\Delta_{nT} = (n-1)a_{1,\theta_0} + Ta_{2,\theta_0},$$

where the biases $a_{1,\theta_0}$ and $a_{2,\theta_0}$ are of $O(1)$ from individual fixed effects and time fixed effects, respectively. Specifically,

$$a_{1,\theta_0} = \begin{pmatrix} -\frac{1}{n-1}\mathrm{tr}(\mathbf{G}_{1,nT}(\frac{1}{T}1_T1_T'\otimes J_n)) \\ \frac{1}{n-1}[\mathrm{tr}(\mathbf{G}_{2,nT}(\frac{1}{T}1_T1_T'\otimes J_n)), \mathrm{tr}(\mathbf{G}_{3,nT}(\frac{1}{T}1_T1_T'\otimes J_n)), 0]' \\ 0_{p\times 1} \\ -\mathrm{vec}\left(\frac{1}{T}\left(\sum\limits_{t=1}^{T-1}\sum\limits_{h=1}^{T-t}\kappa_0'^{(h-1)}\right)\Sigma_{\epsilon 0}^{-1}\right) \\ -\frac{1}{2\sigma^2_{\xi 0}} \\ -\frac{1}{2}\frac{\partial \ln|\Sigma_{\epsilon 0}|}{\partial\alpha} \end{pmatrix}, \quad a_{2,\theta_0} = \begin{pmatrix} -\frac{1}{T}\mathrm{tr}(\mathbf{G}_{1,nT}(I_T\otimes \frac{1}{n}1_n1_n')) \\ 0_{(k_1+2)\times 1} \\ 0_{p\times 1} \\ 0_{p(p+k_2)\times 1} \\ -\frac{1}{2\sigma^2_{\xi 0}} \\ -\frac{1}{2}\frac{\partial \ln|\Sigma_{\epsilon 0}|}{\partial\alpha} \end{pmatrix}.$$

The asymptotic distribution of the unbiased score function can be equivalently represented as

$$\frac{1}{\sqrt{nT}}\frac{\partial \ln L^c_{nT}(\theta_0)}{\partial\theta} - \frac{\Delta_{nT}}{\sqrt{nT}} \xrightarrow{d} N\Big(0,\ \mathcal{I}_{\theta\theta}\Big).$$

Since $\Delta_{nT} = (n-1)a_{1,\theta_0} + Ta_{2,\theta_0}$, it is equivalent to

$$\frac{1}{\sqrt{nT}}\frac{\partial \ln L^c_{nT}(\theta_0)}{\partial\theta} - \sqrt{\frac{n}{T}}a_{1,\theta_0} - \sqrt{\frac{T}{n}}a_{2,\theta_0} \xrightarrow{d} N\Big(0,\ \mathcal{I}_{\theta\theta}\Big).$$

Denoting $\Delta_1 := \sqrt{\frac{n}{T}}a_{1,\theta_0}$ and $\Delta_2 := \sqrt{\frac{T}{n}}a_{2,\theta_0}$,

$$\frac{1}{\sqrt{nT}}\frac{\partial \ln L^c_{nT}(\theta_0)}{\partial\theta} - \Delta_1 - \Delta_2 \xrightarrow{d} N\Big(0,\ \mathcal{I}_{\theta\theta}\Big).$$

Now with large $n$ and $T$, there are three cases for the ratio of $\frac{n}{T}$: (i) If $\frac{n}{T}\to k<\infty$, $\Delta_1$ and $\Delta_2$ do not vanish and so $\frac{\partial \ln L^c_{nT}(\theta_0)}{\partial\theta}$ is not centered around zero. (ii) If $\frac{n}{T}\to 0$, the score function has a degenerating distribution as $\frac{1}{T}\frac{\partial \ln L^c_{nT}(\theta_0)}{\partial\theta} - \sqrt{\frac{n}{T}}\Delta_2 \xrightarrow{p} 0$. (iii) If $\frac{n}{T}\to\infty$, the score function again has a degenerating distribution as $\frac{1}{n}\frac{\partial \ln L^c_{nT}(\theta_0)}{\partial\theta} - \sqrt{\frac{T}{n}}\Delta_1 \xrightarrow{p} 0$. Since a score test requires a non-degenerating distribution, $\frac{n}{T} \to k < \infty$ is parsimoniously adopted and thus $\Delta_1$ and $\Delta_2$ are $O(1)$.

Hence the score function along with the asymptotic rate of $\sqrt{nT}$ has biases, which do not vanish even with large $n$ and $T$. This is not the case in the cross-section where the score function has zero means at $\theta_0$, i.e., the score function does not suffer from biases. Unlike the cross-section, however, the biases arise in a SDPD model due to the incidental parameters of the individual- and time-fixed effects. They do not vanish even with large $n$ and $T$, yielding increased type I error.

#### 3.1.2 Local Parametric Misspecifications

The local alternatives (Neyman 1937) quantify $\eta$, modeling weak ties by spatial dependence parameters drifting towards zero with large $n$ and $T$:

$$H^{\eta}_{a} : \eta_0 = \nu/\sqrt{nT},$$

where $\nu$ are some unknown bounded constant vectors. Note that the score function is supposed to be evaluated under the joint null ($H^{\delta,\eta}_0 := (H^{\delta}_0,\ H^{\eta}_0)$) so that one only needs to estimate a standard two-way fixed effects model. Then the violation of $H^{\eta}_0$ may or may not affect the score function of $\delta$, depending on the relationship between $\delta$ and $\eta$. That is, the score function of $\delta$ would be only affected if $\delta$ and $\eta$ have some dependent patterns so that violation of $H^{\eta}_0$ affects the score function of $\delta$ in some way. If there are no such patterns between $\delta$ and $\eta$, it does not matter whether $H^{\eta}_0$ is violated or not, for the purpose of valid inference on $\delta$. Thus, the presence of local parametric misspecification in $\eta$ may threat the valid inference on $\delta$ if the covariance between the score functions of $(\delta, \eta)$ is nonzero.[9]

Specifically, from (4), the covariance of the two score functions for $(\delta, \eta)$ at $\theta_0$ is given by

$$\mathrm{E}\left( -\frac{\partial^2 \ln L^c_{nT}(\theta_0)}{\partial\delta\partial\eta'}\right) = \Big(\mathrm{E}\varepsilon'_{nT}\mathbf{J}_{nT}\mathbf{Q}_{1,nT} + \mathrm{E}\varepsilon'_{nT}\mathbf{J}_{nT}(\mathbf{G}_{1,nT}\varepsilon_{nT}\delta_0),\ 0_{p\times 1}\Big), \tag{6}$$

where $\varepsilon_{nT} = \mathbf{Z}_{nT} - \mathbf{K}_{nT}\Phi_2$, $\mathbf{Q}_{1,nT} = \mathbf{G}_{1,nT}(\mathbf{X}_{1,nT}\beta_0 + \ell_0(\gamma_0, \rho_0) + \varepsilon_{nT}\delta_0 + c_{1,nT,0} + \tau_{1,nT,0})$, and $\mathbf{G}_{1,nT} = \mathbf{W}_{nT}\mathbf{S}^{-1}_{nT}(\lambda_0, \gamma_0, \rho_0)$ with $\mathbf{S}_{nT}(\lambda, \gamma, \rho) = \mathbf{I}_{nT} - \lambda\mathbf{W}_{nT} - \gamma\mathbf{I}_{nT,-1} - \rho\mathbf{W}_{n,T-1}$. This implies that the endogeneity of the concurrent spatial weight matrix is independent of the spatial dynamic coefficient, given that the error terms are *i.i.d.* over $t$. Therefore, the dependence between the two score functions of $(\delta, \eta)$ comes from the concurrent spatial dependence parameter, $\lambda$.

[9]Since a score function is a function of a parameter, the relationship between the parameters can be equivalently learned from the covariance between the score functions of the parameters.

From (6), the covariance of the two score functions for $(\delta, \eta)$ under the joint null ($H_0^{\delta,\eta}$) would be zero only if $\delta_0$ is zero, i.e., $W$ is exogenous, as well as the expected value of the residuals in the auxiliary equations is zero. Since $\tilde{\theta}$ is estimated under the null that $\lambda = 0$, the expected value of the residuals from the auxiliary equation will be zero only if $\lambda_0 = 0$. If not, $\tilde{\theta}$, which is obtained by omitting the endogenous variable $W_{nT}Y_{nT}$, will result in inconsistent estimates for $(\kappa', \Gamma')'$. This, in turn, will lead to a nonzero expected value of the residuals for $\varepsilon_{nT}$.

Thus, $\delta$ and $\eta$ are independent or the covariance of the two score functions for $(\delta, \eta)$ is zero, only in the very limited situation where concurrent spatial dependence is absent and $W_{nT}$ is given exogenous. Otherwise, in most cases, their covariance is nonzero, meaning $\delta$ and $\eta$ are dependent.

### 3.2 Main Result: The Robust Score Test

Consider the alternatives for $\delta$ and $\eta$ as $H_a^\delta : \delta_0 = \zeta/\sqrt{nT}$ and $H_a^\eta : \eta_0 = \nu/\sqrt{nT}$, respectively, where $\zeta$ and $\nu$ are unknown bounded constants.[10] The first-order Taylor expansions of the score functions give (Appendix A.3)

$$\sqrt{nT}L_\delta(\tilde{\theta}) \overset{a}{=} \sqrt{nT}L_\delta(\theta_0) - \mathcal{I}_{\delta\omega}\mathcal{I}_{\omega\omega}^{-1}\sqrt{nT}L_\omega(\theta_0) + \mathcal{I}_{\delta\cdot\omega}\zeta + \mathcal{I}_{\delta\eta\cdot\omega}\nu,$$

where '$a$' stands for 'asymptotically,' $\mathcal{I}_{\delta\cdot\omega} = \mathcal{I}_{\delta\delta} - \mathcal{I}_{\delta\omega}\mathcal{I}_{\omega\omega}^{-1}\mathcal{I}_{\omega\delta}$ and $\mathcal{I}_{\delta\eta\cdot\omega} = \mathcal{I}_{\delta\eta} - \mathcal{I}_{\delta\omega}\mathcal{I}_{\omega\omega}^{-1}\mathcal{I}_{\omega\eta}$, with $\zeta$ and $\nu$ represent some unknown bounded constant vectors. Note that by definition, $\mathcal{I}_{\psi\cdot\omega}$ is the partial variance of the score function of $\psi \in \{\delta, \eta\}$ conditioning on $L_\omega$ and $\mathcal{I}_{\delta\eta\cdot\omega}$ is the partial covariance between the score functions of $\delta$ and $\eta$, respectively. Since $L_\delta$ and $L_\omega$ are jointly normal distributed:

$$\begin{pmatrix} \sqrt{nT}L_\delta(\theta_0) \\ \sqrt{nT}L_\omega(\theta_0) \end{pmatrix} - \begin{pmatrix} \Delta_{1,\delta}(\theta_0) + \Delta_{2,\delta}(\theta_0) \\ \Delta_{1,\omega}(\theta_0) + \Delta_{2,\omega}(\theta_0) \end{pmatrix} \xrightarrow{d} N\left(0, \begin{pmatrix} \mathcal{I}_{\delta\delta} & \mathcal{I}_{\delta\omega} \\ \mathcal{I}_{\omega\delta} & \mathcal{I}_{\omega\omega} \end{pmatrix}\right), \tag{7}$$

this implies

$$\sqrt{nT}L_\delta(\tilde{\theta}) \xrightarrow{d} N\left(\underbrace{(\Delta_{1,\delta}(\theta_0) + \Delta_{2,\delta}(\theta_0)) - \mathcal{I}_{\delta\omega}\mathcal{I}_{\omega\omega}^{-1}(\Delta_{1,\omega}(\theta_0) + \Delta_{2,\omega}(\theta_0))}_{:=\mathcal{B}_{\delta\omega}} + \mathcal{I}_{\delta\cdot\omega}\zeta + \mathcal{I}_{\delta\eta\cdot\omega}\nu, \mathcal{I}_{\delta\cdot\omega}\right), \tag{8}$$

where $\mathcal{B}_{\psi\omega} := \big(\Delta_{1,\psi}(\theta_0) + \Delta_{2,\psi}(\theta_0)\big) - \mathcal{I}_{\psi\omega}\mathcal{I}_{\omega\omega}^{-1}\big(\Delta_{1,\omega}(\theta_0) + \Delta_{2,\omega}(\theta_0)\big)$ for $\psi \in \{\delta, \eta\}$ is the bias. Thus, $L_\delta(\tilde{\theta})$ is not centered around zero due to the biases from the two-way fixed effects (or incidental parameters) and parametric misspecifications in $\eta$ (i.e., $\nu \neq 0$). This needs to be adjusted so that

[10] Since a score test is developed under the null hypothesis, alternative forms of $\delta$ are not required. However, for the purpose of delineating the noncentrality parameters in Theorem, the alternative hypothesis $H_a^\delta$ is introduced.

the score function of testing parameter ($\delta$) is centered around zero. Otherwise, the score test has positive noncentrality parameter and thus suffers from the over-rejection of the null or increased type I error (Davidson and Mackinnon 1987; Saikkonen 1989).

To this end, I make two adjustments toward the biases from the two-way fixed effects and local parametric misspecifications in $\eta$. Hence there are four cases to be considered: (i) Biased and parametrically misspecified; (ii) Biased but robust to parametric misspecification; (iii) Unbiased but parametrically misspecified; (iv) Unbiased and robust to parametric misspecification. The following Theorem summarizes the results of each adjustment.

**Theorem.** Let $\mathcal{B}_{\psi\omega} := (\Delta_{1,\psi}(\theta_0) + \Delta_{2,\psi}(\theta_0)) - \mathcal{I}_{\psi\omega}\mathcal{I}_{\omega\omega}^{-1}(\Delta_{1,\omega}(\theta_0) + \Delta_{2,\omega}(\theta_0))$, $\psi \in \{\delta, \eta\}$ and $\zeta$ and $\nu$ are some unknown bounded constants. Under the stated assumptions and $\frac{n}{T} \to k < \infty$, the following results hold under $H_a^\delta : \delta_0 = \zeta/\sqrt{nT}$ and $H_a^\eta : \eta_0 = \nu/\sqrt{nT}$.

(i) If the score function of $\delta$ is biased and parametrically misspecified in $\eta$,

$$RS_\delta^{B,P}(\tilde{\theta}) := nTL'_\delta(\tilde{\theta})I_{\delta\cdot\omega}^{-1}(\tilde{\theta})L_\delta(\tilde{\theta}) \xrightarrow{d} \chi_p^2(\varphi_1),$$

where the superscript '$B, P$' stands for 'biased and parametrically misspecified', and $\varphi_1 = \mathcal{B}'_{\delta\omega}\mathcal{I}_{\delta\cdot\omega}^{-1}\mathcal{B}_{\delta\omega} + \zeta'\mathcal{I}_{\delta\cdot\omega}\zeta + \nu'\mathcal{I}_{\eta\delta\cdot\omega}\mathcal{I}_{\delta\cdot\omega}^{-1}\mathcal{I}_{\delta\eta\cdot\omega}\nu + 2(\mathcal{B}_{\delta\omega}\zeta + \mathcal{B}_{\delta\omega}\mathcal{I}_{\delta\cdot\omega}^{-1}\mathcal{I}_{\delta\eta\cdot\omega}\nu + \zeta'\mathcal{I}_{\delta\eta\cdot\omega}\nu)$ is the noncentrality parameter.

(ii) If the score function of $\delta$ is biased but robust to parametric misspecification in $\eta$,

$$RS_\delta^{B}(\tilde{\theta}) := nTL_\delta^{*\prime}(\tilde{\theta})\big(I_{\delta\cdot\omega}(\tilde{\theta}) - I_{\delta\eta\cdot\omega}(\tilde{\theta})I_{\eta\cdot\omega}^{-1}(\tilde{\theta})I_{\eta\delta\cdot\omega}(\tilde{\theta})\big)^{-1}L_\delta^*(\tilde{\theta}) \xrightarrow{d} \chi_p^2(\varphi_2),$$

where the superscript $B$ stands for 'biased', $L_\delta^*(\tilde{\theta}) := L_\delta(\tilde{\theta}) - I_{\delta\eta\cdot\omega}(\tilde{\theta})I_{\eta\cdot\omega}^{-1}(\tilde{\theta})L_\eta(\tilde{\theta})$ represents the biased score function that is adjusted for local parametric misspecification, and $\varphi_2 = (\mathcal{B}_{\delta\omega} - \mathcal{I}_{\delta\eta\cdot\omega}\mathcal{I}_{\eta\cdot\omega}^{-1}\mathcal{B}_{\eta\omega})'(\mathcal{I}_{\delta\cdot\omega} - \mathcal{I}_{\delta\eta\cdot\omega}\mathcal{I}_{\eta\cdot\omega}^{-1}\mathcal{I}_{\eta\delta\cdot\omega})^{-1}(\mathcal{B}_{\delta\omega} - \mathcal{I}_{\delta\eta\cdot\omega}\mathcal{I}_{\eta\cdot\omega}^{-1}\mathcal{B}_{\eta\omega}) + 2\zeta'(\mathcal{I}_{\delta\cdot\omega} - \mathcal{I}_{\delta\eta\cdot\omega}\mathcal{I}_{\eta\cdot\omega}^{-1}\mathcal{I}_{\eta\delta\cdot\omega})'(\mathcal{B}_{\delta\omega} - \mathcal{I}_{\delta\eta\cdot\omega}\mathcal{I}_{\eta\cdot\omega}^{-1}\mathcal{B}_{\eta\omega}) + \zeta'(\mathcal{I}_{\delta\cdot\omega} - \mathcal{I}_{\delta\eta\cdot\omega}\mathcal{I}_{\eta\cdot\omega}^{-1}\mathcal{I}_{\eta\delta\cdot\omega})\zeta$ is the noncentrality parameter.

(iii) If the score function of $\delta$ is unbiased but parametrically misspecified in $\eta$,

$$RS_\delta^{P}(\tilde{\theta}) := nTC'_\delta(\tilde{\theta})I_{\delta\cdot\omega}^{-1}(\tilde{\theta})C_\delta(\tilde{\theta}) \xrightarrow{d} \chi_p^2(\varphi_3),$$

where the superscript $P$ represents 'parametrically misspecified', $C_\delta(\tilde{\theta})$ is the biaspwas -corrected score function, and $\varphi_3 = \zeta'\mathcal{I}_{\delta\cdot\omega}\zeta + 2\zeta'\mathcal{I}_{\eta\delta\cdot\omega}\nu + \nu'\mathcal{I}_{\eta\delta\cdot\omega}\mathcal{I}_{\delta\cdot\omega}^{-1}\mathcal{I}_{\delta\eta\cdot\omega}\nu$ is the noncentrality parameter.

(iv) If the score function of $\delta$ is unbiased and robust to parametric misspecification in $\eta$,

$$RS_\delta^*(\tilde{\theta}) := nTC_\delta^{*\prime}(\tilde{\theta})\big(I_{\delta\cdot\omega}(\tilde{\theta}) - I_{\delta\eta\cdot\omega}(\tilde{\theta})I_{\eta\cdot\omega}^{-1}(\tilde{\theta})I_{\eta\delta\cdot\omega}(\tilde{\theta})\big)^{-1}C_\delta^*(\tilde{\theta}) \xrightarrow{d} \chi_p^2(\varphi_4),$$

where $C_\delta^*(\tilde{\theta}) := C_\delta(\tilde{\theta}) - I_{\delta\eta\cdot\omega}(\tilde{\theta})I_{\eta\cdot\omega}^{-1}(\tilde{\theta})C_\eta(\tilde{\theta})$ is the adjusted bias-corrected score function, and $\varphi_4 := \zeta'(\mathcal{I}_{\delta\cdot\omega} - \mathcal{I}_{\delta\eta\cdot\omega}\mathcal{I}_{\eta\cdot\omega}^{-1}\mathcal{I}_{\eta\delta\cdot\omega})\zeta$ is the noncentrality parameter.

**Proof.** See Appendix A.3.

Note that $RS_\delta^*(\tilde{\theta})$ is the most robust version over the other tests, which are either unadjusted at all or partially adjusted. $RS_\delta^*$ fixes the over-rejection of $H_0^\delta$ towards the two sources that yield non-zero biases in the score function. That is, while $RS_\delta^{B,P}(\tilde{\theta}), RS_\delta^B(\tilde{\theta})$, and $RS_\delta^P(\tilde{\theta})$ suffer from the increased type I error with the positive noncentrality parameter due to the incidental parameters problem or parametric misspecifications in $\eta$, $RS_\delta^*(\tilde{\theta})$ gives asymptotically correct size with its asymptotic null distribution being centered chi-square distribution. Specifically, one may notice that the noncentrality parameter of $RS_\delta^*(\tilde{\theta})$ is free from both $\mathcal{B}_{\psi\omega}$ and $\nu$, while that of other tests is involved with either $\mathcal{B}_{\psi\omega}$ or $\nu$, or both.

## 3.3 Discussion for the Adjustments

Our focus is on elucidating two key adjustments: (i) correcting biases in the score function in the presence of incidental parameters, specifically two-way fixed effects, and (ii) adjusting for local parametric misspecification in spatial dependence parameters when the true parameters are indeed non-zero but locally misspecified with large $n$ and $T$.

### 3.3.1 Bias-correction

Note that from the form of the variance in (8), which is a partial variance concerning the nuisance parameter $\omega$, one can notice that the distribution of $L_\delta(\tilde{\theta})$ is obtained by conditioning on $L_\omega(\theta_0)$. This can be interpreted as that the effects of $L_\omega$ are eliminated because $\omega$ is the pure nuisance parameter. Accordingly, the biases from $L_\omega$ are also scaled out, as presented by the scaled factor $\mathcal{I}_{\delta\omega}\mathcal{I}_{\omega\omega}^{-1}$ on $(\Delta_{1,\omega}(\theta_0) + \Delta_{2,\omega}(\theta_0))$. Notice that the form of biases at $\tilde{\theta}$ are different from those in Qu, Lee, and Yu, as the partition of parameters into three types and the process of conditioning $L_\delta(\tilde{\theta})$ on $L_\omega(\theta_0)$ apply to obtain $B_{\delta\omega}$, while the biases in Qu, Lee, and Yu are given as deterministic quantities of the score function without partition on parameters (i.e., $\mathrm{E}\left(\frac{\partial \ln L_{nT}^c(\theta_0)}{\partial\theta}\right)$). This makes the bias correction in this paper different from that of Qu, Lee, and Yu (2017).

### 3.3.2 Adjustment for Misspecifications

Note that the scale factor $\mathcal{I}_{\delta\eta\cdot\omega}$ of the magnitude of the local parametric misspecification in $\eta$ ($i.e., \nu$) is by definition the partial covariance between $C_\delta(\theta_0)$ and $C_\eta(\theta_0)$ conditioning on $L_\omega(\theta_0)$, as $\omega$ is the pure nuisance parameter and so the effects of $C_\omega(\theta_0)$ on the other score functions are to be eliminated. Geometrically, it is well known that the angle between $C_\delta(\theta_0)$ and $C_\eta(\theta_0)$ is the arccosine of the partial covariance. Since the score functions are normally distributed as in (7), the orthogonality of the score functions is equivalent to their independence.

Now there are two possible cases under $H_0^\delta$: (i) $\mathcal{I}_{\delta\eta\cdot\omega} = 0$, i.e., $C_\delta(\theta_0)$ and $C_\eta(\theta_0)$ are orthogonal or equivalently independent as the bias-corrected score functions are normally distributed. In this case, either $H_0^\eta$ holds or not does not affect the distribution of $C_\delta(\theta_0)$ and thus $RS_\delta^P(\tilde{\theta})$ is valid; (ii) $\mathcal{I}_{\delta\eta\cdot\omega} \neq 0$, i.e., $C_\delta(\theta_0)$ and $C_\eta(\theta_0)$ are neither orthogonal nor independent. In this case, whether $H_0^\eta$ holds or not does affect the distribution of $C_\delta(\theta_0)$ under a presence of local parametric misspecification in $\eta$, i.e., $\nu \neq 0$. $C_\delta(\tilde{\theta})$ being asymptotically not centered around zero yields that $RS_\delta(\tilde{\theta})$ follows $\chi^2$ distribution with non-zero noncentrality parameter.

A robust version of the score test can be constructed by adjusting $C_\delta(\tilde{\theta})$ so that the score test is centered around zero (Bera and Yoon 1993). Even if the local parametric misspecification in $\eta$ presents, which is a deviation from the null, it does not affect the distribution of $C_\delta(\tilde{\theta})$ as long as $\delta$ and $\eta$ are independent, or equivalently, $C_\delta(\tilde{\theta})$ and $C_\eta(\tilde{\theta})$ are orthogonal. Using the linear projection, one may orthogonalize $C_\delta(\tilde{\theta})$ so that $C_\delta^*(\tilde{\theta}) := C_\delta(\tilde{\theta}) - \mathrm{E}(C_\delta(\tilde{\theta})|C_\eta(\tilde{\theta}))$ is orthogonal to $C_\eta(\tilde{\theta})$.

From the proof of Theorem, it turns out that $\mathrm{E}\big(C_\delta(\theta_0)|C_\eta(\theta_0)\big)$ has the form of $\mathcal{I}_{\delta\eta\cdot\omega}\mathcal{I}_{\eta\cdot\omega}^{-1}C_\eta(\theta_0)$. Notice that this is indeed the linear projection coefficient of projecting $C_\delta(\theta_0)$ onto the plane of $C_\eta(\theta_0)$ because $\mathcal{I}_{\delta\eta\cdot\omega}$ and $\mathcal{I}_{\eta\cdot\omega}^{-1}$ are the partial covariance and variance of the score functions concerning $L_\omega(\theta_0)$, respectively.

## 3.4 Remarks

Three key remarks are provided. Other interesting remarks are provided in Appendix A.4.

### 3.4.1 Advantages over other tests

Alternative hypothesis tests can be developed for the same purpose of model selection between efficient and consistent estimations that depend on the endogeneity of $W$. This remark highlights the advantages of my proposed test over other tests.

*(1) $C(\alpha)$ test:* Let $\theta = (\delta', \omega_c')'$, where $\omega_c = (\lambda, \phi_1', \phi_2', \sigma_\xi^2, \alpha')'$ contains all parameters but $\delta$. In $C(\alpha)$

framework, $w_c$ is regarded as relatively nuisance than our testing parameter $\delta$ so that the effects of the function of $w_c$ should be eliminated out from that of $\delta$. The optimal $C(\alpha)$ test for testing $H_0^\delta:\ \delta = 0$ is formulated as

$$C(\alpha)|_{\bar{\theta}} = nT\big(C_\delta(\bar{\theta}) - I_{\delta\omega_c}(\bar{\theta})I^{-1}_{\omega_c\omega_c}(\bar{\theta})C_{\omega_c}(\bar{\theta})\big)' I^{-1}_{\delta\cdot\omega_c}(\bar{\theta})\big(C_\delta(\bar{\theta}) - I_{\delta\omega_c}(\bar{\theta})I^{-1}_{\omega_c\omega_c}(\bar{\theta})C_{\omega_c}(\bar{\theta})\big) \xrightarrow{d} \chi^2_p(\varphi_c),$$

where $\bar{\theta}$ is a $\sqrt{nT}$-consistent estimator for $\theta$ under $H_0^\delta$ and the noncentrality parameter is $\varphi_c = \zeta'\mathcal{I}_{\delta\cdot\omega_c}\zeta$ with $\mathcal{I}_{\delta\cdot\omega_c} = \mathcal{I}_{\delta\delta} - \mathcal{I}_{\delta\omega_c}\mathcal{I}^{-1}_{\omega_c\omega_c}\mathcal{I}_{\omega_c\delta}$. Note that the score test described in Section 2 has an equivalent null hypothesis to that of the $C(\alpha)$ test. Specifically, $H_0^{QL}: \Sigma_\epsilon^{-1/2}\sigma_{v\epsilon}/\sigma_v = 0$ is equivalent to $H_0^\delta:\ \delta = 0$ where $\delta_0 = \Sigma_{\epsilon 0}^{-1}\sigma_{v\epsilon 0}$ and $\hat{\vartheta}$ is $\sqrt{nT}$-consistent under $H_0^{QL}$.

Let $\hat{\theta} = \underset{\theta:\,\delta=0}{\operatorname{argmax}} \ln L^c_{nT}(\theta)$ be the restricted ML estimates under $H_0^\delta$. Since $\hat{\theta}$ is $\sqrt{nT}$-consistent, we can evaluate $C(\alpha)$ test at $\hat{\theta}$. Hypothetically, $RS^*_\delta(\tilde{\theta})$ is expected to be computationally more efficient than $C(\alpha)$ test because it does not require for $\eta$ to be estimated. In contrast, $C(\alpha)$ test requires the restricted ML estimates under $H_0^\delta: \delta_0 = 0$, i.e., $\eta$ needs to be estimated. A comparison over elapsed times in seconds between $RS^*_\delta(\tilde{\theta})$ and $C(\alpha)|_{\hat{\theta}}$ is reported in Section 4.2.

Interestingly, $RS^*_\delta(\tilde{\theta})$ is asymptotically equivalent to the $C(\alpha)$ test (Neyman 1959) of a composite statistical hypothesis involving unknown nuisance parameter (Bera and Yoon 1993). Thus, it addresses the asymptotic optimality of $RS^*_\delta(\tilde{\theta})$, while $RS^*_\delta(\tilde{\theta})$ gives equivalently valid inference as good as $C(\alpha)$ test.

*(2) Wald, likelihood ratio, and (conventional) Rao's score tests:* The proposed score test is evaluated under the joint null for $(\delta', \eta')' = (0', 0')'$, ensuring that the estimates are obtained from non-spatial models. In contrast, the trilogy of tests known as the Wald, likelihood ratio, and (conventional) Rao's score tests for testing the endogeneity of $W$ involve estimates from spatial models. Since the computational cost for spatial models significantly increases with the number of cross-sectional units, they have a higher computational burden than non-spatial models. Thus, the proposed score test is evidently more computationally advantageous than the other three tests.

*(3) Durbin-Wu-Hausman (DWH) test:* The endogeneity of $W$ can be alternatively tested using the framework of instrumental variable (IV) along with the DWH test. This can be a more general approach because the null hypothesis for the DWH test does not necessarily involve the parameter associated with the endogeneity of $W$. Rather, the DWH test deals with the cases featured by consistency and efficiency of the estimates.

The estimates used for formulating the DWH test depend on the approaches taken to address the endogeneity of $W$, where there are two such approaches: First, if using the control function approach—primarily employed in this paper—to handle the endogeneity of $W$, the null hypothesis for the DWH test is equivalent to the null as in my proposed test, i.e., $H_0^\delta: \ \delta = 0$. Thus, the parameter set $\theta$ is identical to that in the proposed test.

Second, one may alternatively approximate $W$ using the two stage least squares estimation using the IVs as in Kelejian and Piras (2014). This process ensures that the covariance between $Y$ and $Z$ is accounted for in advance. In the context of our setup, one can make $W$ exogenous by constructing $W$ based on $\hat{Z}_{nt} = Z_{n,t-1}\hat{\kappa} + X_{2nt}\hat{\Gamma}$. Since *i.i.d.* over $t$, $Z_{n,t-1}$ can be used as an instrumental variable for $Z_{nt}$. In this approach, the parameter set does not include $\delta$ and $\sigma_\xi^2$ as in the control function approach. Instead, there is no specific parameter representing the endogeneity of $W$. Thus, the null hypothesis for the DWH test is equivalent to a general null hypothesis of $H_0: \ W$ is exogenous.

However, although the IV estimation method offers a general approach to addressing the endogeneity of $W$ and simplifies the computational process, there are some caveats. First, the approach to predict exogenous $W$ that uses $Z_{n,t-1}$ as IVs for $Z_{nt}$ does not guarantee the quality of the estimator, which is in particular problematic when the dependence to lagged terms is small. In such cases, it may suffer from the weak IV problem, leading to inflated variances and biased, inconsistent parameter estimates. Thus, researchers are encouraged to find strong IVs for $Z_{nt}$, which may not be straightforward in real-world data. This rather highlights the guaranteed advantage of my proposed test, which is equivalent to $C(\alpha)$ test and thus asymptotically optimal. Second, similar to the trilogy of tests, the DWH test still requires the estimation of spatial models. This may result in a higher computational cost than my proposed test, which only requires estimates from non-spatial models.

#### 3.4.2 Caveats Regarding the Unknown Initial Conditions Problem

Following Qu, Lee, and Yu (2017), I adopted the assumption that the initial values are observable, or, the influence of the initial conditions becomes negligible with a large $T$. Thus, the score tests developed in this paper utilize Information Equality, as initial conditions problem is not considered.

However, if initial values are unknown and the initial conditions problem is of significant concern in analysis, one may consider the following options: (i) ensuring that $T$ is considerably larger than $n$ so that the bias of order $1/T$ diminishes as $n$ goes to infinity (Nickell 1981); (ii) predict the initial values using the idea of jackknife instrumental variables (JIVE) (Philips and Hale 1977; Angrist, Imbens, and Krueger 1999; Blomquist and Dahlberg 1999).

Regarding (ii), the rationale is as follows: Since the essential problem of initial conditions problem lies in the correlation between the initial values and the unobserved effects and such endogeneity of $Y_{i0}$ is not addressed in the current estimation, predicting the initial values using JIVEs may help reducing correlation with their own unobserved effects as one's initial values are represented as a weighted sum of others' values in the first period. The weights can be borrowed from a spatial weight matrix, possibly up to the top 10 largest weights for each individual.

Specifically, the initial values for $i$ can be predicted as $\hat{Y}_{i0} := \frac{1}{n_{j1}} \sum_{j \neq i}^{n} Y_{j1}$ such that $d(i,j) \leq d_0$, where $Y_{j1}$ is the values of dependent variables for $j \neq i$ at time 1, $d(i,j)$ indicates some distance between $i$ and $j$, $d_0$ is some positive value that determines the nearest neighbors, and $n_{j1}$ is the corresponding number of observations of $j$. Let $\mu_i$ represent $i$'s fixed effects. Then the unobserved effects for $\hat{Y}_{i0}$ are represented as

$$\frac{1}{n_{j1}} \Big( \sum_{j \neq i}^{n} \max(0, w_{ji}) \mu_i + \sum_{j \neq i}^{n} \mu_j + \sum_{k \neq i,j}^{n} \max(0, w_{kj}) \mu_j \Big),$$

where the effects of $\mu_j$ from the last two terms are expected to be greater than that of $\mu_i$ in the first term in the parenthesis, given the sparsity of a spatial weight matrix and large-scale networks (i.e., large $n$). In this manner, the endogeneity of initial conditions can be partially addressed for practical purposes. The simulation results (Table 2) demonstrate that this approach has desirable practical properties, making it a favorable empirical strategy.

**Table 2 is here**

### 3.4.3 Some Extensions

Depending on the context, the proposed robust score test can be extended to the following cases.

(i) *Extended $\eta$:* If the dependent variable is known or observed to be *weakly* persistent, $\eta$ may include both spatial and time dependence parameters, i.e., $\eta := (\lambda, \gamma, \rho)$, instead of the default setup $\eta := (\lambda, \rho)$. In this case, $\omega$ no longer includes the time dynamic dependence parameter, $\gamma$.

(ii) *Extended specification:* The endogenous variables from $Z_{nt}$ can also be included in the outcome equation (Qu, Lee, and Yu 2017). This is particularly useful when a researcher is interested in examining the effects of a variable which works as a direct explanatory variable for the outcome as well as an underlying trigger yielding spatial dependence with respect to the outcome. Also, our spatial autoregressive model in (3) can be extended to the spatial Durbin model by including spatial lagged terms in $X_{1nt}$ (i.e., $W_{nt}X_{1nt}$) as regressors. Furthermore, the time dynamic dependent variable $Y_{n,t-1}$ can be added in $X_{2nt}$ in the auxiliary equation in case the spillover in the current period is presumed to arise depending on the previous outcome values. This setup may be useful to study the *sequential* spillovers effects formed by the previous outcome.

(iii) *Adjustment in the presence of Information Inequality:* In cases where the Information Equality does *not* hold due to using an incorrect likelihood (e.g., when initial conditions problem occurs or the errors are non-normal), the asymptotic variance of the score test is affected (Bera et al. 2020).

Applying White (1982) and Bera et al. (2020), an adjusted score test, which is further robust to Information Inequality, can be formulated as

$$RS_{\delta}^{**}(\tilde{\theta}) = nTC_{\delta}^{*\prime}(\tilde{\theta})V_{\delta\cdot\omega}^{-1}(\tilde{\theta})C_{\delta}^{*}(\tilde{\theta}) \xrightarrow{d} \chi_p^2(\varphi),$$

with the noncentrality parameter $\varphi := \zeta'(\mathcal{I}_{\delta\cdot\omega} - \mathcal{I}_{\delta\eta\cdot\omega}\mathcal{I}_{\eta\cdot\omega}^{-1}\mathcal{I}_{\eta\delta\cdot\omega})'\mathcal{V}_{\delta\cdot\omega}^{-1}(\mathcal{I}_{\delta\cdot\omega} - \mathcal{I}_{\delta\eta\cdot\omega}\mathcal{I}_{\eta\cdot\omega}^{-1}\mathcal{I}_{\eta\delta\cdot\omega})\zeta$, where $\mathcal{V}_{\delta\cdot\omega} := \mathcal{K}_{\delta\cdot\omega} + \mathcal{I}_{\delta\eta\cdot\omega}\mathcal{I}_{\eta\cdot\omega}^{-1}\mathcal{K}_{\eta\cdot\omega}\mathcal{I}_{\eta\cdot\omega}^{-1}\mathcal{I}_{\eta\delta\cdot\omega} - \mathcal{I}_{\delta\eta\cdot\omega}\mathcal{I}_{\eta\cdot\omega}^{-1}\mathcal{K}_{\eta\delta\cdot\omega} - \mathcal{K}_{\delta\eta\cdot\omega}\mathcal{I}_{\eta\cdot\omega}^{-1}\mathcal{I}_{\eta\delta\cdot\omega}$ is the adjusted variance, with $\mathcal{K}_{\delta\cdot\omega} := \mathcal{J}_{\delta\delta} + \mathcal{I}_{\delta\omega}\mathcal{I}_{\omega\omega}^{-1}\mathcal{J}_{\omega\omega}\mathcal{I}_{\omega\omega}^{-1}\mathcal{I}_{\omega\delta} - \mathcal{J}_{\delta\omega}\mathcal{I}_{\omega\omega}^{-1}\mathcal{I}_{\omega\delta} - \mathcal{I}_{\delta\omega}\mathcal{I}_{\omega\omega}^{-1}\mathcal{J}_{\omega\delta}$ and the variance-covariance matrix of a score function $\mathcal{J}_{\psi\psi} := \mathrm{E}\frac{1}{nT}\left(\frac{\partial \ln L_{nT}^c(\theta)}{\partial \psi}\frac{\partial \ln L_{nT}^c(\theta)}{\partial \psi'}\right)$. The proofs are provided in Appendix A.5.

# 4 Monte Carlo Simulations

In this section, the performance of the score test statistics are reported through Monte Carlo simulations with different cases of $(n, T)$ to investigate its finite sample properties in size and power. This experiments will get us sense in what empirical conditions the proposed test would work the best, regarding the sample size and the density of a network structure. In our simulations, suppose the dependent variable is weakly persistent, so that $\eta := (\lambda, \rho, \gamma)$ as discussed in Section 3.4.3 (i).

The type I error is set as 0.05, and the dimension of $Z$ is set as $p = 1$ for simplicity. The parameters for the main and auxiliary equations follow the setup in Qu, Lee, and Yu (2017): $\beta_0 = 1, \kappa_0 = 0.2, \Gamma_0 = 0.3, \alpha_0 = 1$, respectively. Data are generated by (1) and (2). The magnitude of local parametric misspecification in $\eta = (\lambda, \gamma, \rho)$ increases by 0.05 from 0 to 0.2, and $\delta_0$ is set from 0 to 0.1 or 0.2, which also increases by 0.05. In summary, we have

$$\begin{aligned}\theta_0 &= (\lambda_0, \gamma_0, \rho_0, \beta_0, \delta_0, \kappa_0, \Gamma_0, \sigma_{\xi 0}^2, \alpha_0)' \\ &= (\lambda_0, \gamma_0, \rho_0, 1, \delta_0, 0.2, 0.3, 1 - \delta_0^2, 1)'.\end{aligned}$$

The initial values of $Y_0, Z_0$ and the deterministic explanatory variables of $\mathbf{X}_{1,nT}, \mathbf{X}_{2,nT}$ are generated from independent standard normal distributions. The joint pdf of the disturbance terms, $(v_{it}, \epsilon_{it})$, follows bivariate normal distribution where the variances are set as 1 for both and the covariance is set as $\delta$.

The spatial weight matrix $W_{nt}$ is generated by the elementwise product of the physical contiguity and the inverse of absolute difference in the stochastic triggers: $W_{nt} = W_n^d \circ W_{nt}^e$, i.e., $w_{ij,nt} = w_{ij}^d \cdot w_{ij,nt}^e$

and is row-normalized afterward. Queen and Rook contiguities are considered for $W_n^d$, characterizing different adjacency: Queen allows interactions over edges & corners, whereas Rook only allows edges. Note that Queen and Rook contiguities are utilized to introduce sparsity in the network so that spatial weight matrices are in alignment with the spatial near-epoch dependence (NED) framework upon which this paper is founded. The economic distance $W_{nt}^e$ is generated by $w_{ij,nt}^e = 1/|z_{it} - z_{jt}|$ if $i \neq j$ or zero otherwise.

## 4.1 Simulation Results

The three scenarios with different types of sample size are considered: $(n, T) = (49, 50)$ as a default, $(n, T) = (100, 50)$ and $(n, T) = (49, 100)$ as the cases where $n$ and $T$ are increased by approximately twice high, respectively. The second and the third scenarios represent a larger network with same time length or the same network size with longer time length than the default scenario, respectively. Each experiment is replicated 1,000 times. The simulation results are mainly reported for $\lambda$, as it is the only parameter subject to the endogenous predictor (i.e., $W_{nt}Y_{nt}$).[11]

Figure 1 reports the size of the score test statistics, highlighting the favorable finite sample properties of my proposed test. Notably, the sizes of the unadjusted score tests (i.e., $RS_\delta^{B,P}$ and $RS_\delta^P$) increase significantly as the magnitude of local parametric misspecification in $\lambda$ grows. In contrast, the size of the biased score test (i.e., $RS_\delta^B$) starts either higher or lower than the true size when the local misspecification in $\lambda$ is very small (ranging from 0 to 0.1). However, the size of my proposed test (i.e., $RS_\delta^*(\tilde{\theta})$) remains relatively stable, approximately up to 0.25 for the Queen contiguity and 0.2 for the Rook contiguity.

Figure 2 reports the power of the score test statistics, which also demonstrates the favorable finite sample properties of my proposed test. While the power of $RS_\delta^{B,P}(\tilde{\theta})$ and $RS_\delta^P(\tilde{\theta})$ appears to increase merely due to the presence of local parametric misspecification, the power of $RS_\delta^*(\tilde{\theta})$ stays comparatively stable regardless of the magnitude of the local parametric misspecifications in $\lambda$.

**Figure 1 is here**

**Figure 2 is here**

**Table 3 is here**

[11]The results of the other parameters of the time or spatial dynamic coefficients ($\gamma$, $\rho$) are reported in Online Appendix OA.1.

**Table 4 is here**

Meanwhile, the error term may deviate from normality when the sample size is not large enough or outliers or extreme values are present. In such cases, using a $t$-distribution for the error terms is often more realistic. Thus, it is worthwhile to explore the $t$-distribution to determine if $RS^*_\delta(\tilde{\theta})$ can better handle extreme values and provide more results in small or moderate sample scenarios.

To this end, the robustness of $RS^*_\delta(\tilde{\theta})$ against non-normality, such as when the error terms follow a bivariate t-distribution, is explored. The results reveal that as the degrees of freedom increase, $RS^*_\delta(\tilde{\theta})$ quickly retrieves its true size (Figure 3 and Table 5). Moreover, its power under non-normal errors also recovers to match that with normal errors as the degrees of freedom increase (Figure 4 and Table 6), demonstrating its robustness and reliability.

**Figure 3 is here**

**Figure 4 is here**

**Table 5 is here**

**Table 6 is here**

Taken together, these simulation results highlight the limitations of biased or unadjusted score tests in certain situations. In contrast, my proposed test consistently offers asymptotically valid inference, irrespective of the extent of local misspecifications. Furthermore, while it performs effectively with a moderate number of cross-sectional units, its performance improves as $(n, T)$ increase (Tables 3 and 4). Additionally, my proposed test demonstrates robustness in scenarios where the normality of the error term is violated, possibly due to small sample sizes or the presence of outliers or extreme values. For instance, even if the error term follows a $t$-distribution, my proposed test still provides valid inference as the degrees of freedom increase, highlighting its practical utility.

## 4.2 Comparison with the $C(\alpha)$ Test

In this subsection, the performances of my proposed test (i.e., $RS^*_\delta(\tilde{\theta})$) and the $C(\alpha)$ test evaluated at the ML estimates under $H_0^\delta$ are compared. While both tests have similar purposes and equivalent asymptotic distributions (i.e., central chi-squared under the null), they differ in their evaluation within spatial models. The $C(\alpha)|_{\hat{\theta}}$ test requires the estimation of a spatial model along with the VAR process (i.e., $\hat{\theta} = \underset{\theta:\, \delta=0}{\operatorname{argmax}} \ln L^c_{nT}(\theta)$), whereas $RS^*_\delta(\tilde{\theta})$ requires estimates from a non-spatial dynamic panel data with two-way fixed effects model along with the VAR process (i.e., $\tilde{\theta} = \underset{\theta:\, \delta=0,\, \eta=0}{\operatorname{argmax}} \ln L^c_{nT}(\theta)$).

In finite samples, the two tests provide nearly identical and valid inferences for size, with minimal differences (Figure 5 and Table 7), regardless of the magnitude of local parametric misspecification in $\lambda$. This equivalence in inference is especially apparent with the Queen contiguity. Regarding power, both tests exhibit similar results of rejection rates of about 0.7 with $(n, T) = (49, 50)$ (Figure 5). The power sharply increases to 0.93 to 0.95 with $(n, T) = (49, 100)$ and $(n, T) = (100, 50)$, respectively (Table 8). As shown in Figure 5, there is little difference between the two tests when $\delta$ slightly deviates from 0, particularly at 0.05, demonstrating good power properties. Overall, these findings indicate that both tests offer nearly equivalent and valid inferences.

**Figure 5 is here**

**Table 7 is here**

**Table 8 is here**

However, the difference in elapsed time between the two tests becomes significant as $n$ or $T$ increases (Figure 6 and Table 9), nearly forming an exponential function. These results imply that my proposed test can obtain equivalent inferences to the $C(\alpha)$ test with lower computational cost by using only a non-spatial model. Since the inclusion of spatial autoregressive terms significantly increases computational expenses with the growth of cross-sectional units, this approach substantially reduces the resources required for valid inference. Therefore, using my proposed test is not only valid but economical compared to other testing options that require estimates from spatial models.

**Figure 6 is here**

**Table 9 is here**

# 5 Empirical Illustration

This section describes a research episode that demonstrates how my proposed test aids in model selection at the early stages of research. This ensures the use of a valid and computationally efficient spatial model before proceeding to the estimation phase. In the subsequent estimation phase, the parameters of the selected model are estimated, and the implications regarding spillover effects are discussed.

Consider the neoclassical growth model (Solow 1956; Swan 1956), which outlines the dynamics of economic growth at a steady state by labor, capital, and technology as

$$Y_{it} = A_{it} K_{it}^{a} L_{it}^{1-a}, \tag{9}$$

where $Y_{it}$ is the total output, $K_{it}$ is the level of reproducible physical capital, and $L_{it}$ is the level of labor. This model implies that the total output is decomposed by the two factors of the physical capital and the level of labor with a working-age population in the form of a Cobb-Douglas production function. All the other factors not explained by physical capital or level of labor are measured by the residuals and understood as the aggregate technology or total factor productivity.

Addressing spatial dependence across economies, a spatially augmented growth model is proposed (Ertur and Koch 2007), which models a country's technology to be interdependent with that of another country in the neighborhood, weighted by the geographic contiguities as

$$A_{it} = \Omega_t k_{it}^{b} \prod_{j \neq i}^{n} A_{jt}^{c w_{ij,t}}, \tag{10}$$

where $\Omega_t = \Omega_0 \exp(\mu t)$ represents some proportion of technological progress as exogenous and identical to all countries, with its constant rate of growth $\mu$; $k_{it} = K_{it}/L_{it}$ is the accumulated physical capital per capita; $b$ $(0 \leq b < 1)$ denotes the connectivity of physical capital externalities or home externalities with respect to knowledge from capital investment; $c$ $(0 \leq c < 1)$ represents the technological interdependence across countries or neighbors. Finally, $w_{ij,t}$ represents the spatial dependence between countries $i$ and $j$ at time $t$, where $0 \leq w_{ij,t} < 1$, $w_{ij,t} = 0$ if $i = j$, and row-normalized such that $\sum_{j \neq i}^{n} w_{ij,t} = 1$ for $i = 1, \ldots, n$. Usually, $w_{ij,t}$ is supposed to be generated by geographic proximity.

However, interactions across economies may additionally be stimulated by *knowledge* distance, where countries with similar knowledge levels interact more easily because they can quickly understand, integrate, and build upon each other's innovations. This compatibility reduces the learning curve and translation efforts, leading to faster and more efficient collaboration. Countries with similar knowledge levels face comparable challenges and opportunities, making shared solutions more relevant. Such interactions create a synergistic environment for exchanging and improving innovations, driving mutual economic growth, and fostering trust and reliability, which encourage sustained collaboration and resource pooling for large-scale projects.

Suppose knowledge is proxied by each country's total trade volumes, calculated as the sum of exports and imports because high trade volumes reflect a country's economic activity and engagement in international markets, indicating technological sophistication and managerial expertise. Additionally, significant trade volumes suggest a competitive economy investing in research and development, closely linked to higher knowledge levels. Therefore, total trade volume effectively encapsulates the elements of economic activity, technological exchange, and human capital that constitute a country's knowledge base.

Now combining (9) and (10) with extension to SDPD model, where a spatial weight matrix might be endogenous (Qu, Lee, and Yu 2017), yields a spatially augmented growth model in SDPD as

$$\begin{aligned}
\ln\left[\frac{Y_{nt}}{L_{nt}}\right] &= \lambda_y W_{nt}\ln\left[\frac{Y_{nt}}{L_{nt}}\right] + \rho W_{n,t-1}\ln\left[\frac{Y_{n,t-1}}{L_{n,t-1}}\right] + \gamma\ln\left[\frac{Y_{n,t-1}}{L_{n,t-1}}\right] \\
&\quad + X_{1nt}\beta + W_{nt}X_{1nt}\lambda_x + c_{n1} + \tau_{t1}1_n + V_{nt}, \\
(W_{nt})_{ij} &= w_{ij,nt} = w_{ij}^d \cdot w_{ij,nt}^e, \\
\ln Z_{nt} &= \ln Z_{n,t-1}\kappa + X_{2nt}\Gamma + c_{n2} + 1_n\tau_{t2}' + \varepsilon_{nt}, \quad t = 1,\ldots,T,
\end{aligned} \tag{11}$$

where $Y_{nt}/L_{nt}$ is the real income per worker at time $t$; $X_{1nt} = (\ln s_{nt},\ \ln(n_{nt}+g+d))'$ with $s_{it}$ being a constant fraction of output saved for country $i = 1,\ldots,n$; $n_{it}$ the exogenous labor growth rate for country $i$; $d$ the annual rate of depreciation of physical capital for all countries and assumed constant; and $g$ being the balanced growth rate; $X_{2nt} = (\ln(Y_{n,t-1}/L_{n,t-1}),\ \ln s_{nt},\ \ln(n_{nt}+g+d))'$ overlaps with $X_{1nt}$ but additionally includes $\ln(Y_{n,t-1}/L_{n,t-1})$ to account for sequential spillovers in $\ln(Y_{nt}/L_{nt})$; $w_{ij}^d$ indicates the geographic proximity, which is predetermined, measured by the great-circle distance; $w_{ij,nt}^e$ indicates the inverse of the absolute difference in non-predetermined variables, which are trade in this application; $Z_{nt}$ is the *total* trade volume as the sum of exports and imports, which measures the knowledge level of an economy; and $c_{nj}$ and $\tau_{nj}$ are the individual and time fixed effects, respectively for $j = 1, 2$. The spatially lagged terms of $W_{nt}\ln(Y_{nt}/L_{nt})$ and $W_{nt}X_{1nt}$ represent the spatial dependence due to the concurrent spillover effects through $W_{nt}$, whether related to the dependent variable or the explanatory variables. Meanwhile, $W_{n,t-1}\ln(Y_{n,t-1}/L_{n,t-1})$ represents the spatial dependence due to the past spillover effects with respect to the dependent variable. Using a control function approach, (11) can be augmented as

$$\begin{aligned}\ln\left[\frac{Y_{nt}}{L_{nt}}\right] = {} & \lambda_y W_{nt}\ln\left[\frac{Y_{n,t-1}}{L_{n,t-1}}\right] + \rho W_{n,t-1}\ln\left[\frac{Y_{n,t-1}}{L_{n,t-1}}\right] + \gamma\ln\left[\frac{Y_{n,t-1}}{L_{n,t-1}}\right] \\ & + X_{1nt}\beta + W_{nt}X_{1nt}\lambda_x + c_{n1} + \tau_{t1}1_n \\ & + (\ln Z_{nt} - \ln Z_{n,t-1}\kappa - X_{2nt}\Gamma - c_{n2} - 1_n\tau_{t2}')\delta + \xi_{nt},\end{aligned}$$

where $\xi_{nt} \sim N(0, \sigma_\xi^2 I_n)$.

The data used in this application are from the Penn World Table version 7.1, with the currency unit being 2005 international dollars (I\$). Suppose we are interested in investigating evidence of spatial dependencies provoked by geographic and knowledge proximities for the years 1950-1990 for 55 countries[12]. To address issues arising from unknown initial conditions when $T(=41)$ is less than $n(=55)$, we predict the initial conditions using the JIVE method, as discussed in Section 3.4.2. The variables are measured similarly to those in Ertur and Koch (2007), except for the non-predetermined trigger of total trade volume and the construction of a spatial weight matrix.

The *total* trade volume is used to measure the knowledge level of an economy, because the knowledge in an economy amplifies with the scale of its trade activities. This approach captures the aggregate impact of trade on knowledge advancement, which is better reflected in the total trade volume rather than per capita figures. To recover the total trade volume as the sum of exports and imports, from the openness[13] defined as the sum of exports and imports divided by Purchasing Power Parity (PPP) converted GDP per capita (Laspeyres) (measured in 2005 International dollar per person)[14], the openness is multiplied by GDP per capita (Laspeyres) and by the total population. Notably, the equation for $Z_{nt}$ indicates that the knowledge level of an economy, proxied by total trade volume, is influenced by the outcome in the previous period. This implies the sequential spillover effects in the outcome itself, as discussed in Section 3.4.3 Some Extensions.

For constructing the spatial weight matrix, geographic distance is measured by the great-circle distance between capitals, and trade distance is measured by the absolute difference in the total trade volumes. The element-wise products of the geographic and trade distances are then generated, with each distance scaled from 0 to 1 to normalize different units. Subsequently, $W_{nt}$ is generated as the inverse of the composite distance, assigning higher weights to closer composite distances. To

[12] The list of 55 countries is provided in Appendix A.7.
[13] *OPENK* in Penn World Table version 7.1.
[14] *RGDPL* in Penn World Table version 7.1.

meet the sparsity required by the near epoch dependence properties, weights below the 80% quantile are truncated, and $W_{nt}$ is row-normalized.

Now, to determine the parameters ($\eta$) that are susceptible to local misspecifications among the spatial or time dynamic coefficients, we examine the autocorrelation function (ACF) plots for the time series in real income per worker for each country (Figures in Online Appendix OA.2). In general, most countries exhibit highly persistent autocorrelation for one-time lagged values. Even for countries with smaller autocorrelation for one-time lagged values, there is still moderate (0.4 to 0.6) to strong persistence (0.6 or above) (Figure 7). Accordingly, this suggests that the time dynamic coefficient ($\gamma$) should not be included as part of $\eta$, as our dependent variable is observed to be not close to zero. Therefore, we only consider spatial dependence parameters as possibly being locally misspecified (i.e., $\eta := (\lambda, \rho)$), leveraging the properties of weak ties in large-scale networks.

**Figure 7 is here**

Now suppose we want to determine which spatial model to use for valid analysis: the conventional SDPD model with an exogenous spatial weight matrix (Lee and Yu, 2012) or the SDPD model addressing an endogenous spatial weight matrix (Qu, Lee, and Yu, 2017), which has a higher computational burden. As long as our $W_{nt}$ is exogenous with respect to our dependent variable, we may use the conventional model for valid inference. However, if this is not the case, we need to employ the SDPD model addressing an endogenous spatial weight matrix for valid inference. Therefore, at this stage, we need to determine whether $W_{nt}$ is endogenous.

A set of hypothesis tests can be conducted to test the endogeneity of $W_{nt}$. However, these tests involve spatial parameters even under the null hypothesis, resulting in greater time requirements compared to non-spatial models. However, leveraging the weak ties properties in large-scale networks, my proposed test gives valid inference comparable to the $C(\alpha)$ test but is computationally efficient since it only requires estimates from non-spatial models. Additionally, my proposed test is equivalent to $C(\alpha)$ test, demonstrating its asymptotic optimality.

Thus, we utilize my proposed test to determine the endogeneity of $W_{nt}$, with the results reported in Table 10. The test results indicate that my proposed robust score test yields a value of 31.209, which exceeds the critical value of $\chi_1^2 = 3.8415$ at the 5% significance level. This leads to the rejection of $H_0^\delta :\ \delta = 0$, suggesting that the spatial weight matrix, which captures spillover effects driven by a combination of geographic and knowledge proximities (with knowledge proxied by total trade volume), is statistically endogenous with respect to real income per worker. It is worth noting that the $C(\alpha)$ test statistic, evaluated at the maximum likelihood estimates under $H_0^\delta$, provides inferences equivalent to those of my proposed test.

**Table 10 is here**

Given the statistical evidence that $W_{nt}$ is endogenous, we select the SDPD model with an endogenous spatial weight matrix for valid inference. The estimation results are reported in Table 11. As detected by the proposed test earlier, our testing parameter $\delta$ is significant, indicating that $W_{nt}$ is endogenous. Notably, the spatial dependence parameters are found to be weak and significant, aligning with our postulation on weak ties in large-scale networks. As observed in the ACF plots, the time dynamic coefficient is significant and moderate.

**Table 11 is here**

Interestingly, all explanatory variables on real income per worker are found to be positive except for the negative spillover effects in saving rates (i.e., $W_{nt} \ln s_{nt}$). This may indicate a complex interplay between internal investment and external competition. Higher saving rates within an economy typically lead to greater capital accumulation, boosting productivity and real income per worker through investments in infrastructure, technology, and education. However, high saving rates in neighboring economies can increase competition for investment and resources, potentially making it more difficult for the original economy to attract foreign investments. Additionally, if neighboring economies are investing heavily in their productivity, this could create competitive pressures, negatively impacting the original economy's industries and real income per worker. Thus, while internal savings foster growth, external savings might introduce competitive disadvantages.

Notably, labor growth rate (i.e., $n_{nt}$) has a positive effect on real income per worker, contrasting with the general findings of the conventional Solow-Swan growth model. In the context of knowledge accumulation, this can be attributed to the fact that a larger labor population often leads to a greater accumulation of knowledge within an economy. This is because a higher labor population size can foster more diverse ideas, innovations, and skills, enhancing the overall knowledge base. The positive correlation between labor population size and knowledge level means that labor population growth facilitates the dissemination and utilization of knowledge, driving technological advancements and productivity improvements. Consequently, these factors collectively contribute to economic improvement and an increase in real income per worker.

Another interesting finding is in the auxiliary equation for the knowledge of an economy, proxied by total trade volume. Notably, the real income per worker in the previous period (i.e., $\ln(Y_{n,t-1}/L_{n,t-1})$)

has positive effects on the current knowledge level of an economy. This indicates that higher income levels lead to increased trade activities, which in turn enhance the economy's knowledge base. The increased income allows for more investments in education, research, and technology, which boosts the overall knowledge within the economy. This process creates a virtuous cycle where higher real income per worker promotes greater knowledge accumulation, which then feeds back into further income growth. This dynamic demonstrates sequential spillover effects, where improvements in real income per worker not only benefit the current period but also lay the groundwork for future economic enhancements through elevated knowledge levels.

In conclusion, these findings underscore the nuanced and multifaceted nature of economic growth. The positive internal effects of savings and population growth highlight the importance of domestic policies that encourage investment in human capital and infrastructure. Meanwhile, the negative external spillover effects from neighboring economies' savings emphasize the need for strategic international economic policies and collaborations. We thus conclude that there is compelling evidence of weak spillover effects in real income per worker, formed by the composite proximities of geographic and knowledge levels of economies. These insights provide a valuable framework for policymakers aiming to enhance economic resilience and foster sustainable growth.

# 6 Conclusion

A non-predetermined spillover framework is exceedingly versatile and broadly applicable in capturing the comprehensive effects of economic interactions because it allows for the incorporation of dynamic and context-specific variables. Traditionally, a spatial weight matrix ($W$) has been primarily employed in the context of predetermined geography. However, $W$ can be extended to accommodate interactions influenced by non-predetermined variables that trigger various types of spillovers.

However, there is a caveat: the non-predetermined spillover framework may result in an endogenous $W$, where the disturbances between the outcome and the non-predetermined triggers are correlated. If $W$ is exogenous with respect to the outcome, one can utilize a SDPD model with an exogenous spatial weight matrix (Lee and Yu 2012) for valid inferences. But if $W$ is endogenous with respect to the outcome, the endogeneity of $W$ must be addressed, necessitating the use of a SDPD model with an endogenous spatial weight matrix (Qu, Lee, and Yu 2017).

Therefore, testing the endogeneity of $W$ before proceeding with full estimation is strongly encouraged to ensure the selection of a valid and efficient model. To this end, I have developed a robust score test to ascertain the endogeneity of $W$ in the weak-tied SDPD model, where the spatial dependence parameters are presumed to be close to zero or exhibit local deviations. My proposed test is advantageous in terms of asymptotic optimality compared to other tests, as it is equivalent to the $C(\alpha)$ test. Notably, it is computationally more efficient than other tests that require estimates from spatial models, as it only requires estimates from non-spatial models. This efficiency is achieved by leveraging the weak ties properties of large-scale networks, where spatial dependence parameters can be asymptotically interpreted to have local alternatives, allowing them to drift towards zero as the network size increases.

Future research can explore several interesting extensions. First, developing tests that are robust against Information Inequality, possibly due to the unknown initial conditions problem or non-normality of errors, is a valuable direction. Second, analytical bias correction for the unknown initial conditions problem is worthwhile. While the large $T$ setup in this paper may mitigate concerns related to initial conditions (Qu, Lee, and Yu 2017), it might still lead to substantial bias unless $T$ is significantly larger than $n$ (Nickell 1981). Additionally, given that many real-world datasets feature short panels, it may be valuable to investigate analytical bias correction for unknown initial conditions in short panel models. This approach could be implemented by extending Yang (2018)'s method to accommodate an endogenous spatial weight matrix.

## Acknowledgements

The author is sincerely grateful for the invaluable feedback from Esfandiar Maasoumi, the former editor-in-chief, Yuya Sasaki, the current editor-in-chief, and the four anonymous referees. Their insightful comments have greatly strengthened this paper.

The author is also deeply grateful to Osman Dogan (Istanbul Technical University), Xiye Yang, Roger Klein (Rutgers University, New Brunswick), Hanbat Jeong (Ohio State University), Xu Lin (Virginia Tech), Anil Bera, EunYi Chung, Bo Li, Xiaofeng Shao (University of Illinois Urbana-Champaign), Kyungchul Kevin Song (University of British Columbia), Italo Santos, and Hongqi Chen (University of Illinois Urbana-Champaign) for their thorough and thoughtful suggestions.

The author is also thankful for the valuable input from conference and seminar attendees, including Heejin Kim, Hongqi Chen, Italo Santos, Mark Daniel Bernhardt, Humberto Martinez-Garcia, Malabika Koley, Wilfred Adohinzin, Eduardo Medina-Cortina, Cihang Wang, Abby Stocker, Cristhian Molina Gonzalez, Flavio Rodrigues, Vinicius Okada Da Silva, Ethan Dee (University of Illinois Urbana-Champaign), Ryan Kim (Johns Hopkins University), Jieun Lee (University of Colorado Boulder), Minjae Yun (University of California, Los Angeles School of Law), Jihyun Oh (Georgetown University), Hyunji Ahn (Rutgers University, New Brunswick), Ahnjeong Kim (North Carolina State University), Nayeon Lim (Boston University), Michael Delgado, Diego Cardoso (Purdue University), Roger Klein, John Landon-Lane, Norman Swanson, Bruce Mizrach, Yuan Liao, Xiye Yang, Ruonan Xu, Rosemary Kaiser, and Carlos Esquivel (Rutgers University, New Brunswick), and Hanbat Jeong (Ohio State University) for their detailed and constructive comments.

## Declaration of Interest Statement

The author declares no conflicts of interest.

## Author Contributions Statement

The author made all contributions to this paper.

## Data Availability

All data and codes are available upon request.

# Reference


Acs, Z. J., Braunerhjelm, P., Audretsch, D. B., & Carlsson, B. (2009). The Knowledge Spillover Theory of Entrepreneurship. *Small Business Economics*, 32, 15-30.

Aghion, P., Bergeaud, A., Gigout, T., Lequien, M., & Melitz, M. (2019). Spreading Knowledge Across the World: Innovation Spillover Through Trade Expansion. Manuscript, Harvard University.

Ali, M. M., & Dwyer, D. S. (2009). Estimating Peer Effects in Adolescent Smoking Behavior: A Longitudinal Analysis. *Journal of Adolescent Health*, 45(4), 402-408.

Amidi, S., Fagheh Majidi, A., & Javaheri, B. (2020). Growth Spillover: A Spatial Dynamic Panel Data and Spatial Cross-Section Data Approaches in Selected Asian Countries. *Future Business Journal*, 6(1), 1-14.

Angrist, J. D., Imbens, G. W., & Krueger, A. B. (1999). Jackknife Instrumental Variables Estimation. *Journal of Applied Econometrics*, 14(1), 57-67.

Angrist, J. D. (2014). The Perils of Peer Effects. *Labour Economics*, 30, 98-108.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Springer, New York.

Anselin, L. (2001). Rao's Score Test in Spatial Econometrics. *Journal of Statistical Planning and Inference*, 97(1), 113-139.

Baltagi, B., & Li, D. (2001). LM Tests for Functional Form and Spatial Error Correlation. *International Regional Science Review*, 24(2), 194–225.

Baltagi, B., Song, S., & Koh, W. (2003). Testing Panel Data Regression Models with Spatial Error Correlation. *Journal of Econometrics*, 117(1), 123–150.

Baltagi, B., Song, S., Jung, B., & Koh, W. (2007). Using Panel Data, Testing for Serial Correlation, Spatial Autocorrelation, and Random Effects. *Journal of Econometrics*, 140(1), 5–51.

Baltagi, B., & Liu, L. (2008). Testing for Random Effects and Spatial Lag Dependence in Panel Data Models. *Statistics & Probability Letters*, 78(18), 3304–3306.

Baltagi, B., Song, S., & Kwon, J. (2009). Testing for Heteroskedasticity and Spatial Correlation in a Random Effects Panel Data Model. *Computational Statistics & Data Analysis*, 53(8), 2897–2922.

Baltagi, B., & Yang, Z. (2013). Standardized LM Tests for Spatial Error Dependence in Linear or Panel Regressions. *The Economic Journal*, 16(1), 103–134.

Bera, A., & Yoon, M. (1993). Specification Testing with Locally Misspecified Alternatives. *Econometric Theory*, 9(4).

Bera, A., Doğan, O., & Taşpınar, S. (2018a). Simple Tests for the Endogeneity of Spatial Weight Matrices. *Regional Science and Urban Economics*, 69, 130-142.

Bera, A., Doğan, O., & Taşpınar, S. (2018b). Testing Spatial Dependence in Spatial Models with Endogenous Weight Matrices. *Journal of Econometric Methods*, 8(1), 1-33.

Bera, A., Doğan, O., Taşpınar, S., & Leiluo, Y. (2019). Robust LM Tests for Spatial Dynamic Panel Data Models. *Regional Science and Urban Economics*, 76, 47-66.

Bera, A., Bilias, Y., Yoon, M., Taşpınar, S., & Doğan, O. (2020). Adjustments of Rao's Score Test for Distributional and Local Parametric Misspecifications. *Journal of Econometric Methods*, 9(1), 20170022.

Billé, A. G., & Rogna, M. (2022). The Effect of Weather Conditions on Fertilizer Applications: A Spatial Dynamic Panel Data Analysis. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 185(1), 3-36.

Blomquist, S., & Dahlberg, M. (1999). Small Sample Properties of LIML and Jackknife IV Estimators: Experiments with Weak Instruments. *Journal of Applied Econometrics*, 14(1), 69-88.

Bonnal, M. (2010). Economic Growth and Labor Standards: Evidence from a Dynamic Panel Data Model. Review of Development Economics, 14(1), 20-33.

Brida, J. G., & Risso, W. (2009). Tourism as a Factor of Long-Run Economic Growth: An Empirical Analysis for Chile. *European Journal of Tourism Research*, 2(2), 178-185.

Brown, D. W., & Konrad, A. M. (2001). Granovetter Was Right: The Importance of Weak Ties to a Contemporary Job Search. *Group & Organization Management*, 26(4), 434-462.

Bursztyn, L., Ederer, F., Ferman, B., & Yuchtman, N. (2014). Understanding Mechanisms Underlying Peer Effects: Evidence from a Field Experiment on Financial Decisions. *Econometrica*, 82(4), 1273-1301.

Card, D., & Giuliano, L. (2013). Peer Effects and Multiple Equilibria in the Risky Behavior of Friends. *Review of Economics and Statistics*, 95(4), 1130-1149.

Cheng, W., & Lee, L. F. (2017). Testing Endogeneity of Spatial and Social Networks. *Regional Science and Urban Economics*, 64, 81-97.

Cliff, Andrew., & Ord, Keith (1972). Testing for Spatial Autocorrelation Among Regression Residuals. *Geographical Analysis*, 4(3), 267-284.

Conley, Timothy G., Ligon, Ethan. (2002). Economic Distance and Cross-Country Spillovers. *Journal of Economic Growth*, 7(2), 157-187.

Conley, Timothy G., Topa, Giorgio. (2002). Socio-Economic Distance and Spatial Patterns in Unemployment. *Journal of Applied Econometrics*, 17(4), 303-327.

Cornelissen, T., Dustmann, C., & Schönberg, U. (2017). Peer Effects in the Workplace. *American Economic Review*, 107(2), 425-456.

Davidson, R., & Mackinnon, J. (1987). Implicit Alternatives and the Local Power of Test Statistics. *Econometrica*, 55(6).

Debarsy, N., & Ertur, C. (2010). Testing for Spatial Autocorrelation in a Fixed Effects Panel Data Model. *Regional Science and Urban Economics*, 40(6), 453–470.

De Long, J. B., & Summers, L. H. (1991). Equipment Investment and Economic Growth. *The Quarterly Journal of Economics*, 106(2), 445-502.

Doğan, O., Taşpınar, S., & Bera, A. (2018). Simple Tests for Social Interaction Models with Network Structures. *Spatial Economic Analysis*, 13(2), 212-246.

Elhorst, J. (2010). Applied Spatial Econometrics: Raising the Bar. *Spatial Economic Analysis*, 5(1), 9-28.

Elhorst, J. (2014). *Spatial Econometrics: from Cross-sectional Data to Spatial Panels*. Springer Briefs in Regional Science. Springer Berlin Heidelberg, New York.

Ertur, C. & Koch, W. (2007). Growth, Technological Interdependence, and Spatial Externalities: Theory and Evidence. *Journal of Applied Econometrics*, 22(6), 1033-1062.

Falk, A., & Ichino, A. (2006). Clean Evidence on Peer Effects. *Journal of Labor Economics*, 24(1), 39-57.

Fowler, J. H., & Christakis, N. A. (2008). Estimating Peer Effects on Health in Social Networks: A Response to Cohen-Cole and Fletcher; Trogdon, Nonnemaker, Pais. *Journal of Health Economics*, 27(5), 1400.

Gaviria, A., & Raphael, S. (2001). School-Based Peer Effects and Juvenile Behavior. *Review of Economics and Statistics*, 83(2), 257-268.

Graham, B. S. (2017). An Econometric Model of Network Formation with Degree Heterogeneity. *Econometrica*, 85(4), 1033-1063.

Granovetter, M. S. (1973). The Strength of Weak Ties. *American Journal of Sociology*, 78(6), 1360-1380.

Greenstone, M., Hornbeck, R., & Moretti, E. (2010). Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings. *Journal of Political Economy*, 118(3), 536-598.

Handcock, M. S., Raftery, A. E., & Tantrum, J. M. (2007). Model-Based Clustering for Social Networks. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 170(2), 301-354.

Heston, A., Summers, R., & Aten, Bettina. (2012). Penn World Table Version 7.1 Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania.

Ho, C., Wang, W., & Yu, J. (2013). Growth Spillover Through Trade: A Spatial Dynamic Panel Data Approach. *Economics Letters*, 120(3), 450–453.

Hoff, P. D., Raftery, A. E., & Handcock, M. S. (2002). Latent Space Approaches to Social Network Analysis. *Journal of the American Statistical Association*, 97(460), 1090-1098.

Jenish, N., & Prucha, I. (2009). Central Limit Theorems and Uniform Laws of Large Numbers for Arrays of Random Fields. *Journal of Econometrics*, 150(1), 86-98.

Jenish, N., & Prucha, I. (2012). On Spatial Processes and Asymptotic Inference Under Near-Epoch Dependence. *Journal of Econometrics*, 170(1), 178-190.

Jeong, H., & Lee, L. F. (2021). Spatial Dynamic Game Models for Coevolution of Intertemporal Economic Decision-Making and Spatial Networks. *Journal of Economic Dynamics and Control*, 129, 104186.

Jochmans, K. (2018). Semiparametric Analysis of Network Formation. *Journal of Business & Economic Statistics*, 36(4), 705-713.

Kelejian, H. H., & Piras, G. (2014). Estimation of Spatial Models with Endogenous Weighting Matrices, and an Application to a Demand Model for Cigarettes. *Regional Science and Urban Economics*, 46, 140-149.

Kelejian, H., & Robinson, D. (1992). Spatial Autocorrelation: A New Computationally Simple Test with an Application to Per-Capita County Police Expenditures. *Regional Science and Urban Economics*, 22(3), 317–331 Special Issue Space and Applied Econometrics.

Kelejian, H., & Prucha, I. (2010). Specification and Estimation of Spatial Autoregressive Models with Autoregressive and Heteroskedastic Disturbances. *Journal of Econometrics*, 157, 53-67.

Korniotis, G. M. (2010). Estimating Panel Models with Internal and External Habit Formation. *Journal of Business & Economic Statistics*, 28(1), 145-158.

Lee, L., & Yu, J. (2010). A Spatial Dynamic Panel Data Model with Both Time and Individual Fixed Effects. *Econometric Theory*, 26(2), 564-597.

Lee, L., & Yu, J. (2012). QML Estimation of Spatial Dynamic Panel Data Models with Time-Varying Spatial Weight Matrices. *Spatial Economic Analysis*, 7(1), 31-74.

Lin, X. (2010). Identifying Peer Effects in Student Academic Achievement by Spatial Autoregressive Models with Group Unobservables. *Journal of Labor Economics*, 28(4), 825-860.

Lin, X. (2014). Peer Effects in Adolescents' Delinquent Behaviors: Evidence from a Binary Choice Network Model. *Regional Science and Urban Economics*, 46, 73-92.

Lucas, R. (1990). Why Doesn't Capital Flow from Rich to Poor Countries?*American Economic Review*, 80(2), 92-96.

Mankiw, N., Romer, D., & Weil, D. (1992). A Contribution to the Empirics of Economic Growth. *Quarterly Journal of Economics*, 107, 407-437.

Moran, PAP. (1950). A Test for the Serial Independence of Residuals. *Biometrika*, 37(1/2), 178-181.

Moreno, R., & Trehan, B. (1997). Location and the Growth of Nations. *Journal of Economic Growth*, 2(4), 399-418.

Moretti, E. (2011). Social Learning and Peer Effects in Consumption: Evidence from Movie Sales. *The Review of Economic Studies*, 78(1), 356-393.

Neyman, J. (1937). Smooth Test for Goodness of Fit. *Scandinavian Actuarial Journal*, 1937(3-4), 149-199.

Neyman, J., & Scott, E. L. (1948). Consistent Estimates Based on Partially Consistent Observations. *Econometrica*, 1-32.

Neyman, J. (1959). Optimal Asymptotic Tests of Composite Hypotheses. *Probability and Statistics*, 213-234.

Nickell, S. (1981). Biases in Dynamic Models with Fixed Effects. *Econometrica*, 1417-1426.

Patacchini, E., & Zenou, Y. (2008). The Strength of Weak Ties in Crime. *European Economic Review*, 52(2), 209-236.

Phillips, G. D. A., & Hale, C. (1977). The Bias of Instrumental Variable Estimators of Simultaneous Equation Systems. *International Economic Review*, 219-228.

Pinkse, J., & Slade, M. (2010). The Future of Spatial Econometrics. *Journal of Regional Science*, 50(1), 103-117.

Qu, X., & Lee, L. (2015). Estimating a Spatial Autoregressive Model with an Endogenous Spatial Weight Matrix. *Journal of Econometrics*, 184(2), 209-232.

Qu, X., Lee, L. F., & Yang, C. (2021). Estimation of a SAR Model with Endogenous Spatial Weights Constructed by Bilateral Variables. *Journal of Econometrics*, 221(1), 180-197.

Qu, X., Lee, L., & Yu, J. (2017). QML Estimation of Spatial Dynamic Panel Data Models with Endogenous Time-Varying Spatial Weight Matrices. *Journal of Econometrics*, 197(2), 173-201.

Roemer, J. E. (1977). The Effect of Sphere of Influence and Economic Distance on the Commodity Composition of Trade in Manufactures. *The Review of Economics and Statistics*, 318-327.

Romer, P. (1989). Capital Accumulation in the Theory of Long-Run Growth. *Modern Business Cycle Theory*, Barro RJ (ed.). Harvard University Press: Cambridge, MA; 51-127.

Romer, P. M. (1990). Endogenous Technological Change. *Journal of Political Economy*, 98(5, Part 2), 71-102.

Saikkonen, P. (1989). Asymptotic Relative Efficiency of the Classical Test Statistics Under Misspecification. *Journal of Economics*, 42(3), 351-369.

Seetaram, N., & Dwyer, L. (2009). Immigration and Tourism Demand in Australia: A Panel Data Analysis. *Anatolia*, 20(1), 212-222.

Solow, R. (1956). A Contribution to the Theory of Economic Growth. *The Quarterly Journal of Economics*, 70(1), 65-94.

Swan, T. (1956). Economic Growth and Capital Accumulation. *Economic Record*, 32(2), 334–361.

Taşpınar, S., Doğan, O., & Bera, A. (2017). GMM Gradient Tests for Spatial Dynamic Panel Data Models. *Regional Science and Urban Economics*, 65, 65-88.

Tokunaga, S., Okiyama, M., & Ikegawa, M. (2017). Impact of Climate Change on Regional Economies Through Fluctuations in Japan's Rice Production: Using Dynamic Panel Data and Spatial CGE Models. *Socioeconomic Environmental Policies and Evaluations in Regional Science: Essays in Honor of Yoshiro Higano*, 557-580.

White, H. (1982). Maximum Likelihood Estimation of Misspecified Models. *Econometrica*, 1-25.

Yang, Z. (2010). A Robust LM Test for Spatial Error Components. *Regional Science and Urban Economics*, 40(5), 299–310 Advances In Spatial Econometrics.

Yang, Z., Yu, J., & Liu, S. (2016). Bias Correction and Refined Inferences for Fixed Effects Spatial Panel Data Models. *Regional Science and Urban Economics*, 61, 52-72.

Yang, Z. (2018). Unified M-Estimation of Fixed-Effects Spatial Dynamic Models with Short Panels. *Journal of Econometrics*, 205(2), 423-447.

Yu, J., de Jong, R., & Lee, L. (2008). Quasi-Maximum Likelihood Estimators for Spatial Dynamic Panel Data with Fixed Effects When Both $n$ and $T$ Are Large. *Journal of Econometrics*, 146(1), 118-134.

Zimmerman, G. M., & Messner, S. F. (2011). Neighborhood Context and Nonlinear Peer Effects on Adolescent Violent Crime. *Criminology*, 49(3), 873-903.

# Appendix

## A.1 Assumptions in Qu, Lee, and Yu (2017)

Following Jenish and Prucha (2009, 2012), consider generalized spatial processes on a possibly unevenly spaced lattice.

**Assumption A.1.** For a sample with $n$ units over $T$ periods, observations are located on a (possibly) unevenly spaced lattice $D \subset \mathbb{R}^{d+1}$, $d \geq 1$ and it is infinitely countable. The location $\ell : \mathcal{N} \times \mathcal{T} \to D_{nT} \subset D$ is a mapping of individual $i \in \mathcal{N} = \{1, \ldots, n\}$ and time $t \in \mathcal{T} = \{1, \ldots, T\}$ to its location $\ell(i,t) \in D_{nT} \subset \mathbb{R}^{d+1}$. For each spatial unit $i$, the location $\ell(i,t)$ is always one unit apart from $\ell(i,t-1)$ with respect to the time dimension. For a fixed $t = 1, \ldots, T$, any two elements in $D$ are separated by at least $\rho_{ct} > 0$ distance from each other, i.e., for any $\ell(i,t), \ell(j,t) \in D$, $\rho_{ij,t} \geq \rho_{ct}$, where $\rho_{ij,t}$ is the distance between $\ell(i,t)$ and $\ell(j,t)$ for a fixed $t$.

**Assumption A.2.** Elements in $W_{nt}$ are nonnegative with zero diagonals. For off-diagonal terms, $w_{ij,t} := h(z_{it}, z_{jt}) \cdot I(\rho_{ij} \leq \rho_c)$ or the row-normalized version of $w_{ij,t} = h(z_{it}, z_{jt}) \cdot I(\rho_{ij} \leq \rho_c) / \sum\limits_{\rho_{ik} \leq \rho_c} h_{ik}(z_{it}, z_{kt})$, where $h_{ij}(\cdot)$ is non-negative, uniformly bounded functions. For two different periods $t$ and $s$, the Lipschitz condition holds so that $|h(z_{it}, z_{jt}) - h(|z_{is}, z_{js})| \leq c_0(|z_{it} - z_{is}| + |z_{jt} - z_{js}|)$.

**Assumption A.3.** $n$ is an increasing function of $T$, and $T$ goes to infinity.

**Assumption A.4.** $\sup\limits_{n,t} ||W_{nt}||_\infty \leq c_w$, $||\Gamma_0||_1 < 1$, and $|\lambda_0|c_w + |\gamma_0| + |\rho_0|c_w < 1$, where $c_w$ is a finite constant.

Assumption A.2 is for technical purposes for sparsity in asymptotic setup. Note that $\rho_{ij}$ represents the time-invariant distance between $i$ and $j$, which is typically the geographic distance. The time-varying feature of $W_{nt}$ comes from $h(z_{it}, z_{jt})$, which captures the proximity between stochastic triggers of $i$ and $j$. Assumption A.3 requires a large $n$ and large $T$ case. As $T$ is large under assumption A.3, the initial condition problem would not be an issue. Assumption A.4 is a standard tool in spatial econometrics to guarantee the stability of the dynamic process by controlling the magnitude of spatial interaction of $W_{nt}$ matrix.

## A.2 Matrix Form

For asymptotic analysis, it is useful to have the likelihood function presented in matrix form. Now one may put the model (1) into a matrix form. Denote

- $\mathbf{Y}_{nT} = \begin{pmatrix} Y_{n1} \\ Y_{n2} \\ \vdots \\ Y_{nT} \end{pmatrix}, \quad \mathbf{Y}_{n,T-1} = \begin{pmatrix} Y_{n0} \\ Y_{n1} \\ \vdots \\ Y_{n,T-1} \end{pmatrix},$

- $\mathbf{W}_{nT} = \begin{pmatrix} W_{n1} & 0 & \cdots & 0 \\ 0 & W_{n2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & W_{nT} \end{pmatrix}, \quad \mathbf{W}_{n,T-1} = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ W_{n1} & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & W_{n,T-1} & 0 \end{pmatrix}$

- $\ell_0(\gamma_0, \rho_0) = \begin{pmatrix} \gamma_0 Y_{n0} + \rho_0 W_{n0} Y_{n0} \\ 0 \\ \vdots \\ 0 \end{pmatrix},$

- $c_{1,nT} = 1_T \otimes c_{1n}, \quad c_{2,nT} = 1_T \otimes c_{2n},$

- $\tau_{1,nT} = \begin{pmatrix} \tau_{11} \\ \tau_{21} \\ \vdots \\ \tau_{T1} \end{pmatrix} \otimes 1_n, \quad \tau_{2,nT} = \begin{pmatrix} \tau'_{12} \\ \tau'_{22} \\ \vdots \\ \tau'_{T2} \end{pmatrix} \otimes 1_n,$

where $\gamma_0$ and $\rho_0$ are the true parameters for $\gamma$ and $\rho$, respectively. $\mathbf{X}_{1,nT}$, $\varepsilon_{nT}$, $\Xi_{nT}$, $\mathbf{Z}_{nT}$ and $\mathbf{X}_{2,nT}$ are defined similarly. From Assumption 2, $\mathbf{V}_{nT} = \varepsilon_{nT}\delta_0 + \Xi_{nT}$ where $\xi_{nt}|e_{nt} \sim N(0, \sigma^2_{\xi_0} I_n)$ and $\delta_0$ and $\sigma^2_{\xi 0}$ represent the true parameter for $\delta$ and $\sigma^2_\xi$, respectively. Hence,

$$\begin{aligned}
\mathbf{Y}_{nT} &= \lambda_0 \mathbf{W}_{nT}\mathbf{Y}_{nT} + \mathbf{R}_{nT}\phi_0 + \ell_0(\gamma_0, \rho_0) + c_{1,nT,0} + \tau_{1,nT,0} + (\mathbf{Z}_{nT} - \mathbf{K}_{nT}\Phi_2 - c_{2,nT,0} - 1_n\tau_{2,nT,0})\delta_0 + \Xi_{nT} \\
&= \lambda_0 \mathbf{W}_{nT}\mathbf{Y}_{nT} + \mathbf{R}_{nT}\phi_0 + \ell_0(\gamma_0, \rho_0) + \overline{c_{1,nT,0}} + \overline{\tau_{1,nT,0}} + (\mathbf{Z}_{nT} - \mathbf{K}_{nT}\Phi_2)\delta_0 + \Xi_{nT},
\end{aligned}$$

where $\overline{c_{1,nT,0}} = c_{1,nT,0} - c_{2,nT,0}\delta_0$ and $\overline{\tau_{1,nT,0}} = \tau_{1,nT,0} - \tau_{2,nT,0}\delta_0$, and $\lambda_0$ and $\beta_0$ are the true parameters for $\lambda$ and $\beta$, respectively.

## A.3 Theorem Proofs

Let $\theta_0 = (\delta_0', \eta_0', \omega_0')'$, $\theta^* = (0', 0', \omega_0')'$, and $\tilde{\theta} = (0', 0', \tilde{\omega}')'$, where $\tilde{\omega}$ is the maximum likelihood (ML) estimator. Let $L_\psi(\theta) = \frac{1}{nT}\frac{\partial \ln L_{nT}^c(\theta)}{\partial \psi}$, $\mathcal{I}_{\psi\psi} := \mathrm{E}\frac{1}{nT}\Big(-\frac{\partial^2 \ln L_{nT}^c(\theta)}{\partial\psi\partial\psi'}\Big)$, and $I_{\psi\psi} = -\frac{1}{nT}\frac{\partial^2 \ln L_{nT}^c(\theta)}{\partial\psi\partial\psi'}$ such that $I_{\psi\psi}(\theta) = \mathcal{I}_{\psi\psi}(\theta) + o_p(1)$ uniformly on $\Theta$.

Consider $H_a^\delta : \delta_0 = \frac{\zeta}{\sqrt{nT}}$ and $H_a^\eta : \eta_0 = \frac{\nu}{\sqrt{nT}}$. The first-order Taylor expansion of the score function, $L_\delta(\tilde{\theta})$, around $\theta_0$ under $H_a^\delta$ and $H_a^\eta$ is

$$\sqrt{nT}L_\delta(\tilde{\theta}) = \sqrt{nT}L_\delta(\theta_0) - \frac{\partial L_\delta(\theta_0)}{\partial\delta}\zeta - \frac{\partial L_\delta(\theta_0)}{\partial\eta}\nu + \sqrt{nT}\frac{\partial L_\delta(\theta_0)}{\partial\omega}(\tilde{\omega} - \omega_0) + o_p(1). \tag{A.1}$$

Similarly, the Taylor expansion of $L_\omega(\theta^*)$ around $\tilde{\theta}$ under $H_a^\delta$ and $H_a^\eta$ is

$$\begin{aligned} \sqrt{nT}L_\omega(\theta^*) &= \sqrt{nT}L_\omega(\tilde{\theta}) + \sqrt{nT}\frac{\partial L_\omega(\tilde{\theta})}{\partial\omega}(\omega_0 - \tilde{\omega}) + o_p(1) \\ &\overset{\mathrm{a}}{=} \mathcal{I}_{\omega\omega}\sqrt{nT}(\tilde{\omega} - \omega_0), \end{aligned} \tag{A.2}$$

where 'a' represents 'asymptotically'. On the other hand, the Taylor expansion of $L_\omega(\theta^*)$ around $\theta_0$ under $H_a^\delta$ and $H_a^\eta$ is

$$\begin{aligned} \sqrt{nT}L_\omega(\theta^*) &= \sqrt{nT}L_\omega(\theta_0) - \frac{\partial L_\omega(\theta_0)}{\partial\delta}\zeta - \frac{\partial L_\omega(\theta_0)}{\partial\eta}\nu + o_p(1) \\ &\overset{\mathrm{a}}{=} \sqrt{nT}L_\omega(\theta_0) + \mathcal{I}_{\omega\delta}\zeta + \mathcal{I}_{\omega\eta}\nu. \end{aligned} \tag{A.3}$$

Since (A.2) and (A.3) are identical,

$$\mathcal{I}_{\omega\omega}\sqrt{nT}(\tilde{\omega} - \omega_0) \overset{a}{=} \sqrt{nT}L_\omega(\theta_0) + \mathcal{I}_{\omega\delta}\zeta + \mathcal{I}_{\omega\eta}\nu,$$

which implies

$$\sqrt{nT}(\tilde{\omega} - \omega_0) \overset{\mathrm{a}}{=} \mathcal{I}_{\omega\omega}^{-1}\sqrt{nT}L_\omega(\theta_0) + \mathcal{I}_{\omega\omega}^{-1}\mathcal{I}_{\omega\delta}\zeta + \mathcal{I}_{\omega\omega}^{-1}\mathcal{I}_{\omega\eta}\nu. \tag{A.4}$$

Using (A.4) for (A.1),

$$\begin{aligned}\sqrt{nT}L_\delta(\tilde\theta) &\overset{\text{a}}{=} \sqrt{nT}L_\delta(\theta_0) + \mathcal{I}_{\delta\delta}\zeta + \mathcal{I}_{\delta\eta}\nu - \mathcal{I}_{\delta\omega}\left(\mathcal{I}_{\omega\omega}^{-1}\sqrt{nT}L_\omega(\theta_0) + \mathcal{I}_{\omega\omega}^{-1}\mathcal{I}_{\omega\delta}\zeta + \mathcal{I}_{\omega\omega}^{-1}\mathcal{I}_{\omega\eta}\nu\right)\\ &= \sqrt{nT}L_\delta(\theta_0) - \mathcal{I}_{\delta\omega}\mathcal{I}_{\omega\omega}^{-1}\sqrt{nT}L_\omega(\theta_0) + (\mathcal{I}_{\delta\delta} - \mathcal{I}_{\delta\omega}\mathcal{I}_{\omega\omega}^{-1}\mathcal{I}_{\omega\delta})\zeta + (\mathcal{I}_{\delta\eta} - \mathcal{I}_{\delta\omega}\mathcal{I}_{\omega\omega}^{-1}\mathcal{I}_{\omega\eta})\nu\\ &:= \sqrt{nT}L_\delta(\theta_0) - \mathcal{I}_{\delta\omega}\mathcal{I}_{\omega\omega}^{-1}\sqrt{nT}L_\omega(\theta_0) + \mathcal{I}_{\delta\cdot\omega}\zeta + \mathcal{I}_{\delta\eta\cdot\omega}\nu,\end{aligned}$$

where $\mathcal{I}_{\delta\cdot\omega} := \mathcal{I}_{\delta\delta} - \mathcal{I}_{\delta\omega}\mathcal{I}_{\omega\omega}^{-1}\mathcal{I}_{\omega\delta}$ and $\mathcal{I}_{\delta\eta\cdot\omega} := \mathcal{I}_{\delta\eta} - \mathcal{I}_{\delta\omega}\mathcal{I}_{\omega\omega}^{-1}\mathcal{I}_{\omega\eta}$. Since the joint distribution of $L_\delta$ and $L_\omega$ are given as

$$\begin{pmatrix}\sqrt{nT}L_\delta(\theta_0)\\ \sqrt{nT}L_\omega(\theta_0)\end{pmatrix} - \begin{pmatrix}\Delta_{1,\delta}(\theta_0)+\Delta_{2,\delta}(\theta_0)\\ \Delta_{1,\omega}(\theta_0)+\Delta_{2,\omega}(\theta_0)\end{pmatrix} \xrightarrow{d} N\left(0, \begin{pmatrix}\mathcal{I}_{\delta\delta} & \mathcal{I}_{\delta\omega}\\ \mathcal{I}_{\omega\delta} & \mathcal{I}_{\omega\omega}\end{pmatrix}\right),$$

this implies

$$\sqrt{nT}L_\delta(\tilde\theta) \xrightarrow{d} N\left(\underbrace{(\Delta_{1,\delta}(\theta_0)+\Delta_{2,\delta}(\theta_0)) - \mathcal{I}_{\delta\omega}\mathcal{I}_{\omega\omega}^{-1}(\Delta_{1,\omega}(\theta_0)+\Delta_{2,\omega}(\theta_0))}_{:=\mathcal{B}_{\delta\omega}} + \mathcal{I}_{\delta\cdot\omega}\zeta + \mathcal{I}_{\delta\eta\cdot\omega}\nu, \mathcal{I}_{\delta\cdot\omega}\right), \tag{A.5}$$

where $\mathcal{B}_{\psi\omega} := (\Delta_{1,\psi}(\theta_0)+\Delta_{2,\psi}(\theta_0)) - \mathcal{I}_{\psi\omega}\mathcal{I}_{\omega\omega}^{-1}(\Delta_{1,\omega}(\theta_0)+\Delta_{2,\omega}(\theta_0))$ for $\psi \in \{\delta, \eta\}$ is the bias.

Since $L_\delta(\tilde\theta)$ is not centered around zero due to the biases from the two-way fixed effects and parametric misspecifications in $\eta$ (i.e., $\nu \neq 0$), there are four cases to be considered for adjustments: (i) Biased[15] and parametrically misspecified; (ii) Biased but robust to parametric misspecification; (iii) Unbiased but parametrically misspecified; (iv) Unbiased and robust to parametric misspecification.

(i) Consider the first case where none of the adjustments are made in (A.5). Then the score test has the form of

$$RS_\delta^{B,P}(\tilde\theta) = nTL'_\delta(\tilde\theta)I_{\delta\cdot\omega}^{-1}(\tilde\theta)L_\delta(\tilde\theta) \xrightarrow{d} \chi_p^2(\varphi_1),$$

where the superscript $B, P$ stands for 'biased and parametrically misspecified' and the non-centrality parameter $\varphi_1 = \mathcal{B}'_{\delta\omega}\mathcal{I}_{\delta\cdot\omega}^{-1}\mathcal{B}_{\delta\omega} + \zeta'\mathcal{I}_{\delta\cdot\omega}\zeta + \nu'\mathcal{I}_{\delta\eta\cdot\omega}\mathcal{I}_{\delta\cdot\omega}^{-1}\mathcal{I}_{\delta\eta\cdot\omega}\nu + 2(\mathcal{B}_{\delta\omega}\zeta + \mathcal{B}_{\delta\omega}\mathcal{I}_{\delta\cdot\omega}^{-1}\mathcal{I}_{\delta\eta\cdot\omega}\nu + \zeta'\mathcal{I}_{\delta\eta\cdot\omega}\nu)$. Notice that $\varphi_1$ is a function of $\mathcal{B}_{\delta\omega}$ and $\nu$ so that it suffers from not only the biases from the two-way fixed effects but also the parametric misspecifications in $\eta$.

(ii) Now consider the second case where the adjustment for the parametric misspecifications in $\eta$ is made. Even though $\nu$ is unknown, it can be canceled out using the joint distribution of $L_\delta$

[15] Biased by the two-way fixed effects (i.e., due to the incidental parameters problem).

and $L_\eta$. Since

$$\begin{pmatrix}\sqrt{nT}L_\delta(\tilde\theta)\\ \sqrt{nT}L_\eta(\tilde\theta)\end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix}\mathcal{B}_{\delta\omega}+\mathcal{I}_{\delta\cdot\omega}\zeta+\mathcal{I}_{\delta\eta\cdot\omega}\nu\\ \mathcal{B}_{\eta\omega}+\mathcal{I}_{\eta\delta\cdot\omega}\zeta+\mathcal{I}_{\eta\cdot\omega}\nu\end{pmatrix}, \begin{pmatrix}\mathcal{I}_{\delta\cdot\omega} & \mathcal{I}_{\delta\eta\cdot\omega}\\ \mathcal{I}_{\eta\delta\cdot\omega} & \mathcal{I}_{\eta\cdot\omega}\end{pmatrix}\right),$$

$\nu$ in the mean of $L_\delta(\tilde\theta)$ can be cancelled out letting

$$L^*_\delta(\tilde\theta) := L_\delta(\tilde\theta) - I_{\delta\eta\cdot\omega}(\tilde\theta)I^{-1}_{\eta\cdot\omega}(\tilde\theta)L_\eta(\tilde\theta),$$

which yields

$$\sqrt{nT}L^*_\delta(\tilde\theta) \xrightarrow{d} N\Big((I_{\delta\cdot\omega}-\mathcal{I}_{\delta\eta\cdot\omega}\mathcal{I}^{-1}_{\eta\cdot\omega}\mathcal{I}_{\eta\delta\cdot\omega})\zeta + (\mathcal{B}_{\delta\omega}-\mathcal{I}_{\delta\eta\cdot\omega}\mathcal{I}^{-1}_{\eta\cdot\omega}\mathcal{B}_{\eta\omega}), \mathcal{I}_{\delta\cdot\omega}-\mathcal{I}_{\delta\eta\cdot\omega}\mathcal{I}^{-1}_{\eta\cdot\omega}\mathcal{I}_{\eta\delta\cdot\omega}\Big).$$

The score test in this case then has a form of

$$RS^B_\delta(\tilde\theta) = nTL^{*\prime}_\delta(\tilde\theta)\Big(I_{\delta\cdot\omega}(\tilde\theta) - I_{\delta\eta\cdot\omega}(\tilde\theta)I^{-1}_{\eta\cdot\omega}(\tilde\theta)I_{\eta\delta\cdot\omega}(\tilde\theta)\Big)^{-1}L^*_\delta(\tilde\theta) \xrightarrow{d} \chi^2_p(\varphi_2),$$

where the superscript $B$ stands for 'biased' and the noncentrality parameter $\varphi_2 = (\mathcal{B}_{\delta\omega} - \mathcal{I}_{\delta\eta\cdot\omega}\mathcal{I}^{-1}_{\eta\cdot\omega}\mathcal{B}_{\eta\omega})'(\mathcal{I}_{\delta\cdot\omega} - \mathcal{I}_{\delta\eta\cdot\omega}\mathcal{I}^{-1}_{\eta\cdot\omega}\mathcal{I}_{\eta\delta\cdot\omega})^{-1}(\mathcal{B}_{\delta\omega} - \mathcal{I}_{\delta\eta\cdot\omega}\mathcal{I}^{-1}_{\eta\cdot\omega}\mathcal{B}_{\eta\omega}) + 2\zeta'(\mathcal{I}_{\delta\cdot\omega} - \mathcal{I}_{\delta\eta\cdot\omega}\mathcal{I}^{-1}_{\eta\cdot\omega}\mathcal{I}_{\eta\delta\cdot\omega})'(\mathcal{B}_{\delta\omega} - \mathcal{I}_{\delta\eta\cdot\omega}\mathcal{I}^{-1}_{\eta\cdot\omega}\mathcal{B}_{\eta\omega}) + \zeta'(\mathcal{I}_{\delta\cdot\omega} - \mathcal{I}_{\delta\eta\cdot\omega}\mathcal{I}^{-1}_{\eta\cdot\omega}\mathcal{I}_{\eta\delta\cdot\omega})\zeta$. Notice that $\varphi_2$ is a function of $\mathcal{B}_{\delta\omega}$ and $\mathcal{B}_{\eta\omega}$ but free from $\nu$ as $RS^B_\delta$ is adjusted for the parametric misspecification in $\eta$.

(iii) Now, let us consider the third case where the bias in the score function is corrected but still parametrically misspecified in $\eta$. From (A.10), I introduce the bias-corrected score function $C_\delta(\tilde\theta)$ of the form

$$\sqrt{nT}C_\delta(\tilde\theta) := \sqrt{nT}L_\delta(\tilde\theta) - \big(\Delta_{1,\delta}(\tilde\theta)+\Delta_{2,\delta}(\tilde\theta)\big) + I_{\delta\omega}(\tilde\theta)I^{-1}_{\omega\omega}(\tilde\theta)\big(\Delta_{1,\omega}(\tilde\theta)+\Delta_{2,\omega}(\tilde\theta)\big),$$

so that the asymptotic distribution of $C_\delta(\tilde\theta)$ is centered around zero under the joint null. The score test in this case has the form of

$$RS^P_\delta(\tilde\theta) = nTC'_\delta(\tilde\theta)I^{-1}_{\delta\cdot\omega}(\tilde\theta)C_\delta(\tilde\theta) \xrightarrow{d} \chi^2_p(\varphi_3),$$

with the noncentrality parameter $\varphi_3 = \zeta'\mathcal{I}_{\delta\cdot\omega}\zeta + 2\zeta'\mathcal{I}_{\delta\eta\cdot\omega}\nu + \nu'\mathcal{I}_{\delta\eta\cdot\omega}\mathcal{I}^{-1}_{\delta\cdot\omega}\mathcal{I}_{\delta\eta\cdot\omega}\nu$. Note that $\varphi_3$ is now free from the biases from the two-way fixed effects but still entails the local parametric misspecification magnitude in $\eta$ (i.e., $\nu$).

(iv) Finally, consider the case where all adjustments are made so that the score function is unbiased and robust to parametric misspecification in $\eta$. This is indeed the combined result of (ii) and (iii).

Similar to (ii), although $\nu$ is unknown, it can be canceled using the joint distribution of the bias-corrected score functions. As obtained in (iii), the bias-corrected score functions of $C_\delta$ and $C_\eta$ are

$$\begin{cases} \sqrt{nT}C_\delta(\tilde{\theta}) := \sqrt{nT}L_\delta(\tilde{\theta}) - \big(\Delta_{1,\delta}(\tilde{\theta}) + \Delta_{2,\delta}(\tilde{\theta})\big) + I_{\delta\omega}(\tilde{\theta})I_{\omega\omega}^{-1}(\tilde{\theta})(\Delta_{1,\omega}(\tilde{\theta}) + \Delta_{2,\omega}(\tilde{\theta})\big), \\ \sqrt{nT}C_\eta(\tilde{\theta}) := \sqrt{nT}L_\eta(\tilde{\theta}) - \big(\Delta_{1,\eta}(\tilde{\theta}) + \Delta_{2,\eta}(\tilde{\theta})\big) + I_{\eta\omega}(\tilde{\theta})I_{\omega\omega}^{-1}(\tilde{\theta})(\Delta_{1,\omega}(\tilde{\theta}) + \Delta_{2,\omega}(\tilde{\theta})\big), \end{cases}$$

so that

$$\begin{pmatrix} \sqrt{nT}C_\delta(\tilde{\theta}) \\ \sqrt{nT}C_\eta(\tilde{\theta}) \end{pmatrix} \xrightarrow{d} N\left( \begin{pmatrix} \mathcal{I}_{\delta\cdot\omega}\zeta + \mathcal{I}_{\delta\eta\cdot\omega}\nu \\ \mathcal{I}_{\eta\delta\cdot\omega}\zeta + \mathcal{I}_{\eta\cdot\omega}\nu \end{pmatrix}, \begin{pmatrix} \mathcal{I}_{\delta\cdot\omega} & \mathcal{I}_{\delta\eta\cdot\omega} \\ \mathcal{I}_{\eta\delta\cdot\omega} & \mathcal{I}_{\eta\cdot\omega} \end{pmatrix} \right).$$

Letting $C_\delta^*(\tilde{\theta}) = C_\delta(\tilde{\theta}) - I_{\delta\eta\cdot\omega}(\tilde{\theta})I_{\eta\cdot\omega}^{-1}(\tilde{\theta})C_\eta(\tilde{\theta})$, $\nu$ is cancelled out and $C_\delta^*(\tilde{\theta})$ has the asymptotic distribution as

$$\sqrt{nT}C_\delta^*(\tilde{\theta}) \xrightarrow{d} N\Big( (\mathcal{I}_{\delta\cdot\omega} - \mathcal{I}_{\delta\eta\cdot\omega}\mathcal{I}_{\eta\cdot\omega}^{-1}\mathcal{I}_{\eta\delta\cdot\omega})\zeta,\ \mathcal{I}_{\delta\cdot\omega} - \mathcal{I}_{\delta\eta\cdot\omega}\mathcal{I}_{\eta\cdot\omega}^{-1}\mathcal{I}_{\eta\delta\cdot\omega} \Big).$$

Thus the score test, in this case, has the form of

$$RS_\delta^*(\tilde{\theta}) = nTC_\delta^{*\prime}(\tilde{\theta}) \Big( I_{\delta\cdot\omega}(\tilde{\theta}) - I_{\delta\eta\cdot\omega}(\tilde{\theta})I_{\eta\cdot\omega}^{-1}(\tilde{\theta})I_{\eta\delta\cdot\omega}(\tilde{\theta}) \Big)^{-1} C_\delta^*(\tilde{\theta}) \xrightarrow{d} \chi_p^2(\varphi_4),$$

with the noncentrality parameter $\varphi_4 = \zeta'(\mathcal{I}_{\delta\cdot\omega} - \mathcal{I}_{\delta\eta\cdot\omega}\mathcal{I}_{\eta\cdot\omega}^{-1}\mathcal{I}_{\eta\delta\cdot\omega})\zeta$. Note that $\varphi_4$ is now free from both the biases from the two-way fixed effects and the parametric misspecification in $\eta$ (i.e., $\nu$). $\square$

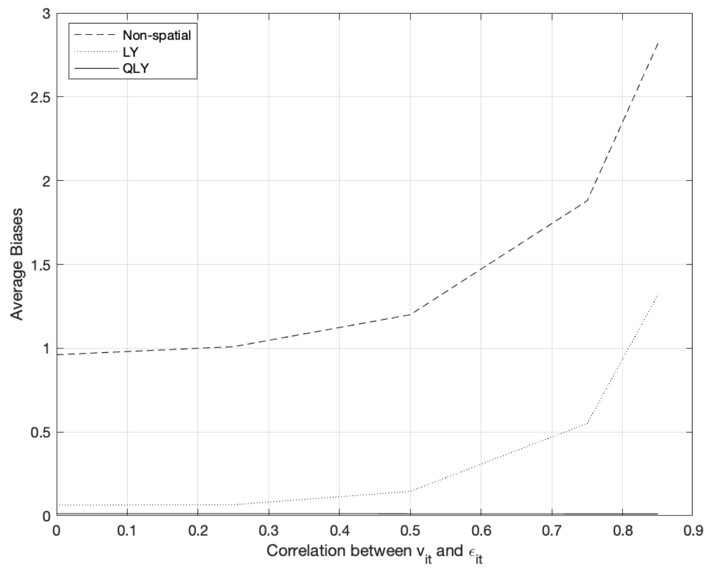
## A.4 Additional Remarks

**Remark A.1.** *(Importance of testing the endogeneity of $W$ in the weak-tied SDPD model)* One might think that the effects of an endogenous $W$ may not be large enough to cause invalid inference as long as $\eta$ is small. This would suggest that ignoring an endogenous $W$ in the weak-tied SDPD model might not be problematic.

To explore this, I conducted a simulation study. To represent a weak-tied SDPD with various levels of endogeneity of $W$, small values were set for the spatial dependence parameters as $(\lambda_0, \rho_0) = (0.25, 0.10)$. Moreover, the correlation between the error terms, which contributes to the endogeneity of $W$, was varied at 0, 0.25, 0.50, 0.75, and 0.85 by changing the value of $\sigma_{v\epsilon 0}$, along with $\Sigma_{\epsilon 0}$ and $\sigma^2_{\xi 0}$ set to 2 and 3, respectively.

The simulation results indicate that ignoring the endogenous $W$, even with small spatial dependence, yield biases (Figure A.1 and A.1). Notably, these biases increase sharply as the correlation between the error terms increases, underscoring the importance of testing for the endogeneity of $W$ in the weak-tied SDPD model. The estimates are presented in full format in Table A.2.

Figure A.1: Simulated Average Biases Across Three Models With Weak-Tied SDPD




Note: $(\lambda, \rho) = (0.25, 0.10)$. Replicated for 1,000 times; Sample sizes $(n, T) = (100, 101)$. Non-spatial represents the standard dynamic panel data model; LY stands for the SDPD with an exogenous spatial weight matrix (Lee and Yu 2012); QLY stands for the SDPD with an endogenous spatial weight matrix (Qu, Lee, and Yu 2017).

Table A.1: Simulation Results in the Weak-Tied SDPD with an Endogenous $W$

| | Average Bias | | |
|---|---|---|---|
| $\mathrm{corr}(v_{it}, \epsilon_{it})$ | Non-spatial | LY | QLY |
| 0.00 | 0.961 | 0.064 | 0.013 |
| 0.25 | 1.009 | 0.066 | 0.013 |
| 0.50 | 1.200 | 0.146 | 0.012 |
| 0.75 | 1.880 | 0.551 | 0.011 |
| 0.85 | 2.818 | 1.319 | 0.011 |

Note: $(\lambda, \rho) = (0.25, 0.10)$. Replicated for 1,000 times; Sample sizes $(n, T) = (100, 101)$. Non-spatial represents the standard dynamic panel data model with two-way fixed effects; LY refers to Lee and Yu (2012), where $W$ is given exogenous in SDPD; QLY refers to Qu, Lee, and Yu (2017), which accounts for an endogenous $W$ in SDPD.

# Table A.2: Comparison of Estimates Across Different Models in the Weak-Tied SDPD

| $corr(v_{it}, \epsilon_{it})$ | | $\lambda$ | $\rho$ | $\gamma$ | | | $\beta$ | | | $\sigma_v^2$ | $\delta$ | $\kappa$ | $\Gamma$ | $\sigma_\xi^2$ | $\Sigma_\epsilon$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.00 | TRUE | 0.250 | 0.100 | 0.050 | 0.300 | -1.000 | 2.500 | 4.000 | -0.200 | 3.000 | 0.000 | 0.200 | 0.300 | 3.000 | 2.000 |
| | Non-spatial | | | 0.070 | 0.313 | -1.028 | 2.543 | 4.061 | -0.204 | 3.791 | | | | | |
| | | | | (0.004) | (0.019) | (0.019) | (0.019) | (0.018) | (0.019) | (0.060) | | | | | |
| | Bias | | | 0.020 | 0.013 | 0.028 | 0.043 | 0.061 | 0.004 | 0.791 | | | | | |
| | LY | 0.245 | 0.100 | 0.049 | 0.301 | -1.000 | 2.501 | 4.002 | -0.200 | 2.946 | | | | | |
| | | (0.005) | (0.006) | (0.003) | (0.017) | (0.017) | (0.017) | (0.017) | (0.018) | (0.044) | | | | | |
| | Bias | 0.005 | 0.000 | 0.001 | 0.001 | 0.000 | 0.001 | 0.002 | 0.000 | 0.054 | | | | | |
| | QLY | 0.248 | 0.099 | 0.050 | 0.301 | -1.000 | 2.501 | 4.001 | -0.200 | 3.004 | 0.000 | 0.199 | 0.300 | 3.004 | 1.998 |
| | | (0.005) | (0.006) | (0.003) | (0.018) | (0.017) | (0.018) | (0.017) | (0.017) | | (0.012) | (0.010) | (0.014) | (0.043) | (0.028) |
| | Bias | 0.002 | 0.001 | 0.000 | 0.001 | 0.000 | 0.001 | 0.001 | 0.000 | | 0.000 | 0.001 | 0.000 | 0.004 | 0.002 |
| 0.25 | TRUE | 0.250 | 0.100 | 0.050 | 0.300 | -1.000 | 2.500 | 4.000 | -0.200 | 3.200 | 0.317 | 0.200 | 0.300 | 3.000 | 2.000 |
| | Non-spatial | | | 0.070 | 0.313 | -1.028 | 2.543 | 4.061 | -0.205 | 4.038 | | | | | |
| | | | | (0.004) | (0.020) | (0.019) | (0.020) | (0.019) | (0.020) | (0.064) | | | | | |
| | Bias | | | 0.020 | 0.013 | 0.028 | 0.043 | 0.061 | 0.005 | 0.837 | | | | | |
| | LY | 0.252 | 0.099 | 0.049 | 0.300 | -0.999 | 2.500 | 4.000 | -0.200 | 3.140 | | | | | |
| | | (0.005) | (0.006) | (0.004) | (0.018) | (0.018) | (0.018) | (0.018) | (0.018) | (0.047) | | | | | |
| | Bias | 0.002 | 0.001 | 0.001 | 0.000 | 0.001 | 0.000 | 0.000 | 0.000 | 0.061 | | | | | |
| | QLY | 0.248 | 0.099 | 0.050 | 0.300 | -1.000 | 2.500 | 4.001 | -0.201 | 3.202 | 0.317 | 0.199 | 0.300 | 3.004 | 1.998 |
| | | (0.005) | (0.006) | (0.003) | (0.018) | (0.017) | (0.018) | (0.017) | (0.017) | | (0.012) | (0.010) | (0.014) | (0.043) | (0.028) |
| | Bias | 0.002 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.001 | 0.001 | | 0.001 | 0.001 | 0.000 | 0.004 | 0.002 |
| 0.50 | TRUE | 0.250 | 0.100 | 0.050 | 0.300 | -1.000 | 2.500 | 4.000 | -0.200 | 4.001 | 0.708 | 0.200 | 0.300 | 3.000 | 2.000 |
| | Non-spatial | | | 0.072 | 0.313 | -1.028 | 2.543 | 4.062 | -0.205 | 5.028 | | | | | |
| | | | | (0.004) | (0.022) | (0.022) | (0.022) | (0.022) | (0.022) | (0.081) | | | | | |
| | Bias | | | 0.022 | 0.013 | 0.028 | 0.043 | 0.062 | 0.005 | 1.027 | | | | | |
| | LY | 0.276 | 0.098 | 0.048 | 0.299 | -0.997 | 2.496 | 3.994 | -0.200 | 3.900 | | | | | |
| | | (0.006) | (0.006) | (0.004) | (0.020) | (0.020) | (0.020) | (0.019) | (0.020) | (0.058) | | | | | |
| | Bias | 0.026 | 0.002 | 0.002 | 0.001 | 0.003 | 0.004 | 0.006 | 0.000 | 0.101 | | | | | |
| | QLY | 0.248 | 0.099 | 0.050 | 0.300 | -0.999 | 2.500 | 4.001 | -0.201 | 4.005 | 0.709 | 0.199 | 0.300 | 3.003 | 1.998 |
| | | (0.005) | (0.006) | (0.003) | (0.018) | (0.017) | (0.018) | (0.017) | (0.017) | | (0.012) | (0.008) | (0.013) | (0.043) | (0.028) |
| | Bias | 0.002 | 0.001 | 0.000 | 0.000 | 0.001 | 0.000 | 0.001 | 0.001 | | 0.001 | 0.001 | 0.000 | 0.003 | 0.002 |
| 0.75 | TRUE | 0.250 | 0.100 | 0.050 | 0.300 | -1.000 | 2.500 | 4.000 | -0.200 | 6.864 | 1.390 | 0.200 | 0.300 | 3.000 | 2.000 |
| | Non-spatial | | | 0.076 | 0.313 | -1.028 | 2.544 | 4.063 | -0.205 | 8.566 | | | | | |
| | | | | (0.005) | (0.029) | (0.029) | (0.029) | (0.029) | (0.029) | (0.141) | | | | | |
| | Bias | | | 0.026 | 0.013 | 0.028 | 0.044 | 0.063 | 0.005 | 1.701 | | | | | |
| | LY | 0.346 | 0.094 | 0.046 | 0.295 | -0.989 | 2.484 | 3.976 | -0.198 | 6.478 | | | | | |
| | | (0.007) | (0.008) | (0.005) | (0.026) | (0.026) | (0.025) | (0.025) | (0.026) | (0.096) | | | | | |
| | Bias | 0.096 | 0.006 | 0.004 | 0.005 | 0.011 | 0.016 | 0.024 | 0.002 | 0.386 | | | | | |
| | QLY | 0.248 | 0.100 | 0.050 | 0.300 | -0.999 | 2.500 | 4.001 | -0.201 | 6.869 | 1.392 | 0.199 | 0.300 | 3.003 | 1.999 |
| | | (0.005) | (0.005) | (0.003) | (0.018) | (0.017) | (0.018) | (0.017) | (0.017) | | (0.013) | (0.007) | (0.010) | (0.043) | (0.028) |
| | Bias | 0.002 | 0.000 | 0.000 | 0.000 | 0.001 | 0.000 | 0.001 | 0.001 | | 0.002 | 0.001 | 0.000 | 0.003 | 0.001 |
| 0.85 | TRUE | 0.250 | 0.100 | 0.050 | 0.300 | -1.000 | 2.500 | 4.000 | -0.200 | 10.821 | 1.978 | 0.200 | 0.300 | 3.000 | 2.000 |
| | Non-spatial | | | 0.080 | 0.314 | -1.028 | 2.544 | 4.063 | -0.206 | 13.454 | | | | | |
| | | | | (0.006) | (0.037) | (0.037) | (0.037) | (0.037) | (0.037) | (0.221) | | | | | |
| | Bias | | | 0.030 | 0.014 | 0.028 | 0.044 | 0.063 | 0.006 | 2.633 | | | | | |
| | LY | 0.414 | 0.089 | 0.043 | 0.292 | -0.982 | 2.472 | 3.958 | -0.197 | 9.783 | | | | | |
| | | (0.007) | (0.009) | (0.006) | (0.032) | (0.032) | (0.031) | (0.030) | (0.032) | (0.146) | | | | | |
| | Bias | 0.164 | 0.011 | 0.007 | 0.008 | 0.018 | 0.028 | 0.042 | 0.003 | 1.038 | | | | | |
| | QLY | 0.248 | 0.099 | 0.050 | 0.300 | -1.000 | 2.500 | 4.000 | -0.200 | 10.826 | 1.980 | 0.199 | 0.300 | 3.003 | 1.999 |
| | | (0.005) | (0.005) | (0.003) | (0.018) | (0.017) | (0.018) | (0.017) | (0.017) | | (0.013) | (0.006) | (0.008) | (0.043) | (0.028) |
| | Bias | 0.002 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | | 0.002 | 0.001 | 0.000 | 0.003 | 0.001 |

Notes: Standard errors are reported in the parentheses. $(n, T) = (100, 101)$, replicated for 1,000 times. Non-spatial indicates a (non-spatial) dynamic panel data model with two-way fixed effects; LY refers to Lee and Yu (2012), where $W$ is given exogenous in SDPD; QLY refers to Qu, Lee, and Yu (2017), which accounts for an endogenous $W$ in SDPD.

**Remark A.2.** *(Tradeoff)* As there is no free lunch, $RS^*_\delta(\tilde{\theta})$ also has a tradeoff. While the over-rejection of the null is resolved to be fixed by the adjustments of bias-correction and orthogonalization towards local parametric misspecifications in $\eta$, the power of $RS^*_\delta(\tilde{\theta})$ is less than that of $RS^P_\delta(\tilde{\theta})$ if the parametric misspecifications in $\eta$ do not present, because then, the adjustment for parametric misspecifications in $\eta$ is redundant and thus inefficient.

Specifically, under $H^\delta_a$ and $H^\eta_0$, the result shows $RS^*_\delta(\tilde{\theta}) \xrightarrow{d} \chi^2_p(\varphi_4)$ and $RS^P_\delta(\tilde{\theta}) \xrightarrow{d} \chi^2_p(\varphi_3)$ with $\varphi_3 - \varphi_4 \geq 0$, indicating $RS^*_\delta(\tilde{\theta})$ asymptotically has less power than $RS^P_\delta(\tilde{\theta})$ when there is no local parametric misspecification in $\eta$, i.e., $\nu = 0$. This can be interpreted as our premium to pay to have less power when there is no local parametric misspecification in $\eta$.

However, simulations show that such power loss turns out to be minimal compared to $RS^P_\delta(\tilde{\theta})$ when there is no local parametric misspecification in $\eta$ (Figure A.3). This assures the shortcoming of $RS^*_\delta(\tilde{\theta})$ is not severe in practice. The tradeoff of $RS^*_\delta(\tilde{\theta})$ is summarized in Figure A.2: While $RS^*_\delta(\tilde{\theta})$ guarantees for type I error to be fixed asymptotically against the unadjusted score tests, one pays the *premium* for less power when there is no local parametric misspecification, which turns out to be negligible in practice. This assures the nice finite sample properties of $RS^*_\delta(\tilde{\theta})$ for valid inference.

Figure A.2: Tradeoff of the Robust Score Test

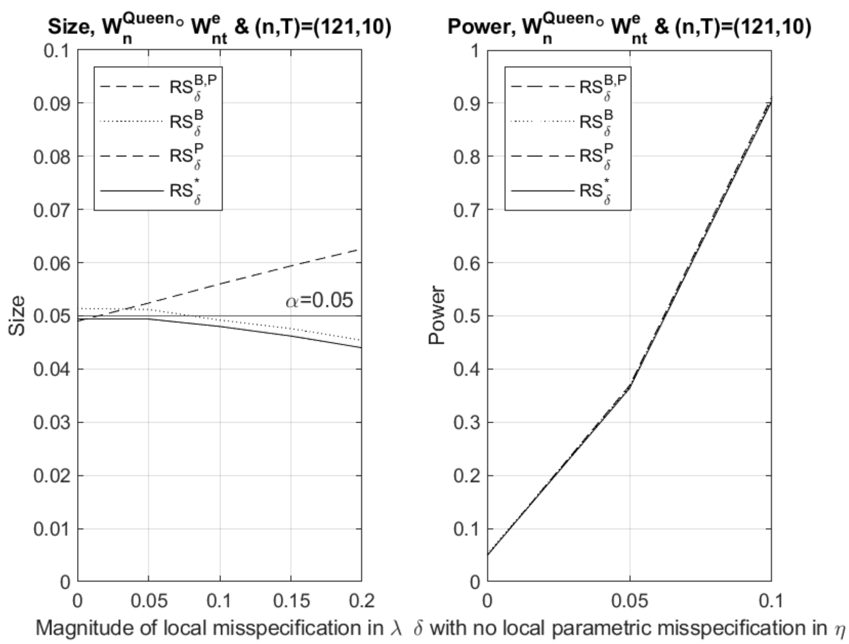
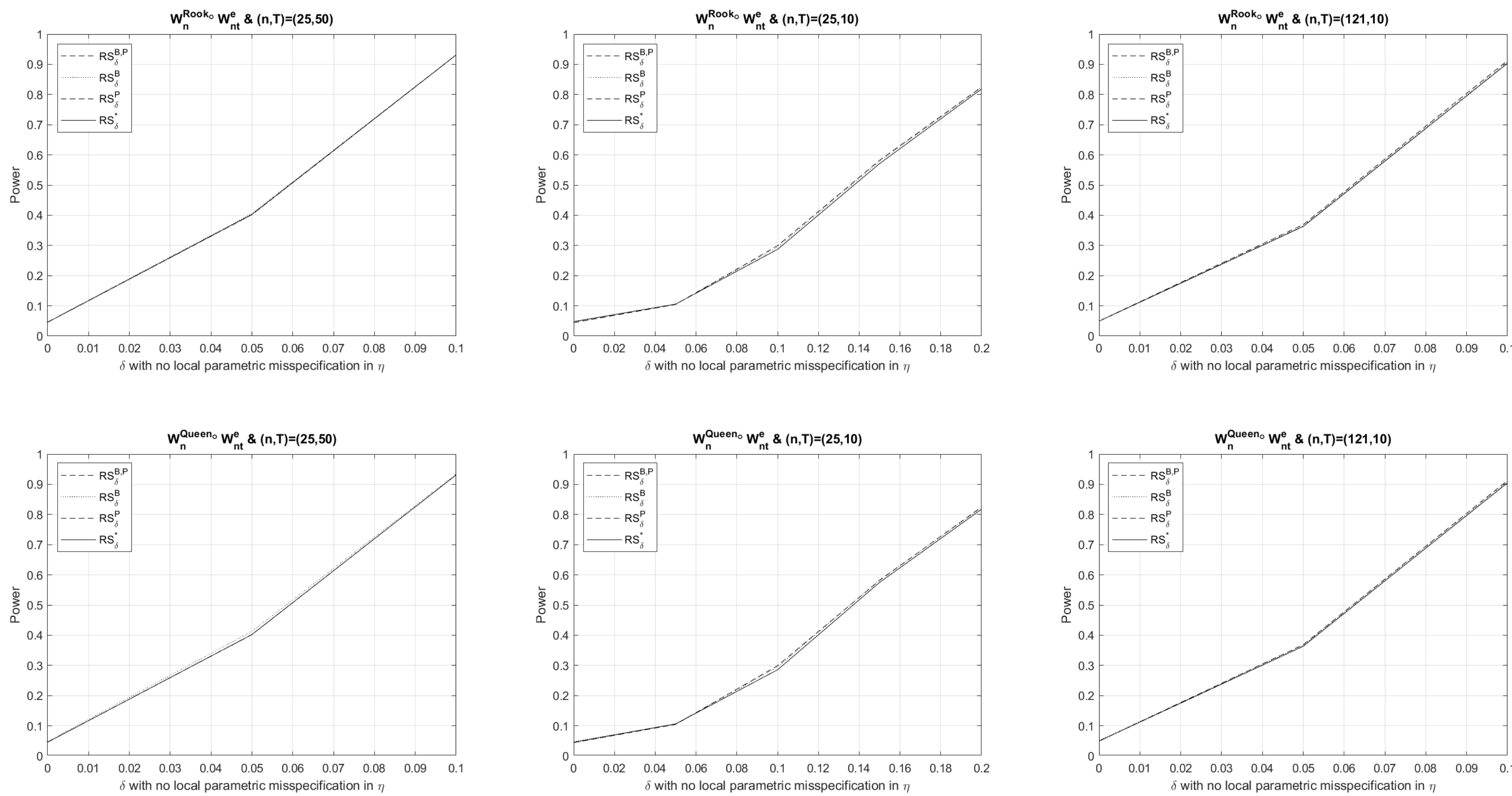


Figure A.3: Power by Sample Size with No Local Parametric Misspecification



**Remark A.3.** *(Comparative Relation With $RS_\delta^P(\tilde{\theta})$)* From the form of $\mathrm{E}(C_\delta(\theta_0)|C_\eta(\theta_0)) = I_{\delta\eta\cdot\omega} I_{\eta\cdot\omega}^{-1} C_\eta(\theta_0)$, the relation between the two score functions can be written as

$$C_\delta(\theta_0) = \rho_1 C_\eta(\theta_0) + C_\delta^*(\theta_0),$$

where $\rho_1 = I_{\delta\eta\cdot\omega} I_{\eta\cdot\omega}^{-1}$ is the linear projection coefficient and $C_\delta^*(\theta_0)$ is the orthogonalized part of $C_\delta(\theta_0)$ characterized as the prediction error in the linear projection of $C_\delta(\theta_0)$ onto the plane of $C_\eta(\theta_0)$.

Now consider the three cases where the sign of $\rho_1$ is either zero, positive, or negative. Note that $RS_\delta^P(\tilde{\theta})$ and $RS_\delta^*(\tilde{\theta})$ are equivalent if $\rho_1$ is zero, i.e., $I_{\delta\eta\cdot\omega} = 0$. Thus, $RS_\delta^P(\tilde{\theta})$ is the threshold of $RS_\delta^*(\tilde{\theta})$ of deciding whether $\rho_1$ is zero or not. Hinted from the approximate form of the Durbin-Watson (DW) test, that $DW \approx 2(1 - \rho_2)$ with the serial correlation $\rho_2$, i.e., 2 is the threshold to determine if $\rho_2$ is zero, positive, or negative, the comparative relation of $RS_\delta^P(\tilde{\theta})$ and $RS_\delta^*(\tilde{\theta})$ is conjectured as

$$RS_\delta^*(\tilde{\theta}) = RS_\delta^P(\tilde{\theta})(1 - \rho_1).$$

Table A.3 summarizes this conjecture on the comparative relation between $RS_\delta^P(\tilde{\theta})$ and $RS_\delta^*(\tilde{\theta})$.

Table A.3: DW Test and Conjecture on Score Test

| | DW Test | | | Score Test | | |
|---|---|---|---|---|---|---|
| Dependence | | $\rho_2$ | | | $\rho_1$ | |
| Cases | <0 | =0 | >0 | <0 | =0 | >0 |
| Comparison | $DW > 2$ | $DW = 2$ | $DW < 2$ | $RS_\delta^* > RS_\delta^P$ | $RS_\delta^* = RS_\delta^P$ | $RS_\delta^* < RS_\delta^P$ |

**Remark A.4.** *(Joint test for $H_0^{\delta,\eta}$)* In case it is of interest to simultaneously test both the endogeneity of $W$ and the appropriateness of local alternatives for $\eta$, one can implement a joint test $H_0^{\delta,\eta}$ : $(\delta', \eta')' = (0', 0')'$. For this, one can utilize the useful equality that decomposes the score test for the joint null into two marginal orthogonal score tests as (Bera and Yoon 1993)

$$RS_{\delta\eta}(\tilde{\theta}) = RS_\delta^*(\tilde{\theta}) + RS_\eta(\tilde{\theta}).$$

Intuitively, this is similar to the cosine rule, as it decomposes the score test statistic into orthogonal terms. This equality further helps researchers ensure the correctness of their algebra when deriving statistics. Unlike other tests, this property is unique to the score test, which is an additional advantage.

**Remark A.5.** *(Robust score test for spatial and time dependencies)* The roles of the testing parameter and the locally misspecified parameters can be switched to develop a robust score test for spatial and time dependencies. In this case, the robust score test determines the presence of spatial or time dependencies circumventing the endogeneity issue of a spatial weight matrix. This test would be useful if a researcher suspects the presence of spatial or time dependencies but is not interested in taking account of the endogeneity of a spatial weight matrix.

## A.5 Proofs for Section 3.4.3. Some Extensions (iii)

Let $\theta_0 = (\delta_0', \eta_0', \omega_0')'$, $\theta^* = (0', 0', \omega_0')'$, and $\tilde{\theta} = (0', 0', \tilde{\omega}')'$, where $\tilde{\omega}$ is the ML estimator. Let $L_\psi(\theta) = \frac{1}{nT}\frac{\partial \ln L_{nT}^c(\theta)}{\partial \psi}$, $\mathcal{I}_{\psi\psi} := \mathrm{E}\frac{1}{nT}\Big(-\frac{\partial^2 \ln L_{nT}^c(\theta)}{\partial\psi\partial\psi'}\Big)$, and $I_{\psi\psi} = -\frac{1}{nT}\frac{\partial^2 \ln L_{nT}^c(\theta)}{\partial\psi\partial\psi'}$ such that $I_{\psi\psi}(\theta) = \mathcal{I}_{\psi\psi}(\theta) + o_p(1)$ uniformly on $\Theta$ and $\mathcal{J}_{\psi\psi} := \mathrm{E}\frac{1}{nT}\Big(\frac{\partial \ln L_{nT}^c(\theta)}{\partial\psi}\frac{\partial \ln L_{nT}^c(\theta)}{\partial\psi'}\Big)$.

Consider $H_a^\delta : \delta_0 = \frac{\zeta}{\sqrt{nT}}$ and $H_a^\eta : \eta_0 = \frac{\nu}{\sqrt{nT}}$. The first-order Taylor expansion of the score function, $L_\delta(\tilde{\theta})$, around $\theta_0$ under $H_a^\delta$ and $H_a^\eta$ is

$$\sqrt{nT}L_\delta(\tilde{\theta}) = \sqrt{nT}L_\delta(\theta_0) - \frac{\partial L_\delta(\theta_0)}{\partial\delta}\zeta - \frac{\partial L_\delta(\theta_0)}{\partial\eta}\nu + \sqrt{nT}\frac{\partial L_\delta(\theta_0)}{\partial\omega}(\tilde{\omega} - \omega_0) + o_p(1). \tag{A.6}$$

Similarly, the Taylor expansion of $L_\omega(\theta^*)$ around $\tilde{\theta}$ under $H_a^\delta$ and $H_a^\eta$ is

$$\begin{aligned}\sqrt{nT}L_\omega(\theta^*) &= \sqrt{nT}L_\omega(\tilde{\theta}) + \sqrt{nT}\frac{\partial L_\omega(\tilde{\theta})}{\partial\omega}(\omega_0 - \tilde{\omega}) + o_p(1)\\ &\overset{\mathrm{a}}{=} \mathcal{I}_{\omega\omega}\sqrt{nT}(\tilde{\omega} - \omega_0),\end{aligned} \tag{A.7}$$

where 'a' represents 'asymptotically'. On the other hand, the Taylor expansion of $L_\omega(\theta^*)$ around $\theta_0$ under $H_a^\delta$ and $H_a^\eta$ is

$$\begin{aligned}\sqrt{nT}L_\omega(\theta^*) &= \sqrt{nT}L_\omega(\theta_0) - \frac{\partial L_\omega(\theta_0)}{\partial\delta}\zeta - \frac{\partial L_\omega(\theta_0)}{\partial\eta}\nu + o_p(1)\\ &\overset{\mathrm{a}}{=} \sqrt{nT}L_\omega(\theta_0) + \mathcal{I}_{\omega\delta}\zeta + \mathcal{I}_{\omega\eta}\nu.\end{aligned} \tag{A.8}$$

Since (A.7) and (A.8) are identical,

$$\mathcal{I}_{\omega\omega}\sqrt{nT}(\tilde{\omega} - \omega_0) \overset{a}{=} \sqrt{nT}L_\omega(\theta_0) + \mathcal{I}_{\omega\delta}\zeta + \mathcal{I}_{\omega\eta}\nu,$$

which implies

$$\sqrt{nT}(\tilde{\omega} - \omega_0) \overset{\mathrm{a}}{=} \mathcal{I}_{\omega\omega}^{-1}\sqrt{nT}L_\omega(\theta_0) + \mathcal{I}_{\omega\omega}^{-1}\mathcal{I}_{\omega\delta}\zeta + \mathcal{I}_{\omega\omega}^{-1}\mathcal{I}_{\omega\eta}\nu. \tag{A.9}$$

Using (A.9) for (A.6),

$$
\begin{aligned}
\sqrt{nT}L_\delta(\tilde{\theta}) &\overset{\text{a}}{=} \sqrt{nT}L_\delta(\theta_0) + \mathcal{I}_{\delta\delta}\zeta + \mathcal{I}_{\delta\eta}\nu - \mathcal{I}_{\delta\omega}\left(\mathcal{I}_{\omega\omega}^{-1}\sqrt{nT}L_\omega(\theta_0) + \mathcal{I}_{\omega\omega}^{-1}\mathcal{I}_{\omega\delta}\zeta + \mathcal{I}_{\omega\omega}^{-1}\mathcal{I}_{\omega\eta}\nu\right) \\
&= \sqrt{nT}L_\delta(\theta_0) - \mathcal{I}_{\delta\omega}\mathcal{I}_{\omega\omega}^{-1}\sqrt{nT}L_\omega(\theta_0) + (\mathcal{I}_{\delta\delta} - \mathcal{I}_{\delta\omega}\mathcal{I}_{\omega\omega}^{-1}\mathcal{I}_{\omega\delta})\zeta + (\mathcal{I}_{\delta\eta} - \mathcal{I}_{\delta\omega}\mathcal{I}_{\omega\omega}^{-1}\mathcal{I}_{\omega\eta})\nu \\
&:= \sqrt{nT}L_\delta(\theta_0) - \mathcal{I}_{\delta\omega}\mathcal{I}_{\omega\omega}^{-1}\sqrt{nT}L_\omega(\theta_0) + \mathcal{I}_{\delta\cdot\omega}\zeta + \mathcal{I}_{\delta\eta\cdot\omega}\nu,
\end{aligned}
$$

where $\mathcal{I}_{\delta\cdot\omega} := \mathcal{I}_{\delta\delta} - \mathcal{I}_{\delta\omega}\mathcal{I}_{\omega\omega}^{-1}\mathcal{I}_{\omega\delta}$ and $\mathcal{I}_{\delta\eta\cdot\omega} := \mathcal{I}_{\delta\eta} - \mathcal{I}_{\delta\omega}\mathcal{I}_{\omega\omega}^{-1}\mathcal{I}_{\omega\eta}$. Since the joint distribution of $L_\delta$ and $L_\omega$ are given as

$$
\begin{pmatrix} \sqrt{nT}L_\delta(\theta_0) \\ \sqrt{nT}L_\omega(\theta_0) \end{pmatrix} - \begin{pmatrix} \Delta_{1,\delta}(\theta_0) + \Delta_{2,\delta}(\theta_0) \\ \Delta_{1,\omega}(\theta_0) + \Delta_{2,\omega}(\theta_0) \end{pmatrix} \xrightarrow{d} N\left(0, \begin{pmatrix} \mathcal{J}_{\delta\delta} & \mathcal{J}_{\delta\omega} \\ \mathcal{J}_{\omega\delta} & \mathcal{J}_{\omega\omega} \end{pmatrix}\right),
$$

this implies

$$
\sqrt{nT}L_\delta(\tilde{\theta}) \xrightarrow{d} N\left(\underbrace{(\Delta_{1,\delta}(\theta_0) + \Delta_{2,\delta}(\theta_0)) - \mathcal{I}_{\delta\omega}\mathcal{I}_{\omega\omega}^{-1}(\Delta_{1,\omega}(\theta_0) + \Delta_{2,\omega}(\theta_0))}_{:=\mathcal{B}_{\delta\omega}} + \mathcal{I}_{\delta\cdot\omega}\zeta + \mathcal{I}_{\delta\eta\cdot\omega}\nu,\ \mathcal{K}_{\delta\cdot\omega}\right), \tag{A.10}
$$

where $\mathcal{B}_{\psi\omega} := (\Delta_{1,\psi}(\theta_0) + \Delta_{2,\psi}(\theta_0)) - \mathcal{I}_{\psi\omega}\mathcal{I}_{\omega\omega}^{-1}(\Delta_{1,\omega}(\theta_0) + \Delta_{2,\omega}(\theta_0))$ is the bias and $\mathcal{K}_{\psi\cdot\omega} := \mathcal{J}_{\psi\psi} + \mathcal{I}_{\psi\omega}\mathcal{I}_{\omega\omega}^{-1}\mathcal{J}_{\omega\omega}\mathcal{I}_{\omega\omega}^{-1}\mathcal{I}_{\omega\psi} - \mathcal{J}_{\psi\omega}\mathcal{I}_{\omega\omega}^{-1}\mathcal{I}_{\omega\psi} - \mathcal{I}_{\psi\omega}\mathcal{I}_{\omega\omega}^{-1}\mathcal{J}_{\omega\psi}$ for $\psi \in \{\delta, \eta\}$.

Let $C_\delta(\tilde{\theta})$ be the *bias-corrected* score function such that $\sqrt{nT}C_\delta(\tilde{\theta}) := \sqrt{nT}L_\delta(\tilde{\theta}) - B_{\delta\omega}(\tilde{\theta})$ and its asymptotic distribution under $H_a^\delta$ and $H_a^\eta$ is given as

$$
\sqrt{nT}C_\delta(\tilde{\theta}) \xrightarrow{d} N(\mathcal{I}_{\delta\cdot\omega}\zeta + \mathcal{I}_{\delta\eta\cdot\omega}\nu,\ \mathcal{K}_{\delta\cdot\omega}).
$$

Note that under $H_0^\delta$ and $H_0^\eta$, where $\tilde{\theta}$ is estimated, $\sqrt{nT}C_\delta(\tilde{\theta})$ is central chi-squared so that the biases from the original score functions are now corrected. Also, its asymptotic distribution reveals specific sources of the local parametric misspecification (i.e., $\mathcal{I}_{\delta\eta\cdot\omega}\nu$) and the distributional misspecification (i.e., $\mathcal{K}_{\delta\cdot\omega}$).

Hence

$$
\begin{pmatrix} \sqrt{nT}C_\delta(\tilde{\theta}) \\ \sqrt{nT}C_\eta(\tilde{\theta}) \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} \mathcal{I}_{\delta\cdot\omega}\zeta + \mathcal{I}_{\delta\eta\cdot\omega}\nu \\ \mathcal{I}_{\eta\delta\cdot\omega}\zeta + \mathcal{I}_{\eta\cdot\omega}\nu \end{pmatrix}, \begin{pmatrix} \mathcal{K}_{\delta\cdot\omega} & \mathcal{K}_{\delta\eta\cdot\omega} \\ \mathcal{K}_{\eta\delta\cdot\omega} & \mathcal{K}_{\eta\cdot\omega} \end{pmatrix}\right),
$$

where $\mathcal{K}_{\delta\eta\cdot\omega} = \mathcal{J}_{\delta\eta} + \mathcal{I}_{\delta\omega}\mathcal{I}_{\omega\omega}^{-1}\mathcal{J}_{\omega\omega}\mathcal{I}_{\omega\omega}^{-1}\mathcal{I}_{\omega\eta} - \mathcal{I}_{\delta\omega}\mathcal{I}_{\omega\omega}^{-1}\mathcal{J}_{\omega\eta} - \mathcal{J}_{\delta\omega}\mathcal{I}_{\omega\omega}^{-1}\mathcal{I}_{\omega\eta}$.

Now letting $C^*_\delta(\tilde{\theta}) := C_\delta(\tilde{\theta}) - I_{\delta\eta\cdot\omega}(\tilde{\theta})I^{-1}_{\eta\cdot\omega}(\tilde{\theta})C_\eta(\tilde{\theta})$, even though it is *unknown*, $\nu$ can be cancelled out. Then $C^*_\delta(\tilde{\theta})$ has the asymptotic distribution as

$$\sqrt{nT}C^*_\delta(\tilde{\theta}) \xrightarrow{d} N\big((\mathcal{I}_{\delta\cdot\omega} - \mathcal{I}_{\delta\eta\cdot\omega}\mathcal{I}^{-1}_{\eta\cdot\omega}\mathcal{I}_{\eta\delta\cdot\omega})\zeta,\ \mathcal{V}_{\delta\cdot\omega}\big),$$

where $\mathcal{V}_{\delta\cdot\omega} := \mathcal{K}_{\delta\cdot\omega} + \mathcal{I}_{\delta\eta\cdot\omega}\mathcal{I}^{-1}_{\eta\cdot\omega}\mathcal{K}_{\eta\cdot\omega}\mathcal{I}^{-1}_{\eta\cdot\omega}\mathcal{I}_{\eta\delta\cdot\omega} - \mathcal{I}_{\delta\eta\cdot\omega}\mathcal{I}^{-1}_{\eta\cdot\omega}\mathcal{K}_{\eta\delta\cdot\omega} - \mathcal{K}_{\delta\eta\cdot\omega}\mathcal{I}^{-1}_{\eta\cdot\omega}\mathcal{I}_{\eta\delta\cdot\omega}$.

Thus, the *robust* score test, which is further robust to Information Inequality, has the form of

$$RS^{**}_\delta(\tilde{\theta}) = nTC^{*\prime}_\delta(\tilde{\theta})V^{-1}_{\delta\cdot\omega}(\tilde{\theta})C^*_\delta(\tilde{\theta}) \xrightarrow{d} \chi^2_p(\varphi),$$

with the noncentrality parameter $\varphi := \zeta'(\mathcal{I}_{\delta\cdot\omega} - \mathcal{I}_{\delta\eta\cdot\omega}\mathcal{I}^{-1}_{\eta\cdot\omega}\mathcal{I}_{\eta\delta\cdot\omega})'\mathcal{V}^{-1}_{\delta\cdot\omega}(\mathcal{I}_{\delta\cdot\omega} - \mathcal{I}_{\delta\eta\cdot\omega}\mathcal{I}^{-1}_{\eta\cdot\omega}\mathcal{I}_{\eta\delta\cdot\omega})\zeta$. □

## A.6 Quantities to Compute the Test Statistics

In this section, specific forms of the quantities to compute the score test statistics are provided. With the local alternatives modeling, the hypotheses are set up as follows:

1. $H^\delta_0 : \delta_0 = 0$ and $H^\eta_0 : \eta_0 = 0$.
2. $H^\delta_a : \delta_0 = \zeta/\sqrt{nT}$ and $H^\eta_a : \eta_0 = \nu/\sqrt{nT}$.

The first joint null hypothesis tests if $W_{nt}$ is endogenous when there is no local parametric misspecification. The restricted ML estimator is denoted by $\tilde{\theta} = (0', 0', \omega')'$ under the joint null hypothesis.

### A.6.1 The Robust Score Test (i.e., $RS^*_\delta(\tilde{\theta})$) when $\eta = (\lambda, \rho)$

If $\eta$ consists of the spatial dependence parameters, $\omega = (\gamma, \beta', \phi_2', \sigma^2_\xi, \alpha')'$. Consider the asymptotic distribution of the test statistic under the joint null, $H^\delta_0$, and $H^\eta_0$. The concentrated log-likelihood function at $\tilde{\theta}$, $\ln L^c_{nT}(\tilde{\theta})$, reduces to

$$\begin{aligned}\ln L^c_{nT}(\tilde{\theta}) = &-\frac{nT}{2}\ln 2\pi - \frac{nT}{2}\ln\tilde{\sigma}^2_\xi - \frac{nT}{2}\ln|\tilde{\Sigma}_\epsilon| \\ &- \frac{1}{2}\mathrm{vec}'(\mathbf{Z}_{nT} - \mathbf{K}_{nT}\tilde{\Phi}_2)(\tilde{\Sigma}^{-1}_\epsilon \otimes \mathbf{J}_{nT})\mathrm{vec}(\mathbf{Z}_{nT} - \mathbf{K}_{nT}\tilde{\Phi}_2) \\ &- \frac{1}{2\tilde{\sigma}^2_\xi}(\mathbf{Y}_{nT} - \tilde{\gamma}\mathbf{Y}_{n,T-1} - \mathbf{X}_{1,nT}\tilde{\beta})'\mathbf{J}_{nT}(\mathbf{Y}_{nT} - \tilde{\gamma}\mathbf{Y}_{n,T-1} - \mathbf{X}_{1,nT}\tilde{\beta}),\end{aligned}$$

which is decomposed into the two parts as

$$\ln L_{nT}^{c}(\tilde{\theta}) = \ln L_{nT}^{C1}(\tilde{\theta}) + \ln L_{nT}^{C2}(\tilde{\theta}),$$

where

$$\begin{cases} \ln L_{nT}^{C1}(\tilde{\theta}) = -\frac{nT}{2}\ln 2\pi - \frac{nT}{2}\ln \tilde{\sigma}_{\xi}^{2} - \frac{1}{2\tilde{\sigma}_{\xi}^{2}}(\mathbf{Y}_{nT} - \tilde{\gamma}\mathbf{Y}_{n,T-1} - \mathbf{X}_{1,nT}\tilde{\beta})'\mathbf{J}_{nT}(\mathbf{Y}_{nT} - \tilde{\gamma}\mathbf{Y}_{n,T-1} - \mathbf{X}_{1,nT}\tilde{\beta}), \\ \ln L_{nT}^{C2}(\tilde{\theta}) = -\frac{nT}{2}\ln|\tilde{\Sigma}_{\epsilon}| - \frac{1}{2}\mathrm{vec}'(\mathbf{Z}_{nT} - \mathbf{K}_{nT}\tilde{\Phi}_{2})(\tilde{\Sigma}_{\epsilon}^{-1} \otimes \mathbf{J}_{nT})\mathrm{vec}(\mathbf{Z}_{nT} - \mathbf{K}_{nT}\tilde{\Phi}_{2}). \end{cases}$$

Using the bias-corrected ML estimator in Qu, Lee, and Yu (2017) under the joint null, one obtains the restricted ML estimators and the residuals, $\Xi_{nT}(\tilde{\theta}) = \mathbf{Y}_{nT} - \tilde{\gamma}\mathbf{Y}_{n,T-1} - \mathbf{X}_{1,nT}\tilde{\beta}$ and $\varepsilon_{nT}(\tilde{\theta}) = \mathbf{Z}_{nT} - \mathbf{K}_{nT}\tilde{\Phi}_{2}$.

The (biased) score functions evaluated at $\tilde{\theta}$ are given as

$$\begin{cases} L_{\delta}(\tilde{\theta}) = \frac{1}{nT}\frac{\partial \ln L_{nT}^{c}(\tilde{\theta})}{\partial \delta} = \frac{1}{nT\tilde{\sigma}_{\xi}^{2}}\left(\varepsilon_{nT}'(\tilde{\theta})\mathbf{J}_{nT}\Xi_{nT}(\tilde{\theta})\right), \\ L_{\eta}(\tilde{\theta}) = \frac{1}{nT}\frac{\partial \ln L_{nT}^{c}(\tilde{\theta})}{\partial \eta} = \frac{1}{nT\tilde{\sigma}_{\xi}^{2}}\left(\Xi_{nT}'(\tilde{\theta})\mathbf{J}_{nT}\mathbf{W}_{nT}\mathbf{Y}_{nT} - \mathrm{tr}(\mathbf{W}_{nT}), (\mathbf{W}_{n,T-1}\mathbf{Y}_{n,T-1})'\mathbf{J}_{nT}\Xi_{nT}(\tilde{\theta})\right)'. \end{cases}$$

Now due to the incidental parameters problem, the bias terms are decomposed by the two parts from the individual- and time heterogeneities as

$$\Delta_{nT}(\tilde{\theta}) = (n-1)a_{1,\theta_0}(\tilde{\theta}) + Ta_{2,\theta_0}(\tilde{\theta}),$$

where

$$a_{1,\theta_0}(\tilde{\theta}) = \begin{pmatrix} -\frac{1}{n-1}\mathrm{tr}\left(\mathbf{G}_{1,nT}\left(\frac{1}{T}1_T 1_T' \otimes J_n\right)\right) \\ \frac{1}{n-1}\left(\mathrm{tr}[\mathbf{G}_{2,nT}\left(\frac{1}{T}1_T 1_T' \otimes J_n\right))], \mathrm{tr}[\mathbf{G}_{3,nT}\left(\frac{1}{T}1_T 1_T' \otimes J_n\right)], 0\right)' \\ 0_{p\times 1} \\ -\mathrm{vec}\left(\frac{1}{T}\left(\sum_{t=1}^{T-1}\sum_{h=1}^{T-t}\tilde{\kappa}'^{(h-1)}\right)\tilde{\Sigma}_{\epsilon}^{-1}\right) \\ -\frac{1}{2\tilde{\sigma}_{\xi}^{2}} \\ -\frac{1}{2}\frac{\partial \ln|\Sigma_{\epsilon 0}|}{\partial \alpha}|_{\tilde{\alpha}} \end{pmatrix}$$

and

$$a_{2,\theta_0} = \begin{pmatrix} -\frac{1}{T}\mathrm{tr}\left(\mathbf{G}_{1,nT}\left(I_T \otimes \frac{1}{n}1_n1_n'\right)\right) \\ 0_{(k_1+2)\times 1} \\ 0_{p\times 1} \\ 0_{p(p+k_2)\times 1} \\ -\frac{1}{2\tilde{\sigma}_\xi^2} \\ -\frac{1}{2}\frac{\partial \ln|\Sigma_{\epsilon 0}|}{\partial \alpha}\Big|_{\tilde{\alpha}} \end{pmatrix},$$

with the order of $\theta = (\lambda, \phi_1', \delta', \phi_2', \sigma_\xi^2, \alpha')'$ and $\phi_1 = (\gamma, \rho, \beta')'$, $\phi_2 = vec(\Phi_2)$ with $\Phi_2 = (\kappa', \Gamma')'$. Denote $\Delta_1 = \sqrt{\frac{n}{T}}a_{1,\theta_0}$ and $\Delta_2 = \sqrt{\frac{T}{n}}a_{2,\theta_0}$. Note that one may regard $\omega = (\gamma, \beta', \sigma_\xi^2)$ because the estimator for $I(\tilde{\theta})$ forms a block diagonal matrix with respect to $(\phi_2', \alpha)'$. Thus, the bias terms necessary for computing the test statistics under the joint null are

$$\begin{aligned}
\Delta_{1,\delta}(\tilde{\theta}) &= 0_{p\times 1}, \\
\Delta_{2,\delta}(\tilde{\theta}) &= 0_{p\times 1}, \\
\Delta_{1,\omega}(\tilde{\theta}) &= \begin{pmatrix} \Delta_{1,\gamma}(\tilde{\theta}) \\ \Delta_{1,\beta}(\tilde{\theta}) \\ \Delta_{1,\sigma_\xi^2}(\tilde{\theta}) \end{pmatrix} = \begin{pmatrix} \mathrm{tr}(\mathbf{G}_{2,nT}\left(\frac{1}{T}1_T1_T' \otimes J_n\right)) \\ 0_{k_1\times 1} \\ \sqrt{\frac{n}{T}}\left(-\frac{1}{2\tilde{\sigma}_\xi^2}\right) \end{pmatrix}, \\
\Delta_{2,\omega}(\tilde{\theta}) &= \begin{pmatrix} \Delta_{2,\gamma}(\tilde{\theta}) \\ \Delta_{2,\beta}(\tilde{\theta}) \\ \Delta_{2,\sigma_\xi^2}(\tilde{\theta}) \end{pmatrix} \\
&= \begin{pmatrix} 0 \\ 0_{k_1\times 1} \\ \sqrt{\frac{T}{n}}\left(-\frac{1}{2\tilde{\sigma}_\xi^2}\right) \end{pmatrix}, \\
\Delta_{1,\eta}(\tilde{\theta}) &= \begin{pmatrix} \Delta_{1,\lambda}(\tilde{\theta}) \\ \Delta_{1,\rho}(\tilde{\theta}) \end{pmatrix} \\
&= \frac{1}{\sqrt{nT}}\begin{pmatrix} -\mathrm{tr}(\mathbf{G}_{1,nT}\left(\frac{1}{T}1_T1_T' \otimes J_n\right)) \\ \mathrm{tr}(\mathbf{G}_{3,nT}\left(\frac{1}{T}1_T1_T' \otimes J_n\right)) \end{pmatrix}, \\
\Delta_{2,\eta}(\tilde{\theta}) &= \begin{pmatrix} \Delta_{2,\lambda}(\tilde{\theta}) \\ \Delta_{2,\rho}(\tilde{\theta}) \end{pmatrix} \\
&= \frac{1}{\sqrt{nT}}\begin{pmatrix} -\mathrm{tr}(\mathbf{G}_{1,nT}\left(I_T \otimes \frac{1}{n}1_n1_n'\right)) \\ 0 \end{pmatrix}.
\end{aligned}$$

Then the bias-corrected score functions are computed as

$$
\begin{aligned}
\sqrt{nT}C_\delta(\tilde\theta) &= \sqrt{nT}L_\delta(\tilde\theta) - (\underbrace{\Delta_{1,\delta}(\tilde\theta)}_{=0} + \underbrace{\Delta_{2,\delta}(\tilde\theta)}_{=0}) + \underbrace{I_{\delta\omega}(\tilde\theta)}_{=0} I^{-1}_{\omega\omega}(\tilde\theta)(\Delta_{1,\omega}(\tilde\theta) + \Delta_{2,\omega}(\tilde\theta)) \\
&= \sqrt{nT}L_\delta(\tilde\theta) = \frac{1}{\sqrt{nT}\tilde\sigma^2_\xi}\left(\varepsilon'_{nT}(\tilde\theta)\mathbf{J}_{nT}\Xi_{nT}(\tilde\theta)\right), \\
\sqrt{nT}C_\eta(\tilde\theta) &= \sqrt{nT}L_\eta(\tilde\theta) - (\Delta_{1,\eta}(\tilde\theta) + \Delta_{2,\eta}(\tilde\theta)) + I_{\eta\omega}(\tilde\theta)I^{-1}_{\omega\omega}(\tilde\theta)(\Delta_{1,\omega}(\tilde\theta) + \Delta_{2,\omega}(\tilde\theta)) \\
&= \frac{1}{\sqrt{nT}\tilde\sigma^2_\xi}\left(\Xi'_{nT}(\tilde\theta)\mathbf{J}_{nT}\mathbf{W}_{nT}\mathbf{Y}_{nT} - tr(\mathbf{W}_{nT}), (\mathbf{W}_{n,T-1}\mathbf{Y}_{n,T-1})'\mathbf{J}_{nT}\Xi_{nT}(\tilde\theta)\right)' \\
&\quad - \left( - \operatorname{tr}\left(\mathbf{G}_{1,nT}\left(\frac{1}{T}1_T1'_T \otimes J_n\right)\right) - \operatorname{tr}\left(\mathbf{G}_{1,nT}\left(I_T \otimes \frac{1}{n}1_n1'_n\right)\right), \operatorname{tr}\left(\mathbf{G}_{3,nT}\left(\frac{1}{T}1_T1'_T \otimes J_n\right)\right)\right)', \\
&\quad + I_{\eta\omega}(\tilde\theta)I^{-1}_{\omega\omega}(\tilde\theta)\begin{pmatrix} \operatorname{tr}\left(\mathbf{G}_{2,nT}\left(\frac{1}{T}1_T1'_T \otimes J_n\right)\right) \\ 0_{k_1\times 1} \\ -\frac{1}{2\tilde\sigma^2_\xi}\left(\sqrt{\frac{n}{T}} + \sqrt{\frac{T}{n}}\right) \end{pmatrix},
\end{aligned}
$$

where the consistent estimator for the information matrix is given as

$$
I(\tilde\theta) = \frac{1}{nT\tilde\sigma^2_\xi}\begin{pmatrix}
\iota_{\lambda\lambda}(\tilde\theta) & * & * & * & * & * \\
\iota_{\phi_1\lambda}(\tilde\theta) & \iota_{\phi_1\phi_1}(\tilde\theta) & * & * & * & * \\
\iota_{\delta\lambda}(\tilde\theta) & 0_{p\times(k_1+2)} & \iota_{\delta\delta}(\tilde\theta) & * & * & * \\
0_{k_{\phi_2}\times 1} & 0_{k_{\phi_2}\times(k_1+2)} & 0_{k_{\phi_2}\times p} & \iota_{\phi_2\phi_2}(\tilde\theta) & * & * \\
\iota_{\sigma^2_\xi\lambda}(\tilde\theta) & 0_{1\times(k_1+2)} & 0_{1\times p} & 0_{1\times k_{\phi_2}} & \iota_{\sigma^2_\xi\sigma^2_\xi}(\tilde\theta) & * \\
0_{J\times 1} & 0_{J\times(k_1+2)} & 0_{J\times p} & 0_{J\times k_{\phi_2}} & 0_{J\times 1} & \iota_{\alpha\alpha}(\tilde\theta)
\end{pmatrix},
$$

with

$$
\begin{aligned}
\iota_{\lambda\lambda}(\tilde{\theta}) &= (\mathbf{W}_{nT}\mathbf{Y}_{nT})'\mathbf{J}_{nT}\mathbf{W}_{nT}\mathbf{Y}_{nT} + \tilde{\sigma}_{\xi}^{2}\mathrm{tr}(\mathbf{G}_{1,nT}^{2}),\\
\iota_{\phi_1\lambda}(\tilde{\theta}) &= \mathbf{R}_{nT}'\mathbf{J}_{nT}\mathbf{W}_{nT}\mathbf{Y}_{nT},\\
\iota_{\phi_1\phi_1}(\tilde{\theta}) &= \mathbf{R}_{nT}'\mathbf{J}_{nT}\mathbf{R}_{nT},\\
\iota_{\delta\lambda}(\tilde{\theta}) &= \varepsilon_{nT}'(\tilde{\theta})\mathbf{J}_{nT}\mathbf{W}_{nT}\mathbf{Y}_{nT},\\
\iota_{\delta\delta}(\tilde{\theta}) &= \varepsilon_{nT}'(\tilde{\theta})\mathbf{J}_{nT}\varepsilon_{nT}'(\tilde{\theta}),\\
\iota_{\phi_2\phi_2}(\tilde{\theta}) &= (\tilde{\sigma}_{\xi}^{2}\tilde{\Sigma}_{\epsilon}^{-1}) \otimes (\mathbf{K}_{nT}'\mathbf{J}_{nT}\mathbf{K}_{nT}),\\
\iota_{\sigma_{\xi}^{2}\lambda}(\tilde{\theta}) &= \mathrm{tr}(\mathbf{G}_{1,nT}),\\
\iota_{\sigma_{\xi}^{2}\sigma_{\xi}^{2}}(\tilde{\theta}) &= \frac{1}{\tilde{\sigma}_{\xi}^{2}}\left(\frac{nT}{2} - T - n + 1\right),\\
\iota_{\alpha\alpha,kj}(\tilde{\theta}) &= \frac{nT}{2}\tilde{\sigma}_{\xi}^{2}\mathrm{tr}\left(\tilde{\Sigma}_{\epsilon}^{-1}\frac{\partial\Sigma_{\epsilon}}{\partial\alpha_k}(\tilde{\theta})\tilde{\Sigma}_{\epsilon}^{-1}\frac{\partial\Sigma_{\varepsilon}}{\partial\alpha_j}(\tilde{\theta})\right) \quad \text{for } k, j = 1, \ldots, J.
\end{aligned}
$$

Accordingly,

$$I_{\delta\delta}(\tilde\theta) = \iota_{\delta\delta}(\tilde\theta)/(nT\tilde\sigma_\xi^2) = \varepsilon'_{nT}(\tilde\theta)\mathbf{J}_{nT}\varepsilon'_{nT}(\tilde\theta)/(nT\tilde\sigma_\xi^2),$$

$$I_{\delta\omega}(\tilde\theta) = \iota_{\delta\omega}(\tilde\theta)/(nT\tilde\sigma_\xi^2) = \begin{pmatrix}\iota_{\delta\gamma}(\tilde\theta) & \iota_{\delta\beta'}(\tilde\theta) & \iota_{\delta\sigma_\xi^2}(\tilde\theta)\end{pmatrix}/(nT\tilde\sigma_\xi^2)$$

$$= \begin{pmatrix}0_{p\times 1} & 0_{p\times k_1} & 0\end{pmatrix}/(nT\tilde\sigma_\xi^2) = 0_{p\times(k_1+2)},$$

$$I_{\delta\eta}(\tilde\theta) = \iota_{\delta\eta}(\tilde\theta)/(nT\tilde\sigma_\xi^2) = \begin{pmatrix}\iota_{\delta\lambda}(\tilde\theta) & \iota_{\delta\rho}(\tilde\theta)\end{pmatrix}/(nT\tilde\sigma_\xi^2)$$

$$= \begin{pmatrix}\varepsilon'_{nT}(\tilde\theta)\mathbf{J}_{nT}\mathbf{W}_{nT}\mathbf{Y}_{nT} & 0_{p\times 1}\end{pmatrix}/(nT\tilde\sigma_\xi^2),$$

$$I_{\omega\omega}(\tilde\theta) = \iota_{\omega\omega}(\tilde\theta)/(nT\tilde\sigma_\xi^2) = \begin{pmatrix}\iota_{\gamma\gamma}(\tilde\theta) & \iota_{\gamma\beta'}(\tilde\theta) & \iota_{\gamma\sigma_\xi^2}(\tilde\theta)\\ \iota_{\beta\gamma}(\tilde\theta) & \iota_{\beta\beta'}(\tilde\theta) & \iota_{\beta\sigma_\xi^2}(\tilde\theta)\\ \iota_{\sigma_\xi^2\gamma}(\tilde\theta) & \iota_{\sigma_\xi^2\beta'}(\tilde\theta) & \iota_{\sigma_\xi^2\sigma_\xi^2}(\tilde\theta)\end{pmatrix}/(nT\tilde\sigma_\xi^2)$$

$$= \begin{pmatrix}\mathbf{R}'_{nT}\mathbf{J}_{nT}\mathbf{R}_{nT(1,1)} & (\mathbf{R}'_{nT}\mathbf{J}_{nT}\mathbf{R}_{nT})_{(1,3:k_1+2)} & 0\\ * & (\mathbf{R}'_{nT}\mathbf{J}_{nT}\mathbf{R}_{nT})_{(3:k_1+2,3:k_1+2)} & 0_{k_1\times 1}\\ * & * & \frac{1}{\tilde\sigma_\xi^2}\left(\frac{nT}{2}-T-n+1\right)\end{pmatrix}/(nT\tilde\sigma_\xi^2),$$

$$I_{\eta\omega}(\tilde\theta) = \iota_{\eta\omega}(\tilde\theta)/(nT\tilde\sigma_\xi^2) = \begin{pmatrix}\iota_{\lambda\gamma}(\tilde\theta) & \iota_{\lambda\beta'}(\tilde\theta) & \iota_{\lambda\sigma_\xi^2}(\tilde\theta)\\ \iota_{\rho\gamma}(\tilde\theta) & \iota_{\rho\beta'}(\tilde\theta) & \iota_{\rho\sigma_\xi^2}(\tilde\theta)\end{pmatrix}/(nT\tilde\sigma_\xi^2)$$

$$= \begin{pmatrix}(\mathbf{W}_{nT}\mathbf{Y}_{nT})'\mathbf{J}_{nT}\mathbf{R}_{nT(1,1)} & (\mathbf{R}'_{nT}\mathbf{J}_{nT}\mathbf{W}_{1,nT}\mathbf{Y}_{nT}\tilde\beta_{(3:k_1+2,1)})' & \mathrm{tr}(\mathbf{G}_{1,nT})\\ \mathbf{R}'_{nT}\mathbf{J}_{nT}\mathbf{R}_{nT(2,1)} & (\mathbf{R}'_{nT}\mathbf{J}_{nT}\mathbf{R}_{nT})_{(2,3:k_1+2)} & 0\end{pmatrix}/(nT\tilde\sigma_\xi^2),$$

$$I_{\eta\eta}(\tilde\theta) = \iota_{\eta\eta}(\tilde\theta)/(nT\tilde\sigma_\xi^2) = \begin{pmatrix}\iota_{\lambda\lambda}(\tilde\theta) & \iota_{\lambda\rho}(\tilde\theta)\\ * & \iota_{\rho\rho}(\tilde\theta)\end{pmatrix}/(nT\tilde\sigma_\xi^2)$$

$$= \begin{pmatrix}(\mathbf{W}_{nT}\mathbf{Y}_{nT})'\mathbf{J}_{nT}\mathbf{W}_{nT}\mathbf{Y}_{nT} + \tilde\sigma_\xi^2\mathrm{tr}(\mathbf{G}^2_{1,nT}) & (\mathbf{W}_{nT}\mathbf{Y}_{nT})'\mathbf{J}_{nT}\mathbf{R}_{nT(1,2)}\\ * & \mathbf{R}'_{nT}\mathbf{J}_{nT}\mathbf{R}_{nT(2,2)}\end{pmatrix}/(nT\tilde\sigma_\xi^2),$$

where $(A)_{(i,j:k)}$ represents the entries in a matrix $A$ located in $i^{th}$ row, $j^{th}$ to $k^{th}$ columns.

### A.6.2 The Robust Score Test (i.e., $RS^*_\delta(\tilde\theta)$) when $\eta = (\lambda, \gamma, \rho)$

If $\eta$ consists of both the spatial dependence parameters and the time dynamic dependence parameter, $\omega = (\beta', \phi'_2, \sigma^2_\xi, \alpha')'$. Consider the asymptotic distribution of the test statistic under the joint null,

$H_0^\delta$, and $H_0^\eta$. The concentrated log-likelihood function at $\tilde{\theta}$, $\ln L_{nT}^c(\tilde{\theta})$, reduces to

$$\begin{aligned}\ln L_{nT}^c(\tilde{\theta}) = &-\frac{nT}{2}\ln 2\pi - \frac{nT}{2}\ln\tilde{\sigma}_\xi^2 - \frac{nT}{2}\ln|\tilde{\Sigma}_\epsilon| \\ &-\frac{1}{2}\mathrm{vec}'(\mathbf{Z}_{nT}-\mathbf{K}_{nT}\tilde{\Phi}_2)(\tilde{\Sigma}_\epsilon^{-1}\otimes\mathbf{J}_{nT})\mathrm{vec}(\mathbf{Z}_{nT}-\mathbf{K}_{nT}\tilde{\Phi}_2) \\ &-\frac{1}{2\tilde{\sigma}_\xi^2}(\mathbf{Y}_{nT}-\mathbf{X}_{1nT}\tilde{\beta})'\mathbf{J}_{nT}(\mathbf{Y}_{nT}-\mathbf{X}_{1nT}\tilde{\beta}),\end{aligned}$$

which is decomposed into the two parts as

$$\ln L_{nT}^c(\tilde{\theta}) = \ln L_{nT}^{C1}(\tilde{\theta}) + \ln L_{nT}^{C2}(\tilde{\theta}),$$

where

$$\begin{cases}\ln L_{nT}^{C1}(\tilde{\theta}) = -\frac{nT}{2}\ln 2\pi - \frac{nT}{2}\ln\tilde{\sigma}_\xi^2 - \frac{1}{2\tilde{\sigma}_\xi^2}(\mathbf{Y}_{nT}-\mathbf{X}_{1,nT}\tilde{\beta})'\mathbf{J}_{nT}(\mathbf{Y}_{nT}-\mathbf{X}_{1,nT}\tilde{\beta}), \\ \ln L_{nT}^{C2}(\tilde{\theta}) = -\frac{nT}{2}\ln|\tilde{\Sigma}_\epsilon| - \frac{1}{2}\mathrm{vec}'(\mathbf{Z}_{nT}-\mathbf{K}_{nT}\tilde{\Phi}_2)(\tilde{\Sigma}_\epsilon^{-1}\otimes\mathbf{J}_{nT})\mathrm{vec}(\mathbf{Z}_{nT}-\mathbf{K}_{nT}\tilde{\Phi}_2).\end{cases}$$

Using the bias-corrected ML estimator in Qu, Lee, and Yu under the joint null, one obtains the restricted ML estimators and the residuals, $\Xi_{nT}(\tilde{\theta}) = \mathbf{Y}_{nT} - \mathbf{X}_{1,nT}\tilde{\beta}$ and $\varepsilon_{nT}(\tilde{\theta}) = \mathbf{Z}_{nT} - \mathbf{K}_{nT}\tilde{\Phi}_2$.

The (biased) score functions evaluated at $\tilde{\theta}$ are given as

$$\begin{cases}L_\delta(\tilde{\theta}) = \frac{1}{nT}\frac{\partial\ln L_{nT}^c(\tilde{\theta})}{\partial\delta} = \frac{1}{nT\tilde{\sigma}_\xi^2}\left(\varepsilon_{nT}'(\tilde{\theta})\mathbf{J}_{nT}\Xi_{nT}(\tilde{\theta})\right), \\ L_\eta(\tilde{\theta}) = \frac{1}{nT}\frac{\partial\ln L_{nT}^c(\tilde{\theta})}{\partial\eta} = \frac{1}{nT\tilde{\sigma}_\xi^2}\Big(\Xi_{nT}'(\tilde{\theta})\mathbf{J}_{nT}\mathbf{W}_{1,nT}\mathbf{Y}_{nT} - \mathrm{tr}(\mathbf{W}_{1,nT}), \mathbf{Y}_{n,T-1}'\mathbf{J}_{nT}\Xi_{nT}(\tilde{\theta}), \\ \qquad\qquad (\mathbf{W}_{n,T-1}\mathbf{Y}_{n,T-1})'\mathbf{J}_{nT}\Xi_{nT}(\tilde{\theta})\Big)'.\end{cases}$$

Now due to the incidental parameters problem, the bias terms are decomposed by the two parts from the individual- and time heterogeneities as

$$\Delta_{nT}(\tilde{\theta}) = (n-1)a_{1,\theta_0}(\tilde{\theta}) + Ta_{2,\theta_0}(\tilde{\theta}),$$

where

$$a_{1,\theta_0}(\tilde{\theta}) = \begin{pmatrix} -\frac{1}{n-1}\mathrm{tr}\left(\mathbf{W}_{nT}\left(\frac{1}{T}1_T 1_T' \otimes J_n\right)\right) \\ \frac{1}{n-1}\left(\mathrm{tr}[\mathbf{I}_{nT,-1}\left(\frac{1}{T}1_T 1_T' \otimes J_n\right))], \mathrm{tr}[\mathbf{W}_{n,T-1}\left(\frac{1}{T}1_T 1_T' \otimes J_n\right)], 0\right)' \\ 0_{p\times 1} \\ -\mathrm{vec}\left(\frac{1}{T}\left(\sum_{t=1}^{T-1}\sum_{h=1}^{T-t}\tilde{\kappa}'^{(h-1)}\right)\tilde{\Sigma}_\epsilon^{-1}\right) \\ -\frac{1}{2\tilde{\sigma}_\xi^2} \\ -\frac{1}{2}\frac{\partial \ln|\Sigma_{\epsilon 0}|}{\partial\alpha}|_{\tilde{\alpha}} \end{pmatrix}$$

and

$$a_{2,\theta_0} = \begin{pmatrix} -\frac{1}{T}\mathrm{tr}\left(\mathbf{W}_{nT}\left(I_T \otimes \frac{1}{n}1_n 1_n'\right)\right) \\ 0_{(k_1+2)\times 1} \\ 0_{p\times 1} \\ 0_{p(p+k_2)\times 1} \\ -\frac{1}{2\tilde{\sigma}_\xi^2} \\ -\frac{1}{2}\frac{\partial \ln|\Sigma_{\epsilon 0}|}{\partial\alpha}|_{\tilde{\alpha}} \end{pmatrix},$$

with the order of $\theta = (\lambda, \phi_1', \delta', \phi_2', \sigma_\xi^2, \alpha')'$ and $\phi_1 = (\gamma, \rho, \beta')'$, $\phi_2 = vec(\Phi_2)$ with $\Phi_2 = (\kappa', \Gamma')'$. Denote $\Delta_1 = \sqrt{\frac{n}{T}}a_{1,\theta_0}$ and $\Delta_2 = \sqrt{\frac{T}{n}}a_{2,\theta_0}$. Denote $\Delta_1 = \sqrt{\frac{n}{T}}a_{1,\theta_0}$ and $\Delta_2 = \sqrt{\frac{T}{n}}a_{2,\theta_0}$. Note that one may regard $\omega = (\beta', \sigma_\xi^2)$ because the estimator for $I(\tilde{\theta})$ forms a block diagonal matrix with respect to $(\phi_2', \alpha)'$. Thus, the bias terms necessary for computing the test statistics under the joint null are

$$
\begin{aligned}
\Delta_{1,\delta}(\tilde{\theta}) &= 0_{p\times 1}, \\
\Delta_{2,\delta}(\tilde{\theta}) &= 0_{p\times 1}, \\
\Delta_{1,\omega}(\tilde{\theta}) &= \begin{pmatrix} \Delta_{1,\beta}(\tilde{\theta}) \\ \Delta_{1,\sigma_\xi^2}(\tilde{\theta}) \end{pmatrix} = \begin{pmatrix} 0_{k_1\times 1} \\ \sqrt{\frac{n}{T}}\left(-\frac{1}{2\tilde{\sigma}_\xi^2}\right) \end{pmatrix}, \\
\Delta_{2,\omega}(\tilde{\theta}) &= \begin{pmatrix} \Delta_{2,\beta}(\tilde{\theta}) \\ \Delta_{2,\sigma_\xi^2}(\tilde{\theta}) \end{pmatrix} \\
&= \begin{pmatrix} 0_{k_1\times 1} \\ \sqrt{\frac{T}{n}}\left(-\frac{1}{2\tilde{\sigma}_\xi^2}\right)) \end{pmatrix}, \\
\Delta_{1,\eta}(\tilde{\theta}) &= \begin{pmatrix} \Delta_{1,\lambda}(\tilde{\theta}) \\ \Delta_{1,\gamma}(\tilde{\theta}) \\ \Delta_{1,\rho}(\tilde{\theta}) \end{pmatrix} \\
&= \frac{1}{\sqrt{nT}} \begin{pmatrix} -\mathrm{tr}(\mathbf{W}_{nT}\left(\frac{1}{T}1_T 1_T' \otimes J_n\right)) \\ \mathrm{tr}(\mathbf{I}_{nT,-1}\left(\frac{1}{T}1_T 1_T' \otimes J_n\right)) \\ \mathrm{tr}(\mathbf{W}_{n,T-1}\left(\frac{1}{T}1_T 1_T' \otimes J_n\right)) \end{pmatrix}, \\
\Delta_{2,\eta}(\tilde{\theta}) &= \begin{pmatrix} \Delta_{2,\lambda}(\tilde{\theta}) \\ \Delta_{2,\gamma}(\tilde{\theta}) \\ \Delta_{2,\rho}(\tilde{\theta}) \end{pmatrix} \\
&= \frac{1}{\sqrt{nT}} \begin{pmatrix} -\mathrm{tr}(\mathbf{W}_{nT}\left(I_T \otimes \frac{1}{n}1_n 1_n'\right)) \\ 0 \\ 0 \end{pmatrix}.
\end{aligned}
$$

Then the bias-corrected score functions are computed as

$$
\begin{aligned}
\sqrt{nT}C_\delta(\tilde\theta) &= \sqrt{nT}L_\delta(\tilde\theta) - (\underbrace{\Delta_{1,\delta}(\tilde\theta)}_{=0} + \underbrace{\Delta_{2,\delta}(\tilde\theta)}_{=0}) + \underbrace{I_{\delta\omega}(\tilde\theta)}_{=0} I^{-1}_{\omega\omega}(\tilde\theta)(\Delta_{1,\omega}(\tilde\theta) + \Delta_{2,\omega}(\tilde\theta)) \\
&= \sqrt{nT}L_\delta(\tilde\theta) = \frac{1}{\sqrt{nT}\tilde\sigma^2_\xi}\left(\varepsilon'_{nT}(\tilde\theta)\mathbf{J}_{nT}\Xi_{nT}(\tilde\theta)\right), \\
\sqrt{nT}C_\eta(\tilde\theta) &= \sqrt{nT}L_\eta(\tilde\theta) - (\Delta_{1,\eta}(\tilde\theta) + \Delta_{2,\eta}(\tilde\theta)) + I_{\eta\omega}(\tilde\theta)I^{-1}_{\omega\omega}(\tilde\theta)(\Delta_{1,\omega}(\tilde\theta) + \Delta_{2,\omega}(\tilde\theta)) \\
&= \frac{1}{\sqrt{nT}\tilde\sigma^2_\xi}\left(\Xi'_{nT}(\tilde\theta)\mathbf{J}_{nT}\mathbf{W}_{nT}\mathbf{Y}_{nT} - tr(\mathbf{W}_{nT}), \mathbf{Y}'_{n,T-1}\mathbf{J}_{nT}\Xi_{nT}(\tilde\theta), (\mathbf{W}_{n,T-1}\mathbf{Y}_{n,T-1})'\mathbf{J}_{nT}\Xi_{nT}(\tilde\theta)\right)' \\
&\quad - \Bigg( - \operatorname{tr}\left(\mathbf{W}_{nT}\left(\frac{1}{T}1_T1'_T \otimes J_n\right)\right) - \operatorname{tr}\left(\mathbf{W}_{nT}\left(I_T \otimes \frac{1}{n}1_n1'_n\right)\right), \\
&\qquad\qquad \operatorname{tr}\left(\mathbf{I}_{nT,-1}\left(\frac{1}{T}1_T1'_T \otimes J_n\right)\right), \operatorname{tr}\left(\mathbf{W}_{n,T-1}\left(\frac{1}{T}1_T1'_T \otimes J_n\right)\right)\Bigg)' \\
&\quad + I_{\eta\omega}(\tilde\theta)I^{-1}_{\omega\omega}(\tilde\theta)\begin{pmatrix} 0_{k_1\times 1} \\ -\frac{1}{2\tilde\sigma^2_\xi}\left(\sqrt{\frac{n}{T}} + \sqrt{\frac{T}{n}}\right)\end{pmatrix},
\end{aligned}
$$

where the consistent estimator for the information matrix is given as

$$
I(\tilde\theta) = \frac{1}{nT\tilde\sigma^2_\xi}\begin{pmatrix}
\iota_{\lambda\lambda}(\tilde\theta) & * & * & * & * & * \\
\iota_{\phi_1\lambda}(\tilde\theta) & \iota_{\phi_1\phi_1}(\tilde\theta) & * & * & * & * \\
\iota_{\delta\lambda}(\tilde\theta) & 0_{p\times(k_1+2)} & \iota_{\delta\delta}(\tilde\theta) & * & * & * \\
0_{k_{\phi_2}\times 1} & 0_{k_{\phi_2}\times(k_1+2)} & 0_{k_{\phi_2}\times p} & \iota_{\phi_2\phi_2}(\tilde\theta) & * & * \\
\iota_{\sigma^2_\xi\lambda}(\tilde\theta) & 0_{1\times(k_1+2)} & 0_{1\times p} & 0_{1\times k_{\phi_2}} & \iota_{\sigma^2_\xi\sigma^2_\xi}(\tilde\theta) & * \\
0_{J\times 1} & 0_{J\times(k_1+2)} & 0_{J\times p} & 0_{J\times k_{\phi_2}} & 0_{J\times 1} & \iota_{\alpha\alpha}(\tilde\theta)
\end{pmatrix},
$$

with

$$
\begin{aligned}
\iota_{\lambda\lambda}(\tilde{\theta}) &= (\mathbf{W}_{nT}\mathbf{Y}_{nT})'\mathbf{J}_{nT}\mathbf{W}_{nT}\mathbf{Y}_{nT} + \tilde{\sigma}_{\xi}^{2}\mathrm{tr}(\mathbf{W}_{nT}^{2}),\\
\iota_{\phi_1\lambda}(\tilde{\theta}) &= \mathbf{R}_{nT}'\mathbf{J}_{nT}\mathbf{W}_{nT}\mathbf{Y}_{nT},\\
\iota_{\phi_1\phi_1}(\tilde{\theta}) &= \mathbf{R}_{nT}'\mathbf{J}_{nT}\mathbf{R}_{nT},\\
\iota_{\delta\lambda}(\tilde{\theta}) &= \varepsilon_{nT}'(\tilde{\theta})\mathbf{J}_{nT}\mathbf{W}_{1,nT}\mathbf{Y}_{nT},\\
\iota_{\delta\delta}(\tilde{\theta}) &= \varepsilon_{nT}'(\tilde{\theta})\mathbf{J}_{nT}\varepsilon_{nT}'(\tilde{\theta}),\\
\iota_{\phi_2\phi_2}(\tilde{\theta}) &= (\tilde{\sigma}_{\xi}^{2}\tilde{\Sigma}_{\epsilon}^{-1}) \otimes (\mathbf{K}_{nT}'\mathbf{J}_{nT}\mathbf{K}_{nT}),\\
\iota_{\sigma_{\xi}^{2}\lambda}(\tilde{\theta}) &= \mathrm{tr}(\mathbf{W}_{nT}),\\
\iota_{\sigma_{\xi}^{2}\sigma_{\xi}^{2}}(\tilde{\theta}) &= \frac{1}{\tilde{\sigma}_{\xi}^{2}}\left(\frac{nT}{2} - T - n + 1\right),\\
\iota_{\alpha\alpha,kj}(\tilde{\theta}) &= \frac{nT}{2}\tilde{\sigma}_{\xi}^{2}\mathrm{tr}\left(\tilde{\Sigma}_{\epsilon}^{-1}\frac{\partial\Sigma_{\epsilon}}{\partial\alpha_k}(\tilde{\theta})\tilde{\Sigma}_{\epsilon}^{-1}\frac{\partial\Sigma_{\epsilon}}{\partial\alpha_j}(\tilde{\theta})\right) \quad \text{for } k, j = 1, \ldots, J.
\end{aligned}
$$

Accordingly,

$$I_{\delta\delta}(\tilde{\theta}) = \iota_{\delta\delta}(\tilde{\theta})/(nT\tilde{\sigma}_\xi^2) = \varepsilon'_{nT}(\tilde{\theta})\mathbf{J}_{nT}\varepsilon'_{nT}(\tilde{\theta})/(nT\sigma_\xi^2),$$

$$\begin{aligned}
I_{\delta\omega}(\tilde{\theta}) = \iota_{\delta\omega}(\tilde{\theta})/(nT\tilde{\sigma}_\xi^2) &= \begin{pmatrix}\iota_{\delta\beta'}(\tilde{\theta}) & \iota_{\delta\sigma_\xi^2}(\tilde{\theta})\end{pmatrix}/(nT\tilde{\sigma}_\xi^2) \\
&= \begin{pmatrix}0_{p\times k_1} & 0\end{pmatrix}/(nT\tilde{\sigma}_\xi^2) = 0_{p\times(k_1+1)},
\end{aligned}$$

$$\begin{aligned}
I_{\delta\eta}(\tilde{\theta}) = \iota_{\delta\eta}(\tilde{\theta})/(nT\tilde{\sigma}_\xi^2) &= \begin{pmatrix}\iota_{\delta\lambda}(\tilde{\theta}) & \iota_{\delta\gamma}(\tilde{\theta}) & \iota_{\delta\rho}(\tilde{\theta})\end{pmatrix}/(nT\tilde{\sigma}_\xi^2) \\
&= \begin{pmatrix}\varepsilon'_{nT}(\tilde{\theta})\mathbf{J}_{nT}\mathbf{W}_{nT}\mathbf{Y}_{nT} & 0_{p\times1} & 0_{p\times1}\end{pmatrix}/(nT\tilde{\sigma}_\xi^2),
\end{aligned}$$

$$\begin{aligned}
I_{\omega\omega}(\tilde{\theta}) = \iota_{\omega\omega}(\tilde{\theta})/(nT\tilde{\sigma}_\xi^2) &= \begin{pmatrix}\iota_{\beta\beta'}(\tilde{\theta}) & \iota_{\beta\sigma_\xi^2}(\tilde{\theta}) \\ \iota_{\sigma_\xi^2\beta'}(\tilde{\theta}) & \iota_{\sigma_\xi^2\sigma_\xi^2}(\tilde{\theta})\end{pmatrix}/(nT\tilde{\sigma}_\xi^2) \\
&= \begin{pmatrix}(\mathbf{R}'_{nT}\mathbf{J}_{nT}\mathbf{R}_{nT})_{(3:k_1+2,3:k_1+2)} & 0_{k_1\times1} \\ 0_{1\times k_1} & \frac{1}{\tilde{\sigma}_\xi^2}\left(\frac{nT}{2} - T - n + 1\right)\end{pmatrix}/(nT\tilde{\sigma}_\xi^2),
\end{aligned}$$

$$\begin{aligned}
I_{\eta\omega}(\tilde{\theta}) = \iota_{\eta\omega}(\tilde{\theta})/(nT\tilde{\sigma}_\xi^2) &= \begin{pmatrix}\iota_{\lambda\beta'}(\tilde{\theta}) & \iota_{\lambda\sigma_\xi^2}(\tilde{\theta}) \\ \iota_{\gamma\beta'}(\tilde{\theta}) & \iota_{\gamma\sigma_\xi^2}(\tilde{\theta}) \\ \iota_{\rho\beta'}(\tilde{\theta}) & \iota_{\rho\sigma_\xi^2}(\tilde{\theta})\end{pmatrix}/(nT\tilde{\sigma}_\xi^2) \\
&= \begin{pmatrix}(\mathbf{R}'_{nT}\mathbf{J}_{nT}\mathbf{W}_{nT}\mathbf{Y}_{nT}\tilde{\beta}_{(3:k_1+2,1)})' & \mathrm{tr}(\mathbf{W}_{nT}) \\ (\mathbf{R}'_{nT}\mathbf{J}_{nT}\mathbf{R}_{nT})_{(1,3:k_1+2)} & 0 \\ (\mathbf{R}'_{nT}\mathbf{J}_{nT}\mathbf{R}_{nT})_{(2,3:k_1+2)} & 0\end{pmatrix}/(nT\tilde{\sigma}_\xi^2),
\end{aligned}$$

$$I_{\eta\eta}(\tilde{\theta}) = \iota_{\eta\eta}(\tilde{\theta})/(nT\tilde{\sigma}_\xi^2) = \begin{pmatrix}\iota_{\lambda\lambda}(\tilde{\theta}) & \iota_{\lambda\gamma}(\tilde{\theta}) & \iota_{\lambda\rho}(\tilde{\theta}) \\ * & \iota_{\gamma\gamma}(\tilde{\theta}) & \iota_{\gamma\rho}(\tilde{\theta}) \\ * & * & \iota_{\rho\rho}(\tilde{\theta})\end{pmatrix}/(nT\tilde{\sigma}_\xi^2)$$

$$= \begin{pmatrix}(\mathbf{W}_{nT}\mathbf{Y}_{nT})'\mathbf{J}_{nT}\mathbf{W}_{nT}\mathbf{Y}_{nT} + \tilde{\sigma}_\xi^2\mathrm{tr}(\mathbf{W}_{nT}^2) & (\mathbf{W}_{nT}\mathbf{Y}_{nT})'\mathbf{J}_{nT}\mathbf{R}_{nT(1,1)} & (\mathbf{W}_{nT}\mathbf{Y}_{nT})'\mathbf{J}_{nT}\mathbf{R}_{nT(1,2)} \\ * & \mathbf{R}'_{nT}\mathbf{J}_{nT}\mathbf{R}_{nT(1,1)} & \mathbf{R}'_{nT}\mathbf{J}_{nT}\mathbf{R}_{nT(1,2)} \\ * & * & \mathbf{R}'_{nT}\mathbf{J}_{nT}\mathbf{R}_{nT(2,2)}\end{pmatrix}/(nT\tilde{\sigma}_\xi^2),$$

where $(A)_{(i,j:k)}$ represents the entries in a matrix $A$ located in $i^{th}$ row, $j^{th}$ to $k^{th}$ columns.

#### A.6.3 The $C(\alpha)$ test at the maximum likelihood estimates under $H_0^\delta$ (i.e., $C(\alpha)|_{\hat{\theta}}$)

Let $\theta = (\delta', \Psi')'$ where $\Psi = (\lambda, \phi_1', \phi_2', \sigma_\xi^2, \alpha')'$ and $\hat{\theta} = \underset{\theta:\,\delta=0}{\mathrm{argmax}} \ln L_{nT}^c(\theta)$ be the restricted maximum likelihood (ML) estimates under $H_0^\delta$. Remark that the estimator for the information matrix at $\hat{\theta}$, $I(\hat{\theta})$ forms a block diagonal matrix with respect to $(\phi_2', \alpha')'$, where one thus may regard $\Psi = (\lambda, \phi_1', \sigma_\xi^2)'$.

The concentrated log-likelihood under $H_0^\delta$ is then

$$\begin{aligned}\ln L_{nT}^c(\hat\theta) &= -\frac{nT}{2}\ln 2\pi + \ln|\mathbf{S}_{nT}(\hat\lambda,\hat\gamma,\hat\rho)| - \frac{nT}{2}\ln\hat\sigma_\xi^2 - \frac{nT}{2}\ln|\hat\Sigma_\epsilon| \\ &\quad - \frac{1}{2}\mathrm{vec}'(\mathbf{Z}_{nT}-\mathbf{K}_{nT}\hat\Phi_2)\cdot(\hat\Sigma_\epsilon^{-1}\otimes\mathbf{J}_{nT})\cdot\mathrm{vec}(\mathbf{Z}_{nT}-\mathbf{K}_{nT}\hat\Phi_2) \\ &\quad - \frac{1}{2\sigma_\xi^2}\big(\mathbf{S}_{nT}(\hat\lambda,\hat\gamma,\hat\rho)\mathbf{Y}_{nT} - \mathbf{X}_{1,nT}\hat\beta - \ell_0(\hat\gamma,\hat\rho)\big)'\cdot\mathbf{J}_{nT}\cdot\big(\mathbf{S}_{nT}(\hat\lambda,\hat\gamma,\hat\rho)\mathbf{Y}_{nT} - \mathbf{X}_{1,nT}\hat\beta - \ell_0(\hat\gamma,\hat\rho)\big).\end{aligned}$$

The centered score function at $\hat\theta$ is

$$\sqrt{nT}C_\delta(\hat\theta) = \sqrt{nT}L_\delta(\hat\theta) - (\Delta_{1,\delta}(\hat\theta)+\Delta_{2,\delta}(\hat\theta)) + I_{\delta\Psi}(\hat\theta)I_{\Psi\Psi}^{-1}(\hat\theta)(\Delta_{1,\Psi}(\hat\theta),\Delta_{2,\Psi}(\hat\theta)),$$

with

$$\Delta_{1,\Psi}(\hat\theta) = \begin{pmatrix}\Delta_{1,\lambda}(\hat\theta)\\ \Delta_{1,\phi_1}(\hat\theta)\\ \Delta_{1,\sigma_\xi^2}(\hat\theta)\end{pmatrix} = \sqrt{\frac{n}{T}}\begin{pmatrix} -\frac{1}{n-1}\mathrm{tr}\left(\hat{\mathbf{G}}_{1,nT}\left(\frac{1}{T}1_T1_T'\otimes J_n\right)\right)\\ \frac{1}{n-1}\left(\mathrm{tr}\left(\hat{\mathbf{G}}_{2,nT}\left(\frac{1}{T}1_T1_T'\otimes J_n\right)\right),\mathrm{tr}\left(\hat{\mathbf{G}}_{3,nT}\left(\frac{1}{T}1_T1_T'\otimes J_n\right)\right),0\right)'\\ -\frac{1}{2\hat\sigma_\xi^2}\end{pmatrix},$$

$$\Delta_{2,\Psi}(\hat\theta) = \begin{pmatrix}\Delta_{2,\lambda}(\hat\theta)\\ \Delta_{2,\phi_1}(\hat\theta)\\ \Delta_{2,\sigma_\xi^2}(\hat\theta)\end{pmatrix} = \sqrt{\frac{T}{n}}\begin{pmatrix} -\frac{1}{T}\mathrm{tr}\left(\hat{\mathbf{G}}_{1,nT}\left(I_T\otimes\frac{1}{n}1_n1_n{}'\right)\right)\\ 0_{(k_1+2)\times 1}\\ -\frac{1}{2\hat\sigma_\xi^2}\end{pmatrix},$$

and

$$\begin{aligned}I_{\delta\delta}(\hat\theta) &= \iota_{\delta\delta}(\hat\theta)/(nT\hat\sigma_\xi^2) = \varepsilon_{nT}'(\hat\theta)\mathbf{J}_{nT}\varepsilon_{nT}'(\hat\theta)/(nT\hat\sigma_\xi^2),\\ I_{\delta\Psi}(\hat\theta) &= \begin{pmatrix}\iota_{\delta\lambda}(\hat\theta) & \iota_{\delta\phi_1}(\hat\theta) & \iota_{\delta\sigma_\xi^2}(\hat\theta)\end{pmatrix}/(nT\hat\sigma_\xi^2) = \begin{pmatrix}\varepsilon_{nT}'\mathbf{J}_{nT}(\mathbf{W}_{1,nT}\mathbf{Y}_{nT}) & 0_{p\times(k_1+2)} & 0_{p\times 1}\end{pmatrix}/(nT\hat\sigma_\xi^2),\\ I_{\Psi\Psi}(\hat\theta) &= \begin{pmatrix}\iota_{\lambda\lambda}(\hat\theta) & \iota_{\lambda\phi_1}(\hat\theta) & \iota_{\lambda\sigma_\xi^2}(\hat\theta)\\ \iota_{\phi_1\lambda}(\hat\theta) & \iota_{\phi_1\phi_1}(\hat\theta) & \iota_{\phi_1\sigma_\xi^2}(\hat\theta)\\ \iota_{\sigma_\xi^2\lambda}(\hat\theta) & \iota_{\sigma_\xi^2\phi_1}(\hat\theta) & \iota_{\sigma_\xi^2\sigma_\xi^2}(\hat\theta)\end{pmatrix}/(nT\hat\sigma_\xi^2)\\ &= \begin{pmatrix}(\mathbf{W}_{nT}\mathbf{Y}_{nT})'\mathbf{J}_{nT}\mathbf{W}_{nT}\mathbf{Y}_{nT} + \hat\sigma_\xi^2\mathrm{tr}(\hat{\mathbf{G}}_{1,nT}^2) & \mathbf{R}_{nT}'\mathbf{J}_{nT}(\mathbf{W}_{nT}\mathbf{Y}_{nT}) & \mathrm{tr}(\hat{\mathbf{G}}_{1,nT})\\ * & \mathbf{R}_{nT}'\mathbf{J}_{nT}\mathbf{R}_{nT} & 0_{(k_1+2)\times 1}\\ * & * & \frac{1}{\hat\sigma_\xi^2}\left(\frac{nT}{2}-T-n+1\right)\end{pmatrix}/(nT\hat\sigma_\xi^2).\end{aligned}$$

## A.7 List of 55 Countries

| A-C | D-G | H-M | N-P | S-Z |
|---|---|---|---|---|
| Argentina | Democratic Republic of the Congo | Honduras | Nigeria | South Africa |
| Australia | Denmark | India | Nicaragua | Spain |
| Austria | Dominican Republic | Ireland | Netherlands | Sri Lanka |
| Belgium | Ecuador | Israel | Norway | Sweden |
| Bolivia | Egypt | Italy | New Zealand | Switzerland |
| Brazil | El Salvador | Japan | Pakistan | Thailand |
| Canada | Ethiopia | Kenya | Panama | Trinidad and Tobago |
| Chile | Finland | Morocco | Peru | Turkey |
| Colombia | France | Mexico | Philippines | Uganda |
| Costa Rica | Greece | Mauritius | Portugal | United Kingdom |
| | Guatemala | | Paraguay | United States |
| | | | | Uruguay |
| | | | | Venezuela |