On Testing for Spatial or Social Network Dependence in Panel Data Allowing for Network Variability*

Xiaodong Liu† Ingmar R. Prucha‡

July 21, 2024

*We thank the Editor Aureo de Paula, the Associate Editor and two anonymous referees for their helpful comments and suggestions.

†Department of Economics, University of Colorado Boulder, Boulder, CO 80309, USA. E-mail: xiaodong.liu@colorado.edu.

‡Department of Economics, University of Maryland, College Park, MD 20742, USA. E-mail: prucha@umd.edu.

Abstract

The paper introduces robust generalized Moran $\mathcal{I}$ tests for network-generated cross-sectional dependence in a panel data setting where unit-specific effects can be random or fixed. Network dependence may originate from endogenous variables, exogenous variables, and/or disturbances, and the network dependence is allowed to vary over time. The formulation of the test statistics also aims at accommodating situations where the researcher is unsure about the exact nature of the network. Unit-specific effects are eliminated using the Helmert transformation, which is well known to yield time-orthogonality for linear forms of transformed disturbances. Given the specification of our test statistics, these orthogonality properties also extend to the quadratic forms that underlie our test statistics. This greatly simplifies the expressions for the asymptotic variances of our test statistics and their estimation. Monte Carlo simulations suggest that the generalized Moran $\mathcal{I}$ tests introduced in this paper have the proper size and can provide substantial improvement in robustness when the researcher faces uncertainty about the specification of the network topology.

JEL Classification: C12, C21, C23.

Key Words: Test for network dependence, panel data, spatial and social networks, Moran $\mathcal{I}$ test, $\mathcal{I}_u^2(q)$ test, $\mathcal{I}_y^2(q)$ test, Lagrange Multiplier (LM) test, time-varying network dependence, robustness.

1 Introduction

The paper introduces new tests for network-generated cross-sectional dependence in a panel data setting, where unit-specific effects can be random or fixed. The test statistics are geared to allow for time-varying network dependence. This is important since in many applications the network structure can change over time. An example in a macro setting would be a growth model with spillovers among countries or regions related to, e.g., the relative size of trade, which changes over time. An example in a micro setting would be a situation where there are potential spillovers among friends, but friendships change over time.

Tests for network-generated cross-sectional dependence generally assume knowledge of the nature of the underlying network. However, in applied work, researchers are often unsure about the exact nature of the network that generates spillovers. For example, for a growth model as mentioned above, spillovers may be related to trade, but could also be related to geographic proximity, language resemblance, similarity in industrial sector composition, etc. Or in the friendship network example mentioned above, spillovers can arise from various factors, including similarities in upbringing, educational background, income, and more. Therefore another important feature of our new tests is that they are robust, in the sense that they accommodate situations where the researcher is uncertain about the exact nature of the underlying network. The new tests are not sequential and properly sized.

Our test statistics involve both linear forms and quadratic forms. Towards formulating our test statistic, unit-specific effects are eliminated using the Helmert transformation. The Helmert transformation is well known to yield time-orthogonality for linear forms of transformed disturbances. Upon our adopting appropriate specifications of the quadratic forms that underlie our test statistics, these orthogonality properties are seen to also extend to the quadratic forms. Thus an important advantage of our choice of transformation is that it greatly simplifies the expressions for the asymptotic variances of our test statistics and their estimation.

The foundational principles of the proposed tests draw inspiration from a test introduced by Moran (1950) for spatial correlation of a single variable within a simple cross-sectional setting, and its subsequent generalization to a test for cross-sectional correlation in the disturbances of a linear regression model.[1] Burridge (1980) showed that the Moran $\mathcal{I}$ test can be interpreted as a Lagrange Multiplier (LM) test if the disturbance process under the alternative hypothesis is either a spatial autoregressive or spatial moving average process of order one. King (1980; 1981) demonstrated that the Moran $\mathcal{I}$ test is a locally best invariant test, when the alternative is one-sided, and the errors come from an elliptical distribution. A more detailed discussion of the optimality properties of the Moran $\mathcal{I}$ test can be found in Hillier and Martellosio (2018), including a discussion of conditions under which the Moran $\mathcal{I}$ test is a uniformly most powerful invariant test. We note that, while designed for a general setting, our test statistics also have interpretations as LM tests for specific alternatives. Kelejian and Prucha (2001) introduced a central limit theorem for spatial and social networks and established that the Moran $\mathcal{I}$ test statistic is asymptotically distributed $N(0,1)$. In line with this result, we establish that our test statistics, which contain the squared Moran $\mathcal{I}$ test statistic as a special case, are asymptotically χ^2 distributed.

In situations where a researcher is uncertain about the network structure, they could apply multiple Moran $\mathcal{I}$ tests corresponding to different potential network specifications, using, e.g., Bonferroni-Holm adjustments of the individual Moran $\mathcal{I}$ tests. However, within the context of cross-sectional data Liu and Prucha (2018) report on Monte Carlo sim-

[1] See Durbin and Watson (1950; 1951) for a corresponding test in a time series setting.

ulations that show that such an approach can lead to sizable distortions of the desired overall significance level. Within their setting Liu and Prucha (2018) proposed to overcome this problem by combining, loosely speaking, several Moran $\mathcal{I}$ test statistics into a single test statistic. The resulting test statistics then have the structure of a quadratic form in linear forms and quadratic forms; see also Robinson's (2008) test for correlation in the disturbances where the statistics are defined as quadratic forms of quadratic forms (within the context of regression models where all regressors are exogenous). The test statistics introduced in this paper for panel data built on work by Liu and Prucha (2018) for cross-sectional data. Monte Carlo simulations suggest that the generalized Moran $\mathcal{I}$ tests introduced in this paper have the proper size and can provide substantial improvement in robustness when the researcher faces uncertainty about the specification of the network topology.

Our tests are applicable to detect potential cross-sectional dependence from both spatial and social networks, where the network dependence may originate from the dependent variable, exogenous variables, and/or disturbances. That is, using the terminology coined in Manski (1993) for social networks, network dependence may stem from endogenous peer effects, contextual effects, and/or correlated effects. We formulate different tests for different potential sources of network effects. The tests should have wide applicability. In particular, suppose an empirical researcher reports on estimation results obtained under the assumption of the absence of network effects, based on their a priori beliefs, in one or all of the channels mentioned above. We then envision the proposed tests as useful and robust tools for checking this assumption, similar to the use of other tests a researcher may report on in support of other assumptions underlying the empirical research. Furthermore, some past empirical research may have overlooked network effects. The proposed tests can also be used to check the validity of studies that do not account for network effects.

For example, Kelly (2019) analyzes 27 persistence studies published in leading economic journals and suggests the use of Moran $\mathcal{I}$ test in a procedure that guards against spurious regressions due to spatial dependence. Our tests can be particularly advantageous in this situation, as researchers are often unsure about the exact form of the spatial weight matrix that characterizes potential network spillovers to implement the Moran $\mathcal{I}$ test. As is well known, in general pre-testing may affect the actual distribution of post-model-selection estimators, and consequently we caution against the use of our tests as pre-tests, except in situations where the researcher has access to an independent sample for that purpose.[2]

There is increasing interest in the analysis of networks.[3] The need to account for potential network dependencies was recognized early in the regional science, urban economics, and geography literature. An important class of network models was introduced by Cliff and Ord (1973; 1981). The original focus of the authors was on spatial networks. However, since the formulation of those models only depends on a measure of distance and not on geographic location, those models can also be applied to other networks. In particular, those models can be applied to classical social networks, where information on the distance between units is collected in an adjacency matrix, recognizing that formally an adjacency-type matrix can be seen as a special case of a spatial weight matrix employed in the formulation of Cliff-Ord network models. For recent contributions to the social network literature see, e.g., Bramoullé, Djebbari and Fortin (2009), Lee, Liu and Lin (2010), Blume, Brock, Durlauf and Jayaraman (2015), de Paula (2017) and Kuersteiner and Prucha (2020).[4]

[2] See, in particular, Leeb and Pötscher (2003) and the subsequent related literature. In light of the pretesting problem, it seems prudent to allow for general forms of network effects in empirical research, and only impose their absence (or partial absence) if considered unlikely, accompanied by a test (such as the test proposed in this paper) for their absence.

[3] E.g., Kolaczyk (2009) remarked that "... during the decade surrounding the turn of the 21st century network-centric analysis ... has reached new levels of prevalence and sophistication."Applications range widely from physical and mathematical sciences to social sciences and humanities.

[4] We note that Cliff-Ord network models also cover simple group-average models as special cases; see,

Early contributions to the literature on testing for network dependence in panel data include Baltagi, Song and Koh (2003), Baltagi, Song, Jung and Koh (2007), Baltagi and Liu (2008), Baltagi, Song and Kwon (2009), and He and Lin (2015), who considered LM tests for spatial error correlation and/or spatial lag dependence for random effects models. Baltagi and Yang (2013) provided standardized LM tests for spatial error correlation in both random effects and fixed effects panel data models.

For spatial dependence in fixed effects panel data models, Debarsy and Ertur (2010) derived LM test statistics and their likelihood ratio (LR) counterparts. Taşpınar, Doğan and Bera (2017) derived GMM gradient tests based on the GMM approach in Lee and Yu (2014), focusing on spatial lag dependence in the endogenous variables only. Yang (2021) proposed adjusted quasi score (AQS) tests based on his M-estimation method in Yang (2018). We note that, in contrast to our paper, none of the above papers (except for Taşpınar et al., 2017) considers higher-order spatial/network lags and none considers time-varying network dependence. Additionally, different from our paper, all of the above papers postulate very specific model structures under the alternative.

Our asymptotic results are derived under the assumption that the cross-sectional dimension n tends to infinity, while the time dimension T is fixed. For cases where T also tends to infinity, Pesaran (2004) developed a cross-section dependence (CD) test, exploiting the additional information from the time dimension. Pesaran, Ullah and Yamagata (2008) proposed a bias-adjusted LM test based on finite sample approximation in the context of a heterogeneous panel data model. Baltagi, Feng and Kao (2012) suggested a simple bias-corrected LM test based on the asymptotic bias of the scaled version of the LM test in the context of a fixed effects homogeneous panel data model. Baltagi, Kao and Peng (2016) further extended Baltagi et al. (2012) to allow for unknown forms of serial corre-

e.g., Lee (2007), Davezies, D'Haultfoeuille and Fougère (2009), and Carrell, Sacerdote and West (2013).

lation. Bera, Doğan, Taşpınar and Leiluo (2019) developed adjusted LM tests for spatial lag dependence in the dependent variable in a maximum likelihood (ML) framework, assuming the absence of spatial lag dependence in the disturbances. In addition to requiring $T \to \infty$, none of the above papers (except for Bera et al., 2019) considers time-varying network dependence.

The paper is organized as follows: In Section 2, we define our new test statistics regarding different types of network-generated dependence in a general panel data setting and establish the limiting distribution of the proposed test statistics. In Section 3, we show the connection of the proposed test statistics with LM test statistics. In Section 4, we discuss the implementation of the tests with endogenous weight matrices. In Section 5, we report on Monte Carlo simulation results regarding the small sample properties of our test statistics. Concluding remarks are given in Section 6. All technical details are relegated to the appendices and a supplementary online appendix. In the online appendix we also report on additional Monte Carlo simulations.

For a square matrix $A = (a_{ij})$, let $\mathring{A} = (A + A')/2$, let $\text{vec}_D(A)$ denote the column vector of the diagonal elements of A, and let A^- denote the Moore Penrose generalized inverse. In abuse of conventional notation we also write $A \subset B$, if the columns of A are a subset of the columns of B. Next let A be an $n \times n$ matrix with $\sup_n \sum_{j=1}^n |a_{ij}| < \infty$ and $\sup_n \sum_{i=1}^n |a_{ij}| < \infty$, then we say, abusing language slightly, that the row and column sums of the matrix are uniformly bounded in absolute value. For $n \times m$ matrices $B_1, \cdots, B_T$, the corresponding Helmert transformed matrices are defined as $B_t^+ = \sum_{\tau=1}^T \pi_{t\tau} B_\tau$, for $t = 1, \cdots, T-1$, where $\pi_{t\tau}$ are the weights of the Helmert transformation, given in Section 2.2.

2 Test Statistics for Network Dependence in Panel Data

In the following, we define our new test statistics regarding different types of network-generated dependence (or spillovers) within a panel data setting. We differentiate between testing for network dependence in the disturbances and in the dependent variable. Of course, network dependence in the disturbances will generally lead to network dependence in the dependent variable. However, even in the absence of spillovers in the disturbances, network dependence in the dependent variable may also arise from spillovers in the endogenous variable and/or exogenous variables. We refer to these tests as the $\mathcal{I}_u^2$ and $\mathcal{I}_y^2$ tests respectively. These tests are related in spirit to tests introduced in Liu and Prucha (2018) for a single cross-sectional dataset.

2.1 Motivation and Intuition

Suppose a set of panel data for n individuals and T periods is generated by the following linear panel data model ($t = 1, \ldots, T$)

$$y_t = Z_t \delta_0 + u_t, \tag{1}$$

where $y_t = (y_{1t}, \ldots, y_{nt})'$ is an $n \times 1$ vector of observations on the dependent variable, $Z_t = [z_{it,k}]$ is an $n \times K_Z$ matrix of observations on K_Z endogenous and/or nonstochastic exogenous regressors, where we collect the latter in an $n \times K_X$ matrix X_t, and where $u_t = (u_{1t}, \ldots, u_{nt})'$ is an $n \times 1$ vector of regression disturbances.

To accommodate network effects, we allow for all variables to depend on the sample size n, i.e, to form triangular arrays, though we suppress the index n on respective variables for simplicity of notation. Consequently, as a simple example, the above specification allows for the data to be generated as

$$y_{it} = \overline{y}_{it}\delta_{1,0} + x_{it}\delta_{2,0} + \overline{x}_{it}\delta_{3,0} + u_{it},$$

with $\overline{y}_{it} = \sum_{i=1}^{n} w_{ij,t}y_{jt}$ and $\overline{x}_{it} = \sum_{i=1}^{n} w_{ij,t}x_{jt}$, where x_{it} denotes an exogenous regressor and $w_{ij,t}$ is a weight with $w_{ii,t} = 0$. The weight $w_{ij,t}$ is typically thought to be determined by a measure of geographical, economic, or social proximity between units i and j, contingent upon the context.[5] The weighted averages $\overline{y}_{it}$ and $\overline{x}_{it}$ are the conduits for network spillovers. In the spatial literature, weighted averages of the above form would typically be called spatial lags. For social interaction models, adopting the terminology of Manski (1993), those averages would be said to represent the endogenous peer effect and the contextual effect respectively. A simple example of a disturbance process with spillovers would be

$$u_{it} = \rho_0\overline{u}_{it} + \mu_i + \epsilon_{it},$$

with $\overline{u}_{it} = \sum_{i=1}^{n} w_{ijt}u_{jt}$, where μ_i denotes unit specific fixed effects and ϵ_{it} an idiosyncratic disturbance term. In the social interaction literature, the weighted average $\overline{u}_{it}$ would typically be said to represent the correlated effect.

In the following discussion, the weighted averages are expressed more compactly as $\overline{y}_t = (\overline{y}_{1t}, \ldots, \overline{y}_{nt})' = W_t y_t$, $\overline{x}_t = (\overline{x}_{1t}, \ldots, \overline{x}_{nt})' = W_t x_t$, and $\overline{u}_t = (\overline{u}_{1t}, \ldots, \overline{u}_{nt})' = W_t u_t$, where $W_t = [w_{ij,t}]$ denotes the $n \times n$ weight matrix with zero diagonal elements.

As remarked in the Introduction, the foundational principles of the proposed tests draw inspiration from a test introduced by Moran (1950) for cross-sectional data. This test, and its generalization to test for cross-sectional dependence in the disturbances of a

[5] For example, in the social interaction literature, the weights are often chosen as $w_{ij,t} = 1/n_{i,t}$ if j is a "friend" of i and zero otherwise, where $n_{i,t}$ denotes the number of friends of i in period t. Let $\mathcal{I}_i$ denote the index set of friends of i, then $\overline{y}_{it} = n_{i,t}^{-1}\sum_{j\in\mathcal{I}_i} y_{jt}$ and $\overline{x}_{it} = n_{i,t}^{-1}\sum_{j\in\mathcal{I}_i} x_{jt}$ represent arithmetic peer averages of the endogenous and exogenous variables, respectively.

linear regression model, is typically referred to as the Moran $\mathcal{I}$ test. The Moran $\mathcal{I}$ test statistic in essence assumes knowledge of the network structure, represented by the weight matrix W_t. The Moran $\mathcal{I}$ test statistics is then formed as an appropriately normalized quadratic form of the disturbances u_t, with W_t in the middle of the quadratic form. Since the diagonal elements of W_t are zero, the expected value of the test statistic is zero if the disturbances are cross-sectionally uncorrelated and generally nonzero otherwise. A problem is that the power of the test will depend on whether or not the weight matrix W_t properly represents the network structure. Consequently, our generalized test statistic is geared towards accommodating situations where the researcher is unsure about the network structure. For example, as discussed in the Introduction, in a growth model spillovers between countries or regions may be associated with different measures of proximity. Our test statistics accommodate different potential network structures by combining quadratic forms that correspond to different weight matrices. They contain the (squared) Moran $\mathcal{I}$ test statistic as a special case.

Since our panel data model allows for fixed effects, we first eliminate the fixed effects by transforming the data using the Helmert transformation. Our test statistics are then formed in terms of the Helmert-transformed variables. By using the Helmert transformation and showing that this transformation leads to the orthogonality of both linear and quadratic forms, we are able to obtain relatively simple expressions for the variance-covariance (VC) matrix of linear and quadratic forms, which in turn simplifies the expressions for our test statistics.

2.2 Model and Helmert Transformation

As discussed above, in general, the model in (1) allows for both endogenous and exogenous regressors in Z_t, as well as for spillovers in the form of network lags. Thus Z_t could be composed of (a subset of) the columns of $[W_t y_t, X_t, W_t X_t, Y_t^o, W_t Y_t^o]$, where X_t is a matrix of exogenous variables and Y_t^o is a matrix of "outside" endogenous variables. More generally, Z_t could also contain network lags corresponding to different weight matrices.

Under the assumption that there is no network-generated correlation in the disturbance process we maintain that, for $t = 1, \ldots, T$,

$$u_t = \mu + \epsilon_t \tag{2}$$

where $\mu = (\mu_1, \ldots, \mu_n)'$ is the vector of unit-specific effects and $\epsilon_t = (\epsilon_{1t}, \ldots, \epsilon_{nt})'$ is the vector of idiosyncratic disturbances.

We do not maintain any specific assumptions regarding the unit-specific effects, which can be fixed or random, and eliminate them by transforming the observations. In general, let $\pi_t = (\pi_{t1}, \cdots, \pi_{tT})$ be the tth row of a $(T-1) \times T$ transformation matrix Π with $\sum_{\tau=1}^{T} \pi_{t\tau} = 0$, then the individual effects can be eliminated by a transformation of the form

$$u_t^+ = \sum_{\tau=1}^{T} \pi_{t\tau} u_\tau = \sum_{\tau=1}^{T} \pi_{t\tau} \epsilon_\tau = \epsilon_t^+.$$

The transformations satisfying the above condition include the one-period forward-differencing, the differencing from the sample average, and the Helmert transformation. In particular, the Helmert transformation was introduced to the panel data literature by Arellano and Bover (1995), and corresponds to $\pi_{ts} = 0$ for $s < t$, $\pi_{tt} = \sqrt{(T-t)/(T-t+1)}$, and $\pi_{ts} = -\sqrt{(T-t)/(T-t+1)}/(T-t)$ for $s > t$.[6] In the following, we adopt the convention that $\Pi = [\pi_{t\tau}]$ refers specifically to the Helmert transformation matrix and Helmert transformed variables are denoted with a superscript "+".

The Helmert transformation is a forward-differencing transformation, which is ortho-

[6] A generalization of the Helmert transformation to accommodate time-varying individual effects was introduced in Kuersteiner and Prucha (2020).

normal in that $\pi_t \pi_t' = 1$ and $\pi_t \pi_s' = 0$ for $t \neq s$. As a consequence, the transformed disturbances u_t^+, and linear functions of u_t^+, are uncorrelated over time. Our test statistics involve quadratic forms of the transformed disturbances. Existing results on variances and covariances of quadratic forms given in, e.g., Kelejian and Prucha (2010) and Kuersteiner and Prucha (2020), imply that the orthogonality property of the Helmert transformation also extends to quadratic forms, provided that the diagonal elements of the weight matrices of the quadratic forms are zero. In more detail, let A and B be non-stochastic symmetric zero-diagonal $n \times n$ matrices. Then, by Lemma A.1 in Appendix A, in general $\mathrm{Cov}(u_t^{\times\prime} A u_t^{\times}, u_s^{\times\prime} B u_s^{\times}) \neq 0$ for $t \neq s$, where $u_t^{\times}$ represents the disturbances obtained from a transformation that is not orthogonal. In contrast, for the Helmert transformation, $\mathrm{Cov}(u_t^{+\prime} A u_t^{+}, u_s^{+\prime} B u_s^{+}) = 2\sigma_0^4 \mathrm{tr}(AB)(\sum_{\tau=1}^{T} \pi_{t\tau}\pi_{s\tau})^2 = 0$ as $\sum_{\tau=1}^{T} \pi_{t\tau}\pi_{s\tau} = 0$ for $t \neq s$. As our test statistics are based on linear and quadratic forms of the transformed disturbances, using the Helmert transformation greatly simplifies the expression of the test statistic. Apply the Helmert transformation to (1) and (2) yields, for $t = 1, \ldots, T-1$,

$$y_t^+ = Z_t^+ \delta_0 + u_t^+, \quad \text{and} \quad u_t^+ = \epsilon_t^+. \tag{3}$$

Let H_t denote an instrumental variable (IV) matrix.[7] With a little abuse of notation, define $y^+ = [y_1^{+\prime}, \cdots, y_{T-1}^{+\prime}]'$, $Z^+ = [Z_1^{+\prime}, \cdots, Z_{T-1}^{+\prime}]'$ and $H^+ = [H_1^{+\prime}, \cdots, H_{T-1}^{+\prime}]'$. Then δ_0 can be estimated by the 2SLS estimator $\widehat{\delta} = (\widehat{Z}^{+\prime}\widehat{Z}^{+})^{-1}\widehat{Z}^{+\prime}y^+$ with $\widehat{Z}^+ = H^+(H^{+\prime}H^+)^{-1}H^{+\prime}Z^+$, and σ_0^2 can be estimated by $\widehat{\sigma}^2 = [n(T-1)]^{-1}\sum_{t=1}^{T-1}\widehat{u}_t^{+\prime}\widehat{u}_t^{+} = [n(T-1)]^{-1}\widehat{u}^{+\prime}\widehat{u}^{+}$ with $\widehat{u}_t^+ = y_t^+ - Z_t^+\widehat{\delta}$ and $\widehat{u}^+ = [\widehat{u}_1^{+\prime}, \cdots, \widehat{u}_{T-1}^{+\prime}]'$.

[7] We defer the discussion on possible choices for H_t to the following subsections.

2.3 The $\mathcal{I}_u^2(q)$ Test Statistic

Suppose the researcher wants to test for network-generated correlation in the model disturbances, where the underlying structure of the network at period t is represented by an $n \times n$ zero-diagonal weight matrix $W_t = [w_{ij,t}]$. More specifically, the researcher wants to test the null hypothesis that the VC matrix of u_t is proportional to the identity matrix, i.e.,

$$H_0^u : \mathrm{VC}(u_t|\mu) = \sigma_0^2 I_n, \qquad \text{for } t = 1, \cdots, T, \tag{4}$$

against the alternative that the disturbances are cross-sectionally correlated.[8] Since the Helmert transformation is an orthogonal transformation, we have $\mathrm{VC}(u_t^+) = \sigma_0^2 I_n$ under the null hypothesis, which implies $\mathrm{E}(u_t^{+\prime} W_t^* u_t^+) = \sigma_0^2 \mathrm{tr}(W_t^*) = 0$ with $W_t^* = \sum_{\tau=1}^{T} \pi_{t\tau}^2 W_\tau$.[9] On the other hand, in general, $\mathrm{E}(u_t^{+\prime} W_t^* u_t^+) \neq 0$ under the alternative hypothesis. This basic idea, which is in line with that underlying the Moran $\mathcal{I}$ test for cross-sectional data, motivates the following test statistic for H_0^u:

$$\mathcal{I} = \widehat{\Psi}_Q^{-1/2} \widehat{V}_Q, \tag{5}$$

where $\widehat{V}_Q = \sum_{t=1}^{T-1} \widehat{u}_t^{+\prime} W_t^* \widehat{u}_t^+$ and $n^{-1}\widehat{\Psi}_Q$ is a consistent estimator for the limiting VC matrix of $n^{-1/2}\widehat{V}_Q$. For simplicity of presentation, we refer to the above test statistic as the Moran $\mathcal{I}$ test statistic, while noting that it would be more appropriate to refer to it as a Moran $\mathcal{I}$-type test statistic.

We emphasize that the weights $w_{ij,t}$ are generally considered to be reflective of some measure of proximity between units, but do not depend on explicit indexing of units by location. By extending the notion of proximity from geographical proximity to economic

[8] The weight matrix W_t is taken to be non-stochastic and suppressed from any information set for simplicity.

[9] The motivation to use W_t^* instead of other linear combinations of $W_1, \cdots, W_T$ to construct the test statistics will be given in Section 3.

proximity, technological proximity, social proximity, etc., the Moran $\mathcal{I}$ test statistic becomes useful for testing for dependence not only within the context of spatial networks, but for a much wider class of networks, including social networks.

One practical problem with the Moran $\mathcal{I}$ test statistic defined in (5) is that empirical researchers are often unsure about the specification of W_t. Thus it is of interest to consider a generalized Moran $\mathcal{I}$ test for situations where the researcher is not sure whether $W_{t,1}, W_{t,2}, \cdots$, or $W_{t,q}$, or some linear combination of those matrices properly model the network topology. Towards introducing such a generalization, let

$$\widehat{V}_Q = \begin{bmatrix} \widehat{u}^{+\prime} W_1^* \widehat{u}^+ \\ \vdots \\ \widehat{u}^{+\prime} W_q^* \widehat{u}^+ \end{bmatrix}, \quad \text{and} \quad \widehat{\Phi}_Q = 2\widehat{\sigma}^4 \begin{bmatrix} \mathrm{tr}(\mathring{W}_1^* \mathring{W}_1^*) & \cdots & \mathrm{tr}(\mathring{W}_1^* \mathring{W}_q^*) \\ \vdots & \ddots & \vdots \\ \mathrm{tr}(\mathring{W}_q^* \mathring{W}_1^*) & \cdots & \mathrm{tr}(\mathring{W}_q^* \mathring{W}_q^*) \end{bmatrix}, \tag{6}$$

where $W_r^* = \mathrm{diag}_{t=1}^{T-1}\{W_{t,r}^*\}$ is a block diagonal matrix with the tth diagonal block being $W_{t,r}^* = \sum_{\tau=1}^{T} \pi_{t\tau}^2 W_{\tau,r}$ for $r = 1, \cdots, q$; and let

$$\widehat{\Psi}_Q = \widehat{\Phi}_Q + \widehat{\Sigma}_Q, \tag{7}$$

where $\widehat{\Phi}_Q$ is defined in (6) and $\widehat{\Sigma}_Q$ is a $q \times q$ matrix with the (r,s)th element being $4\widehat{\sigma}^2 \widehat{u}^{+\prime} \mathring{W}_r^* (Z^+ - \widehat{Z}^+)(\widehat{Z}^{+\prime} \widehat{Z}^+)^{-1} (Z^+ - \widehat{Z}^+)' \mathring{W}_s^* \widehat{u}^+$. Under the regularity conditions of Theorem 1 below, $n^{-1}\widehat{\Psi}_Q$ is seen as a consistent estimator for the limiting VC matrix of $n^{-1/2}\widehat{V}_Q$. Suppose $\widehat{\Psi}_Q$ is nonsingular. Then the proposed generalized Moran $\mathcal{I}$ statistic for H_0^u is defined as

$$\mathcal{I}_u^2(q) = \widehat{V}_Q' \widehat{\Psi}_Q^{-1} \widehat{V}_Q. \tag{8}$$

The above statistic generalizes the (squared) Moran $\mathcal{I}$ test statistic. It may be viewed as combining q Moran $\mathcal{I}$ test statistics in a way that controls the significance level of the overall test. As such it represents an attractive alternative to q sequential Moran $\mathcal{I}$ tests. For $q = 1$, the $\mathcal{I}_u^2(q)$ test statistic reduces to the square of the Moran $\mathcal{I}$ test statistic defined in (5).

To formally establish the asymptotic properties of the proposed test statistic, we assume the following conditions hold under the null hypothesis.

Assumption 1 The innovations ϵ_{it} are i.i.d. with mean zero, variance σ_0^2, and finite $(4+\kappa)$th moments for some $\kappa > 0$, for $i = 1, \cdots, n$ and $t = 1, \cdots, T$.

Assumption 2 The zero-diagonal weight matrices $W_{t,r}$ are non-stochastic with row and column sums uniformly bounded in absolute value, for $r = 1, \cdots, q$ and $t = 1, \cdots, T$.

Assumption 3 Let $A = \mathrm{diag}_{t=1}^{T-1}\{A_t\}$, where A_t is non-stochastic and the row and column sums of A_t are uniformly bounded in absolute value, then $n^{-1}Z^{+\prime}AZ^{+} = O_p(1)$ and $n^{-1}Z^{+\prime}A\epsilon^{+} = n^{-1}\mathrm{E}(Z^{+\prime}A\epsilon^{+}) + o_p(1)$, where $n^{-1}\mathrm{E}(Z^{+\prime}A\epsilon^{+}) = O(1)$. In particular, the assumption holds for $A = \mathring{W}_r^*$ and $A = I_{n(T-1)}$.

Assumption 4 The instrument matrix H_t includes the exogenous regressors X_t. The elements of H_t are non-stochastic and uniformly bounded for $t = 1, \cdots, T$. $\mathrm{plim}_{n\to\infty} n^{-1}H^{+\prime}Z^{+}$ is finite with full column rank and $\lim_{n\to\infty} n^{-1}H^{+\prime}H^{+}$ is finite and nonsingular.

The assumptions are generally in line with the spatial econometrics literature. If, in contrast to what is assumed in Assumption 2, the weight matrices $W_{t,r}$ are endogenous and stochastic, then the test statistic could be constructed based on a non-stochastic analog of $W_{t,r}$.[10]

Assumption 3 is kept general, to allow for spatial lags of the dependent variable and "outside" endogenous variables in the regressor matrix Z_t under the null hypothesis. More

[10]Such an analog could be obtained from a projection of $W_{t,r}$ on the exogenous variables in the process that generates $W_{t,r}$. See Section 4 and the Monte Carlo study in the online appendix for further discussion.

specifically, let $\underline{X}_t$ represent the matrix of system-wide exogenous variables, where the ith row of $\underline{X}_t$ only contains exogenous characteristics of unit i. Under H_0^u, the regressor matrix Z_t could be composed of (a subset of) the linearly independent columns of $[W_{t,1}y_t, \cdots, W_{t,q}y_t, X_t, W_{t,1}X_t, \cdots, W_{t,q}X_t,\ Y_t^o, W_{t,1}Y_t^o, \cdots, W_{t,q}Y_t^o]$, where $X_t \subset \underline{X}_t$ and Y_t^o is a matrix of "outside" endogenous variables. Given that Assumption 3 is high level we provide additional discussions of this assumption, including a discussion of sufficient conditions, in the online appendix.

Assumption 4 is standard for the IV matrix H_t. To discuss possible choices for IVs, we adopt the notation that $[A_j]_{j=1}^m := [A_1, \cdots, A_m]$ for any set of conformable matrices $A_1, \cdots, A_m$. Given this notational convention, the IV matrix H_t may be composed of the linearly independent columns of $\underline{X}_t$, $[W_{t,j}\underline{X}_t]_{j=1}^q$, $\cdots$, $[W_{t,j_1}W_{t,j_2}\cdots W_{t,j_R}\underline{X}_t]_{j_1,j_2,\cdots,j_R=1}^q$, for some fixed constant R. See Drukker, Egger and Prucha (2023) for a more in-depth discussion of instrument selection for systems with spatial or social network structures, including their Appendix F, which discusses scenarios where identification and instruments would be weak. The following theorem gives the asymptotic distribution of the $\mathcal{I}_u^2(q)$ test statistic under H_0^u.

Theorem 1. *If the null hypothesis H_0^u and Assumptions 1-4 hold, then $n^{-1}\widehat{\Psi}_Q = n^{-1}\Psi_Q + o_p(1)$,*[11] *where $n^{-1}\Psi_Q$ is non-stochastic and converges to the limiting VC matrix of $n^{-1/2}\widehat{V}_Q$ under H_0^u. Furthermore, provided that the smallest eigenvalues of $n^{-1}\Psi_Q$ are bounded away from zero, $\mathcal{I}_u^2(q) = \widehat{V}_Q'\widehat{\Psi}_Q^{-1}\widehat{V}_Q \xrightarrow{d} \chi^2(q)$.*

The above theorem is derived by showing that the elements of $n^{-1/2}\widehat{V}_Q$ are asymptotically equivalent to linear-quadratic forms, rather than just quadratic forms, in the innovations ϵ. The linear part, in general, stems from the presence of cross-sectionally correlated endogenous regressors. To establish the limiting distribution of $n^{-1/2}\widehat{V}_Q$, the

[11] Explicit expressions for Ψ_Q are given in the proof.

proof applies the CLT for vectors of linear-quadratic forms given in Kelejian and Prucha (2001; 2010). We note that the proof of the CLT in Kelejian and Prucha (2001; 2010) is based on rewriting the linear-quadratic forms as a sum of martingale differences, checking the conditions of the CLT for martingale differences in Hall and Heyde (1981) are satisfied, and deriving simplified expressions for the limiting VC matrix.

In the spirit of the outer product gradient approach of Born and Breitung (2011), an alternative approach to prove the above theorem would be to rewrite the elements $n^{-1/2}\widehat{V}_Q$ as martingale differences, and then verify that, under the maintained assumptions, the conditions of the CLT for martingale differences in Hall and Heyde (1981) are satisfied. However, formally checking those conditions of the CLT for martingale differences is involved, and the advantage of using the CLT in Kelejian and Prucha (2001; 2010) is that those conditions have been verified under a set of easy-to-check basic conditions. The estimator for the VC matrix used in normalizing $n^{-1/2}\widehat{V}_Q$ is based on the normalized limit of the sample variance of the martingale differences. Alternatively one could use the sample variance of the martingale differences as an estimator. However, the latter estimator does not set terms, which can be seen to go to zero in probability, to zero.

2.4 The $\mathcal{I}_y^2(q)$ Test Statistic

Now, suppose the researcher wants to test for a more general form of network-generated dependence. Such dependence could come from the dependence of an individual's dependent variable on the dependent variable, exogenous variables, and/or disturbances of other individuals in the network. More specifically, suppose the researcher wants to test the null hypothesis $(t = 1, \cdots, T)$

$$H_0^y : \mathrm{VC}(y_t|Z_t, \mu) = \sigma_0^2 I_n, \quad \text{and} \quad \mathrm{E}(y_t|Z_t, \mu) = Z_t\delta_0 + \mu, \tag{9}$$

where the ith row of the regressor matrix Z_t only contains exogenous and endogenous characteristics specific to the ith individual.[12] That is, the researcher wants to test that, conditional on individual characteristics Z_t and individual effects μ, (i) the dependent variable is uncorrelated across individuals, and (ii) the expected value of the dependent variable of an individual only depends on the characteristics specific to that individual, and thus is not affected by the characteristics of other individuals.

In line with our motivation of the $\mathcal{I}_u^2(q)$ test statistic, suppose again that the empirical researcher is not sure whether the weight matrices $W_{t,1}, W_{t,2}, \cdots$, or $W_{t,q}$ or some linear combination of those matrices properly model the network structure at period t. Let $\bar{H}_{t,r} = W_{t,r}H_t$, for $r = 1, \cdots, q$, where the IV matrix H_t contains linearly independent columns of the matrix of system-wide exogenous variables $\underline{X}_t$. Then clearly the following linear and quadratic moment conditions $\mathrm{E}(\bar{H}_{t,r}^{\prime +} u_t^+) = 0$ and $\mathrm{E}(u_t^{+\prime} W_{t,r}^* u_t^{+\prime}) = 0$, for $r = 1, \cdots, q$, hold under the null hypothesis but may not generally hold under the alternative. This motivates the following $\mathcal{I}_y^2(q)$ test statistic. With a little abuse of notation, define $\bar{H}_r^+ = [\bar{H}_{1,r}^{+\prime}, \cdots, \bar{H}_{T-1,r}^{+\prime}]'$. Let

$$\widehat{V} = \begin{bmatrix} \widehat{V}_L \\ \widehat{V}_Q \end{bmatrix} \quad \text{and} \quad \widehat{\Phi} = \begin{bmatrix} \widehat{\Phi}_L & 0 \\ 0 & \widehat{\Phi}_Q \end{bmatrix},$$

where $\widehat{V}_Q$ and $\widehat{\Phi}_Q$ are defined in (6),

$$\widehat{V}_L = \begin{bmatrix} \bar{H}_1^{+\prime}\widehat{u}^+ \\ \vdots \\ \bar{H}_q^{+\prime}\widehat{u}^+ \end{bmatrix}, \quad \text{and} \quad \widehat{\Phi}_L = \widehat{\sigma}^2 \begin{bmatrix} \bar{H}_1^{+\prime} M_{\widehat{Z}Z}' M_{\widehat{Z}Z} \bar{H}_1^+ & \cdots & \bar{H}_1^{+\prime} M_{\widehat{Z}Z}' M_{\widehat{Z}Z} \bar{H}_q^+ \\ \vdots & \ddots & \vdots \\ \bar{H}_q^{+\prime} M_{\widehat{Z}Z}' M_{\widehat{Z}Z} \bar{H}_1^+ & \cdots & \bar{H}_q^{+\prime} M_{\widehat{Z}Z}' M_{\widehat{Z}Z} \bar{H}_q^+ \end{bmatrix}, \tag{10}$$

[12] The linear dependence of $\mathrm{E}(y_t|Z_t, \mu)$ on Z_t is only maintained for ease of exposition, and the assumption could be extended to allow for $\mathrm{E}(y_t|Z_t, \mu)$ to depend nonlinearly on Z_t.

with $M_{\widehat{Z}Z} = I_{n(T-1)} - \widehat{Z}^{+}(\widehat{Z}^{+\prime}\widehat{Z}^{+})^{-1}Z^{+\prime}$. The $qK_H \times 1$ vector $\widehat{V}_L$ collects the linear moments, with K_H denoting the number of columns in H_t, and the $q \times 1$ vector $\widehat{V}_Q$ collects the quadratic moments. Under the regularity conditions of Theorem 2 below, $n^{-1}\widehat{\Phi}$ is seen as a consistent estimator for the limiting VC matrix of $n^{-1/2}\widehat{V}$. Then the proposed generalized Moran $\mathcal{I}$ statistic for H_0^y is defined as

$$\mathcal{I}_y^2(q) = \widehat{V}'\widehat{\Phi}^{-1}\widehat{V}. \tag{11}$$

To formally establish the asymptotic properties of the test statistic, we maintain similar assumptions as Theorem 1, with the following modifications. First, for the $\mathcal{I}_u^2(q)$ test, Z_t may include spatial lags of endogenous variables such as $W_r y$ and $W_r Y^o$, where Y^o is a matrix of "outside" endogenous variables, and thus $\mathrm{E}(Z^{+\prime}A\epsilon^{+})$ may not be zero even if $\mathrm{tr}(A) = 0$. Hence, in Assumption 3, we only assume $n^{-1}\mathrm{E}(Z^{+\prime}A\epsilon^{+}) = O(1)$. In contrast, since the $\mathcal{I}_y^2(q)$ test is designed to detect the presence of network-generated dependence in the dependent variable, spatial lags of endogenous variables are not allowed as regressors under H_0^y. We modify Assumption 3 accordingly as follows.

Assumption 3' Assumption 3 holds. Furthermore, $\mathrm{E}(Z^{+\prime}A\epsilon^{+}) = 0$ when $\mathrm{tr}(A) = 0$.

Second, as $\mathcal{I}_y^2(q)$ also uses linear moment conditions, we expand Assumption 4 as follows.

Assumption 4' Assumption 4 holds. Furthermore, $\mathrm{plim}_{n\to\infty} n^{-1}\bar{H}_r^{+\prime}Z^{+}$ is finite with full column rank and $\lim_{n\to\infty} n^{-1}\bar{H}_r^{+\prime}\bar{H}_r^{+}$ is finite and nonsingular, for $r = 1, \cdots, q$.

The following theorem gives the asymptotic distribution of the $\mathcal{I}_y^2(q)$ test statistic under H_0^y.

Theorem 2. *If the null hypothesis H_0^y and Assumptions 1, 2, 3' and 4' hold, then* $n^{-1}\widehat{\Phi} = n^{-1}\Phi + o_p(1)$,[13] *where $n^{-1}\Phi$ is non-stochastic and converges to the limiting VC matrix of $n^{-1/2}\widehat{V}$ under H_0^y. Furthermore, provided that the smallest eigenvalues of $n^{-1}\Phi$ are bounded away from zero,* $\mathcal{I}_y^2(q) = \widehat{V}'\widehat{\Phi}^{-1}\widehat{V} \xrightarrow{d} \chi^2((K_H+1)q)$.

A discussion analogous to that given after Theorem 1 also applies here. For reasons of generality, the formulation of Theorem 2 does not assume an explicit form of the underlying data-generating process of Z_t. However, in Lemma A.4 we establish that the parts of Assumptions 3' and 4' relating to Z_t hold if

$$Z_t = \underline{X}_t\Gamma + \Xi + E_t, \tag{12}$$

where $\underline{X}_t$ is a matrix of fully observable exogenous regressors, Ξ is a matrix of individual effects, and E_t is a matrix of i.i.d. innovations with zero mean and finite $(4+\kappa)$th moments for some $\kappa > 0$, and if $\text{plim}_{n\to\infty} n^{-1}H^{+\prime}\underline{X}^{+}\Gamma$ and $\lim_{n\to\infty} n^{-1}\bar{H}_r^{+\prime}\underline{X}^{+}\Gamma$, for $r = 1, \cdots, q$, are finite with full column rank, with $\underline{X}^{+} = [\underline{X}_1^{+\prime}, \cdots, \underline{X}_{T-1}^{+\prime}]'$.

We note that if Z_t is generated by (12) one could additionally exploit the moment condition $\mathrm{E}(u_t^{+\prime}W_{t,r}^{*}e_{j,t}^{+}) = 0$,[14] with $e_{j,t}^{+}$ denoting the jth column of E_t^{+}, to improve the power of the test (see Liu and Prucha, 2018). However, the advantage of the $\mathcal{I}_y^2(q)$ test statistic as defined in (11) is that it is easy to compute, less demanding on data requirements, and robust to potential misspecification in (12).

3 Equivalence Relationships with LM Test Statistics

In this subsection, we show that, for important specific alternatives, our Moran $\mathcal{I}$ test statistics are equivalent to LM test statistics, with proofs relegated to the online appendix.

[13] Explicit expressions for Φ are given in the proof.
[14] See the proof of Lemma A.4 of Appendix A.

In showing this equivalence we hope to provide additional justification for the selection of the moments employed in forming our generalized Moran $\mathcal{I}$ test statistics and also for the manner in which the moments are aggregated into a single statistic. More specifically, we first specify an elongated vector of moments that interacts the transformed disturbances with time leads and lags of the exogenous variables and weight matrices. Corresponding to the elongated moment vectors, we then derive the GMM LM test statistics (Newey and West, 1987) when the data are generated by a Cliff-Ord type model with higher-order spatial lags under the alternative hypothesis. Implicit in the formulation of the GMM LM test statistics is the construction of a shortened moment vector based on an optimal weighting of the original moments. We then establish the equivalence of the generalized Moran $\mathcal{I}$ test statistics with the GMM LM test statistics. We also show that the proposed test statistics are identical to the ML LM (Rao's score) test statistics when the underlying network structure is invariant over time, i.e., $W_{t,r} = W_{1,r}$ for $r = 1, \cdots, q$ and $t = 1, \cdots, T$.

For the following discussion, we focus on the case where under the null hypothesis the data are generated by

$$y_t = X_t \beta_0 + u_t, \quad \text{and} \quad u_t = \mu + \epsilon_t, \qquad \text{for } t = 1, \cdots, T, \tag{13}$$

where X_t is an $n \times K_X$ matrix of observations on K_X exogenous variables. In this simple case, the IV matrix is identical to the regressor matrix (i.e., $H_t = X_t$) and $\widehat{X}^+ = H^+(H^{+\prime}H^+)^{-1}H^{+\prime}X^+ = X^+$. Consequently, the 2SLS estimator used to estimate (3) degenerates to the OLS estimator $\widehat{\beta} = (X^{+\prime}X^+)^{-1}X^{+\prime}y^+$ with the estimation residuals $\widehat{u}_t^+ = y_t^+ - X_t^+\widehat{\beta}$. Clearly $\widehat{\Sigma}_Q = 0$ under this setup, and thus the $\widehat{\Psi}_Q$ defined in (7) becomes $\widehat{\Phi}_Q$. Furthermore, since $\bar{H}_r^+$ becomes $\bar{X}_r^+ = [\bar{X}_{1,r}^{+\prime}, \cdots, \bar{X}_{T-1,r}^{+\prime}]'$, where $\bar{X}_{t,r} = W_{t,r}X_t$ for $r = 1, \cdots, q$, and $M_{\widehat{Z}Z}$ becomes $M_X = I - X^+(X^{+\prime}X^+)^{-1}X^{+\prime}$, the matrices $\widehat{V}_L$ and $\widehat{\Phi}_L$

defined in (10) turn into

$$\widehat{V}_L = \begin{bmatrix} \bar{X}_1^{+\prime}\widehat{u}^+ \\ \vdots \\ \bar{X}_q^{+\prime}\widehat{u}^+ \end{bmatrix} \quad \text{and} \quad \widehat{\Phi}_L = \widehat{\sigma}^2 \begin{bmatrix} \bar{X}_1^{+\prime}M_X\bar{X}_1^+ & \cdots & \bar{X}_1^{+\prime}M_X\bar{X}_q^+ \\ \vdots & \ddots & \vdots \\ \bar{X}_q^{+\prime}M_X\bar{X}_1^+ & \cdots & \bar{X}_q^{+\prime}M_X\bar{X}_q^+ \end{bmatrix}. \tag{14}$$

3.1 Equivalence of $\mathcal{I}_u^2(q)$ and LM Test Statistics

Suppose under the alternative hypothesis the data are generated by

$$y_t = X_t\beta_0 + u_t, \quad \text{and} \quad u_t = \sum_{r=1}^{q} \rho_{r0} W_{t,r} u_t + \mu + \epsilon_t, \qquad \text{for } t = 1, \cdots, T. \tag{15}$$

Clearly under this setup the null hypothesis H_0^u defined in (4) can be formulated equivalently as $H_0^u : \rho_0 = 0$, where $\rho_0 = (\rho_{10}, \cdots, \rho_{q0})'$.

Let $R_t(\rho) = I_n - \sum_{r=1}^{q} \rho_r W_{t,r}$ and $R_t = R_t(\rho_0)$. Applying the Cochrane-Orcutt transformation and then the Helmert transformation to (15) yields

$$(R_t y_t)^+ = (R_t X_t)^+ \beta_0 + \epsilon_t^+, \qquad \text{for } t = 1, \cdots, T-1. \tag{16}$$

Under the maintained assumptions, $\mathrm{E}(X_s'\epsilon_t^+) = 0$ and $\mathrm{E}(\epsilon_t^{+\prime} W_{s,r} \epsilon_t^+) = \sigma^2 \mathrm{tr}(W_{s,r}) = 0$ for $r = 1, \cdots, q$, $s = 1, \cdots, T$, and $t = 1, \cdots, T-1$, which suggests the following empirical moment function

$$g_t(\theta) = [X_1, \cdots, X_T, W_{1,1}\epsilon_t^+(\theta), \cdots, W_{T,1}\epsilon_t^+(\theta), \cdots, W_{1,q}\epsilon_t^+(\theta), \cdots, W_{T,q}\epsilon_t^+(\theta)]'\epsilon_t^+(\theta), \tag{17}$$

with $\epsilon_t^+(\theta) = [R_t(\rho)(y_t - X_t\beta)]^+$ and $\theta = (\rho', \beta')'$. Let $g(\theta) = [g_1(\theta)', \cdots, g_{T-1}(\theta)']'$, then $\mathrm{E}[g(\theta_0)] = 0$. The corresponding GMM LM test statistic (Newey and West, 1987) for

$H_0^u : \rho_0 = 0$ is given by

$$\text{GMM-LM}_u = g'(\widehat{\theta})\widehat{\Omega}^{-1}\widehat{G}(\widehat{G}'\widehat{\Omega}^{-1}\widehat{G})^{-1}\widehat{G}'\widehat{\Omega}^{-1}g(\widehat{\theta}), \tag{18}$$

where $\widehat{G}$ and $\widehat{\Omega}$ are respectively $G = -\mathrm{E}[\frac{\partial g(\theta)}{\partial \theta}|_{\theta_0}]$ and $\Omega = \mathrm{E}[g(\theta_0)g(\theta_0)']$ evaluated at the restricted estimators $\widehat{\theta} = (0, \widehat{\beta}')'$ and $\widehat{\sigma}^2$.[15] When $\mathrm{E}(\epsilon_t|\mu) = 0$ and $\mathrm{E}(\mu) = 0$, the following proposition shows that the GMM LM test statistic defined in (18) is identical to the generalized Moran $\mathcal{I}$ test statistic $\mathcal{I}_u^2(q)$. The proofs of the propositions in this section are given in the online appendix.

Proposition 1. *Under the maintained assumptions,* $\text{GMM-LM}_u = \mathcal{I}_u^2(q)$.

In the special case where the underlying network structure is invariant over time, i.e., $W_{t,r} = W_{1,r}$ for $r = 1, \cdots, q$ and $t = 1, \cdots, T$, the Helmert transformed model (16) becomes

$$R_1 y_t^+ = R_1 X_t^+ \beta_0 + \epsilon_t^+, \qquad \text{for } t = 1, \cdots, T-1. \tag{19}$$

Under the assumption that $\epsilon = [\epsilon_1', \cdots, \epsilon_T']' \sim N(0, \sigma_0^2 I_{nT})$, we have $\epsilon^+ = [\epsilon_1^{+\prime}, \cdots, \epsilon_{T-1}^{+\prime}]' \sim N(0, \sigma_0^2 I_{n(T-1)})$, and the log-likelihood function of the Helmert transformed model (19) is

$$\ln L(\theta, \sigma^2) = -\frac{n(T-1)}{2}\ln(2\pi\sigma^2) + (T-1)\ln|R_1(\rho)| - \frac{1}{2\sigma^2}\sum_{t=1}^{T-1}\epsilon_t^+(\theta)'\epsilon_t^+(\theta),$$

where $\epsilon_t^+(\theta) = [R_1(\rho)(y_t - X_t\beta)]^+ = R_1(\rho)(y_t^+ - X_t^+\beta)$. The LM test statistic for H_0^u :

[15] The explicit expressions for G and Ω are given in the proof of Proposition 1.

$\rho_0 = 0$ is given by[16]

$$\text{ML-LM}_u = \begin{bmatrix} \frac{\partial \ln L}{\partial \theta'} \\ \frac{\partial \ln L}{\partial \sigma^2} \end{bmatrix}'_{\widehat{\theta},\widehat{\sigma}^2} \left[-\text{E} \left(\begin{array}{cc|} \frac{\partial^2 \ln L}{\partial \theta \partial \theta'} & \frac{\partial^2 \ln L}{\partial \sigma^2 \partial \theta'} \\ \frac{\partial^2 \ln L}{\partial \theta \partial \sigma^2} & \frac{\partial^2 \ln L}{(\partial \sigma^2)^2} \end{array}_{\theta_0, \sigma_0^2} \right) \right]^{-1}_{\widehat{\theta},\widehat{\sigma}^2} \begin{bmatrix} \frac{\partial \ln L}{\partial \theta'} \\ \frac{\partial \ln L}{\partial \sigma^2} \end{bmatrix}_{\widehat{\theta},\widehat{\sigma}^2}. \tag{20}$$

The following proposition shows that the ML LM test statistic defined in (20) is identical to the generalized Moran $\mathcal{I}$ test statistic $\mathcal{I}_u^2(q)$ with time invariant weight matrices.

Proposition 2. *Suppose $W_{t,r} = W_{1,r}$ for $r = 1, \cdots, q$ and $t = 1, \cdots, T$. Under the maintained assumptions,* $\text{ML-LM}_u = \mathcal{I}_u^2(q)$.

3.2 Equivalence of $\mathcal{I}_y^2(q)$ and LM Test Statistics

Now, suppose under the alternative hypothesis the data are generated by

$$y_t = \sum_{r=1}^{q} \lambda_{r0} W_{t,r} y_t + X_t \beta_0 + \sum_{r=1}^{q} W_{t,r} X_t \gamma_{r0} + u_t, \text{ and } u_t = \sum_{r=1}^{q} \rho_{r0} W_{t,r} u_t + \mu + \epsilon_t, \quad \text{for } t = 1, \cdots, T. \tag{21}$$

Under this setup, the null hypothesis H_0^y defined in (9) can be formulated equivalently as $H_0^y : \rho_0 = \lambda_0 = 0$ and $\gamma_{10} = \cdots = \gamma_{q0} = 0$, where $\lambda_0 = (\lambda_{10}, \cdots, \lambda_{q0})'$.

Let $S_t(\lambda) = I_n - \sum_{r=1}^{q} \lambda_r W_{t,r}$ and $S_t = S_t(\lambda_0)$. Applying the Cochrane-Orcutt transformation and then the Helmert transformation to (21) yields

$$(R_t S_t y_t)^+ = (R_t X_t)^+ \beta_0 + \sum_{r=1}^{q} (R_t W_{t,r} X_t)^+ \gamma_{r0} + \epsilon_t^+, \qquad \text{for } t = 1, \cdots, T-1. \tag{22}$$

Under the maintained assumptions, $\text{E}(X_s' \epsilon_t^+) = 0$, $\text{E}(X_s' W_{s,r}' \epsilon_t^+) = 0$ and $\text{E}(\epsilon_t^{+\prime} W_{s,r} \epsilon_t^+) = \sigma^2 \text{tr}(W_{s,r}) = 0$ for $r = 1, \cdots, q$, $s = 1, \cdots, T$, and $t = 1, \cdots, T-1$, which implies the

[16] The explicit expression for the LM test statistic is given in the proof of Proposition 2.

following empirical moment function

$$g_t(\vartheta) = [H, W_{1,1}\epsilon_t^+(\vartheta), \cdots, W_{T,1}\epsilon_t^+(\vartheta), \cdots, W_{1,q}\epsilon_t^+(\vartheta), \cdots, W_{T,q}\epsilon_t^+(\vartheta)]'\epsilon_t^+(\vartheta) \tag{23}$$

with $H = [X_1, \cdots, X_T, W_{1,1}X_1, \cdots, W_{T,1}X_T, \cdots, W_{1,q}X_1, \cdots, W_{T,q}X_T]$, $\epsilon_t^+(\vartheta) = \{R_t(\rho)[S_t(\lambda)y_t - X_t\beta - \sum_{r=1}^q W_{t,r}X_t\gamma_r]\}^+$ and $\vartheta = (\lambda', \rho', \gamma_1', \cdots, \gamma_q', \beta')'$. Let $g(\vartheta) = [g_1(\vartheta)', \cdots, g_{T-1}(\vartheta)']'$, then $\mathrm{E}[g(\vartheta_0)] = 0$. The corresponding GMM LM test statistic (Newey and West, 1987) for $H_0^y : \rho_0 = \lambda_0 = 0$ and $\gamma_{10} = \cdots = \gamma_{q0} = 0$ is given by

$$\text{GMM-LM}_y = g'(\widehat{\vartheta})\widehat{\Omega}^{-1}\widehat{G}(\widehat{G}'\widehat{\Omega}^{-1}\widehat{G})^{-}\widehat{G}'\widehat{\Omega}^{-1}g(\widehat{\vartheta}), \tag{24}$$

where $\widehat{G}$ and $\widehat{\Omega}$ are respectively $G = -\mathrm{E}[\frac{\partial g(\vartheta)}{\partial \vartheta}|_{\vartheta_0}]$ and $\Omega = \mathrm{E}[g(\vartheta_0)g(\vartheta_0)']$ evaluated at the restricted estimators $\widehat{\vartheta} = (0, \widehat{\beta}')'$ and $\widehat{\sigma}^2$.[17] When $\mathrm{E}(\epsilon_t|\mu) = 0$ and $\mathrm{E}(\mu) = 0$, the following proposition shows that the GMM LM test statistic defined in (24) is identical to the generalized Moran $\mathcal{I}$ test statistic $\mathcal{I}_y^2(q)$.

Proposition 3. *Under the maintained assumptions,* $\text{GMM-LM}_y = \mathcal{I}_y^2(q)$.

In the special case where the underlying network structure is invariant over time, i.e., $W_{t,r} = W_{1,r}$ for $r = 1, \cdots, q$ and $t = 1, \cdots, T$, the Helmert transformed model (22) becomes

$$R_1 S_1 y_t^+ = R_1 X_t^+\beta + \sum_{r=1}^q R_1 W_{1,r} X_t^+\gamma_r + \epsilon_t^+, \qquad \text{for } t = 1, \cdots, T-1. \tag{25}$$

Under the assumption that $\epsilon = [\epsilon_1', \cdots, \epsilon_T']' \sim N(0, \sigma_0^2 I_{nT})$, we have $\epsilon^+ = [\epsilon_1^{+\prime}, \cdots, \epsilon_{T-1}^{+\prime}]' \sim$

[17] The explicit expressions for G and Ω are given in the proof of Proposition 3.

$N(0, \sigma_0^2 I_{n(T-1)})$, and the log-likelihood function of the Helmert transformed model (25) is

$$\ln L(\vartheta, \sigma^2) = -\frac{n(T-1)}{2}\ln(2\pi\sigma^2) + (T-1)\ln|R_1(\rho)| + (T-1)\ln|S_1(\lambda)| - \frac{1}{2\sigma^2}\sum_{t=1}^{T-1}\epsilon_t^+(\vartheta)'\epsilon_t^+(\vartheta),$$

where $\epsilon_t^+(\vartheta) = \{R_1(\rho)[S_1(\lambda)y_t - X_t\beta - \sum_{r=1}^q W_{1,r}X_t\gamma_r]\}^+ = R_1(\rho)[S_1(\lambda)y_t^+ - X_t^+\beta - \sum_{r=1}^q W_{1,r}X_t^+\gamma_r]$. The LM test statistic for $H_0^y : \rho_0 = \lambda_0 = 0$ and $\gamma_{10} = \cdots = \gamma_{q0} = 0$ is given by[18]

$$\text{ML-LM}_y = \begin{bmatrix} \frac{\partial \ln L}{\partial \vartheta'} \\ \frac{\partial \ln L}{\partial \sigma^2} \end{bmatrix}'_{\widehat{\vartheta}, \widehat{\sigma}^2} \left[-\mathrm{E}\left(\begin{array}{cc} \frac{\partial^2 \ln L}{\partial \vartheta \partial \vartheta'} & \frac{\partial^2 \ln L}{\partial \sigma^2 \partial \vartheta'} \\ \frac{\partial^2 \ln L}{\partial \vartheta \partial \sigma^2} & \frac{\partial^2 \ln L}{(\partial \sigma^2)^2} \end{array} \Bigg|_{\vartheta_0, \sigma_0^2} \right) \right]^{-1}_{\widehat{\vartheta}, \widehat{\sigma}^2} \begin{bmatrix} \frac{\partial \ln L}{\partial \vartheta'} \\ \frac{\partial \ln L}{\partial \sigma^2} \end{bmatrix}_{\widehat{\vartheta}, \widehat{\sigma}^2}. \tag{26}$$

The following proposition shows that the ML LM test statistic defined in (26) is identical to the generalized Moran $\mathcal{I}$ test statistic $\mathcal{I}_y^2(q)$ with time invariant weight matrices.

Proposition 4. *Suppose* $W_{t,r} = W_{1,r}$ *for* $r = 1, \cdots, q$ *and* $t = 1, \cdots, T$. *Under the maintained assumptions,* $\text{ML-LM}_y = \mathcal{I}_y^2(q)$.

4 Implementation with Endogenous Weight Matrices

In the following we provide a brief discussion on how to implement the generalized Moran $\mathcal{I}$ test when the weight matrix W_t is endogenous in the sense that its elements may be correlated with the error term u_t of the main regression.[19] Suppose the (i, j)th element of W_t is generated by

$$w_{ij,t} = f_{ij,t}(\zeta_t, \eta_t), \tag{27}$$

[18] The explicit expression for the LM test statistic is given in the proof of Proposition 4.

[19] For notational simplicity, we focus on the case with a single endogenous weight matrix W_t. The same argument can be easily extended to the case with multiple endogenous weight matrices $W_{t,1}, \cdots, W_{t,q}$.

where (i) ζ_t is a matrix of exogenous and/or endogenous characteristics, in the sense that they may be correlated with the individual effects μ, but not with the idiosyncratic disturbances ϵ_t, and (ii) η_t is a matrix of potentially endogenous characteristics in the sense that they may also be correlated with the idiosyncratic disturbances ϵ_t.

In light of the above it proves useful to distinguish between two cases of endogeneity of the weight matrix. First, consider the case where the elements of W_t only depend on ζ_t. Since the fixed effects are eliminated by the Helmert transformation, the weight matrix W_t becomes exogenous in the transformed model, and the generalized Moran $\mathcal{I}$ test based on W_t still has the proper size. Second, consider the case where the elements of W_t also depend on η_t. In this case we may attempt to implement the generalized Moran $\mathcal{I}$ test with the endogenous weight matrix W_t replaced by (in the transformed model) exogenous auxiliary weight matrices, which are obtained from a projection of $w_{ij,t}$ onto observed elements of ζ_t (and/or other exogenous variables). For example, suppose $w_{ij,t}$ is determined by a homophily link formation model where individuals with similar characteristics are more likely to form a link, and where the formation process may also depend on endogenous components. Then, we could construct auxiliary exogenous weight matrices based on the similarity between individuals in observed exogenous characteristics (e.g., whether two individuals are of the same gender, race, etc.). As the auxiliary weight matrices are exogenous, the generalized Moran $\mathcal{I}$ test based on the auxiliary weight matrices has the proper size. We conduct Monte Carlo simulations to investigate the performance of the generalized Moran $\mathcal{I}$ test when the weight matrices are endogenous in the online appendix.

5 Monte Carlo Study

In this section, we report on results from a Monte Carlo study of the finite sample properties of the $\mathcal{I}_u^2(q)$ and $\mathcal{I}_y^2(q)$ test statistics. Additional Monte Carlo results are reported in an online appendix.

For our Monte Carlo experiments we utilize two sets of weight matrices, $W_{t,1}$ and $W_{t,2}$, for $t = 1, \dots, T$. To generate those weight matrices we partition n individuals into equal-sized groups with m individuals in each group. Let $\xi_{t,r}^*$ (for $r = 1, 2$) be two independent $n \times 1$ vectors of random variables generated as $\xi_{t,r}^* = \phi_r \xi_{t-1,r}^* + v_{t,r}$, with the initial condition $\xi_{0,r}^* \sim N(0, (1 - \phi_r^2)^{-1} I_n)$ and error term $v_{t,r} \sim N(0, I_n)$. Let $\xi_{t,r}$ (for $r = 1, 2$) be the standardized $\xi_{t,r}^*$ given by $\xi_{t,r} = (1 - \phi_r^2)^{1/2} \xi_{t,r}^*$. Let $D_{t,r}$ (for $r = 1, 2$) be an $n \times n$ zero-diagonal matrix of indicator variables with the (i, j)th element being one if and only if individuals i and j are in the same group and $\left|\xi_{it,r} - \xi_{jt,r}\right| \leq c$, where $\xi_{it,r}$ denotes the ith element of $\xi_{t,r}$ and c a cutoff distance. For an exemplary interpretation, $\xi_{t,r}$ may be thought of as representing some exogenous characteristics of the individuals and the elements of $D_{t,r}$ as reflecting network links based on homophily. For our simulations we set $\phi_1 = 0$ and $\phi_2 = 0.5$ so that $D_{t,1}$ is independent over time and $D_{t,2}$ has a moderate correlation over time. We set the group size to $m = 50$ and the cutoff distance to $c = 0.2$, which generates somewhat sparse networks. The weight matrix $W_{t,r}$ (for $r = 1, 2$) is then obtained by row-sum normalizing $D_{t,r}$ so that each non-zero row of $W_{t,r}$ sums to one.

The proposed $\mathcal{I}^2(q)$ tests utilize quadratic moment conditions with weighting matrices $W_{t,r}^* = \sum_{\tau=1}^T \pi_{t\tau}^2 W_{\tau,r}$, where the $\pi_{t\tau}$ denote the Helmert coefficients. In Section 5.1, we investigate the performance of the proposed $\mathcal{I}^2(q)$ tests relative to alternative tests, which do not use the Helmert weighting.

In Section 5.2, we report on the size, power, and the trade-offs in power of $\mathcal{I}^2(2)$ tests based on both $W_{t,1}$ and $W_{t,2}$, relative to the $\mathcal{I}^2(1)$ tests based on only $W_{t,1}$ or $W_{t,2}$. We find all tests to be properly sized. We also find that the $\mathcal{I}_u^2(2)$ test based on both $W_{t,1}$ and $W_{t,2}$ offers a substantial degree of robustness as compared to the $\mathcal{I}_u^2(1)$ test.

In providing Monte Carlo results on the performance of the $\mathcal{I}_u^2(q)$ and $\mathcal{I}_y^2(q)$ tests we consider a number of different generating processes for the $n \times 1$ vectors of disturbances u_t and endogenous variables y_t. For all simulations, u_t is generated as

$$u_t = \rho_1 W_{t,1} u_t + \rho_2 W_{t,2} u_t + \mu + \epsilon_t.$$

As covariates in the generating processes for y_t we consider spatial lags of y_t, exogenous variables collected in a $n \times 2$ matrix X_t, and an outside endogenous variable y_t^0 generated as

$$y_t^0 = X_t \delta + \xi + e_t.$$

with $\delta = (1,1)'$. The elements of the $n \times 2$ matrix X_t are drawn independently as Uniform$[0,3]$, the individual effects μ and ξ are drawn independently from $N(0, I_n)$ and the random innovations ϵ_t and e_t are generated as dependent $N(0, I_n)$ with $cov(\epsilon_t, e_t) = 0.5 I_n$. Each Monte Carlo experiment is based on 50,000 repetitions.

In addition to the simulation results presented below, in the online appendix, we also report on the performance of the $\mathcal{I}_u^2(q)$ and $\mathcal{I}_y^2(q)$ test statistics when q is large. In particular, we compare the performance of the proposed $\mathcal{I}_u^2(q)$ and $\mathcal{I}_y^2(q)$ tests with the Holm test (Holm, 1979). We find that the Holm test tends to under-reject the null hypothesis. Not surprisingly, the downward size distortion of the Holm test tends to be more severe when q is large and the correlation between $W_{t,r}$ and $W_{t,s}$ is high. In the online appendix we also report on the performance of the $\mathcal{I}_u^2(q)$ and $\mathcal{I}_y^2(q)$ tests when the weight matrices are endogenous. In line with our discussion in the Section 4, we consider two forms of endogeneity. The first case arises when the weight matrix W_t is correlated with the individual effects μ, but not with the idiosyncratic disturbances ϵ_t. The results confirm that in this case we can still use the weight matrix W_t in forming our test statistics. The second case arises when the weight matrix W_t is also correlated with the idiosyncratic disturbances ϵ_t. In this latter case we see that the use of the actual weight matrices W_t can result in substantial size distortions. However, replacing the actual weight matrices with "approximated/projected" weight matrices, which only depend on exogenous variables, yields properly sized tests. The power of those tests will, of course, be application specific and depend on the quality of the approximations.

5.1 Performance of $\mathcal{I}^2(q)$ Tests Relative to Non-Helmert Weighted Tests

In Tables 1-4 below we report on the performance of the $\mathcal{I}_u^2(q)$ and $\mathcal{I}_y^2(q)$ tests relative to alternative tests, which do not use the Helmert weighting. For the $\mathcal{I}_u^2(q)$ test, y_t is generated as

$$y_t = X_t\beta + u_t,$$

and, for the $\mathcal{I}_y^2(q)$ test, y_t is generated as

$$y_t = \lambda_1 W_{t,1} y_t + \lambda_2 W_{t,2} y_t + X_t\beta + W_{t,1} X_t \gamma_1 + W_{t,2} X_t \gamma_2 + u_t,$$

with $\beta = (1,1)'$. We consider a number of scenarios corresponding to different values for the spatial lag parameters as detailed in the tables. As described in Section 2, the proposed $\mathcal{I}_u^2(q)$ and $\mathcal{I}_y^2(q)$ tests utilize the quadratic moments composing $\widehat{V}_Q$, where $\widehat{V}_Q$ is defined by (**6**). Observe that those quadratic moments are based on $W_r^* = \operatorname{diag}_{t=1}^{T-1}\left\{W_{t,r}^*\right\}$, where $W_{t,r}^* = \sum_{\tau=1}^{T} \pi_{t\tau}^2 W_{\tau,r}$ is a weighted average of the spatial weight matrices for periods $t = 1, \ldots, T$, using the squared Helmert coefficients as weights.

As a first alternative test we consider a test where $W_{t,r}^*$ is replaced by a simple time average of the spatial weight matrices, i.e., $\overline{W}_r = T^{-1} \sum_{t=1}^{T} \mathring{W}_{t,r}$. Clearly, this test can also be viewed as being obtained by replacing the matrices W_r^* by $I_{T-1} \otimes \overline{W}_r$ and replacing $\mathrm{tr}(\mathring{W}_r^* \mathring{W}_s^*)$ by $(T-1)\mathrm{tr}(\overline{W}_r \overline{W}_s)$ in $\widehat{\Phi}_Q$ defined in (**6**). In the following we refer to this alternative test as the "time-average weighted" $\mathcal{I}^2$ test.

As a second alternative test we consider a test where $W_{t,r}^*$ is replaced by the first period spatial weight matrix $W_{1,r}$, thus ignoring the time variation of the spatial weight matrices. Clearly, this test can also be viewed as being obtained by replacing the matrices W_r^* by $I_{T-1} \otimes W_{1,r}$ and replacing $\mathrm{tr}(\mathring{W}_r^* \mathring{W}_s^*)$ by $(T-1)\mathrm{tr}(\mathring{W}_{1,r} \mathring{W}_{1,s})$ in $\widehat{\Phi}_Q$ defined in (**6**). In the following we refer to this alternative test as the "initial-period weighted" $\mathcal{I}^2$ test. Of course, in settings where the spatial weight matrices do not vary over time, the proposed $\mathcal{I}_u^2(q)$ and $\mathcal{I}_y^2(q)$ tests and the alternative tests are all identical.

[Insert Tables 1-4 here]

Tables 1-4 show that the $\mathcal{I}_u^2(q)$ and $\mathcal{I}_y^2(q)$ tests as defined in Section 2, and locally labeled in those tables as "Helmert weighted" $\mathcal{I}_u^2(q)$ and $\mathcal{I}_y^2(q)$ tests for clarity, have the correct size and strictly higher power than the other two alternative tests. Henceforth, we focus our attention on the performance of the $\mathcal{I}_u^2(q)$ and $\mathcal{I}_y^2(q)$ tests as defined previously with $W_{t,r}^* = \sum_{\tau=1}^{T} \pi_{t\tau}^2 W_{\tau,r}$.

5.2 Relative Performance of $\mathcal{I}^2(1)$ and $\mathcal{I}^2(2)$ Tests

In the following we report on the performance of the $\mathcal{I}^2(2)$ tests based on both $W_{t,1}$ and $W_{t,2}$, relative the the performance of the $\mathcal{I}^2(1)$ tests based on only $W_{t,1}$ or $W_{t,2}$ for different scenarios.

For the $\mathcal{I}_u^2(q)$ test we consider four different data generating processes for y_t:

$$
\begin{aligned}
y_t &= X_t\beta + u_t, && (28)\\
y_t &= \beta y_t^0 + u_t, && (29)\\
y_t &= \lambda W_{t,1} y_t + X_t\beta + u_t, && (30)\\
y_t &= \lambda W_{t,1} y_t + X_t\beta + W_{t,1} X_t \gamma + u_t. && (31)
\end{aligned}
$$

Under the null hypothesis H_0^u we have $u_t = \mu + \epsilon_t$. In the first scenario where y_t is generated by (28), all covariates are exogenous. We set $\beta = (1,1)'$, and report the simulation results on the $\mathcal{I}_u^2(q)$ test for this scenario in Table 5. The second scenario, where y_t is generated by (29), allows for an "outside" endogenous covariate y_t^0. We set $\beta = 1$ and use X_t as instruments for y_t^0. The corresponding simulation results are reported in Table 6. In the last two scenarios, we allow for endogenous covariates in the form of spatial lags of y_t under the null. For (30), we set $\lambda = 0.5$ and $\beta = (1,1)'$, and use $W_{t,1}X_t$ as instruments for $W_{t,1}y_t$. For (31), we set $\lambda = 0.5$ and $\beta = \gamma = (1,1)'$, and use $W_{t,1}^2 X_t$ as additional instruments for $W_{t,1}y_t$. The corresponding simulation results are reported in Tables 7 and 8 respectively.

[Insert Tables 5-8 here]

Overall, we find that the actual sizes of the $\mathcal{I}_u^2(q)$ tests are close to the asymptotic nominal size of 0.05 and that the power increases as the magnitudes of the spatial autoregressive parameters (cross sectional correlation) increases. We also find that the $\mathcal{I}_u^2(2)$ tests based on both $W_{t,1}$ and $W_{t,2}$ can offer a substantial degree of robustness as compared to the $\mathcal{I}_u^2(1)$ tests. Consider the case where $\rho_1 \neq 0$ and $\rho_2 = 0$. In this scenario, $W_{t,1}$ correctly models the network topology, whereas $W_{t,2}$ does not. We expect the $\mathcal{I}_u^2(1)$ test based on $W_{t,1}$ to outperform other tests. Indeed, compared to the $\mathcal{I}_u^2(1)$ test based on $W_{t,1}$, we find a substantial power loss for the $\mathcal{I}_u^2(1)$ test based on the "wrong" weight matrix $W_{t,2}$, while there is only a modest power loss for the $\mathcal{I}_u^2(2)$ test based both $W_{t,1}$ and $W_{t,2}$. The results suggest that, unless a researcher is very sure which weight matrices properly model the network topology, using an $\mathcal{I}_u^2(q)$ test that combines several candidate weight matrices can be an attractive and robust approach.

To explore the performance of the $\mathcal{I}_y^2(q)$ test we consider two different data generating processes for y_t:

$$y_t = \lambda_1 W_{t,1} y_t + \lambda_2 W_{t,2} y_t + X_t\beta + W_{t,1} X_t \gamma_1 + W_{t,2} X_t \gamma_2 + u_t, \tag{32}$$

$$y_t = \lambda_1 W_{t,1} y_t + \lambda_2 W_{t,2} y_t + \beta y_t^0 + \gamma_1 W_{t,1} y_t^0 + \gamma_2 W_{t,2} y_t^0 + u_t. \tag{33}$$

We set $\beta = (1,1)'$ in (32) and $\beta = 1$ in (33). In the scenario where y_t is generated by (32), the null hypothesis H_0^y corresponds to $y_t = X_t\beta + \mu + \varepsilon_t$ and in this scenario all covariates are exogenous under the null. By contrast, in the scenario where y_t is generated by (33), the null hypothesis H_0^y corresponds to $y_t = \beta y_t^0 + \mu + \varepsilon_t$ and in this scenario we allow for an endogenous covariate y_t^0 under the null. The endogenous covariate y_t^0 is an "outside endogenous variable" in the sense of a "classical" simultaneous equation system. The simulation results for the two scenarios are reported in Tables 9 and 10 respectively.

[Insert Tables 9 and 10 here]

Similar to the simulation results for the $\mathcal{I}_u^2(q)$ tests, we find that the actual sizes of the $\mathcal{I}_y^2(q)$ tests are close to the asymptotic nominal size of 0.05 and the power of the tests increases as the amount of cross sectional dependence increases. Furthermore, when cross sectional dependence is based on $W_{t,1}$ but not on $W_{t,2}$, the power loss of the $\mathcal{I}_y^2(2)$ test using both $W_{t,1}$ and $W_{t,2}$ is much less than that of the $\mathcal{I}_u^2(1)$ test based on the "wrong" weight matrix $W_{t,2}$. This indicates that the $\mathcal{I}_y^2(q)$ test combining multiple candidate weight matrices can provide the empirical researcher some important level of robustness when they are unsure about which weight matrices properly model the network topology.

6 Conclusion

In this paper, we introduce generalizations of the Moran $\mathcal{I}$ tests for network-generated cross-sectional dependence in a panel data setting with unit-specific fixed or random effects and time-varying network structures. The tests are applicable to both spatial and social network structures. Our tests are intuitively motivated. They are geared towards situations where researchers are uncertain as to how to choose among multiple potential spatial weight matrices or adjacency matrices. While our tests are intuitive, they are also shown to be equivalent to Lagrange Multiplier tests for specific, but widely considered, model formulations under the alternative hypothesis. We establish the limiting distribution of the test statistics and the rejection regions of the tests for a given significance level under fairly general assumptions, which should make the test useful in a wide range of empirical research. Our test statistics are relatively simple and easy to compute. This simplicity is, in particular, due to adopting the Helmert transformation to eliminate unit-specific effects, which may be random or fixed, and by selecting the quadratic moments such that the diagonal elements of the weight matrices of the quadratic forms are zero.

We also conduct Monte Carlo experiments to investigate the finite sample performance of the proposed tests. Overall, the results suggest that the proposed tests perform well with proper size and reasonable power. The loss in power from using more weight matrices than needed is mostly modest. This suggests that the tests can indeed provide robustness against uncertainty about the proper choice of the spatial weight matrices or adjacency matrices. We also discuss how the testing framework can be extended to the case where network formation is endogenous.

References

Arellano, M. and Bover, O. (1995). Another look at the instrumental variable estimation of error-components models, *Journal of Econometrics* **68**: 29–51.

Baltagi, B. H., Feng, Q. and Kao, C. (2012). A Lagrange Multiplier test for cross-sectional dependence in a fixed effects panel data model, *Journal of Econometrics* **170**: 164–177.

Baltagi, B. H., Kao, C. and Peng, B. (2016). Testing cross-sectional correlation in large panel data models with serial correlation, *Econometrics* **4**: 1–24.

Baltagi, B. H. and Liu, L. (2008). Testing for random effects and spatial lag dependence in panel data models, *Statistics & Probability Letters* **78**: 3304–3306.

Baltagi, B. H., Song, S. H., Jung, B. C. and Koh, W. (2007). Testing for serial correlation, spatial autocorrelation and random effects using panel data, *Journal of Econometrics* **140**: 5–51.

Baltagi, B. H., Song, S. H. and Koh, W. (2003). Testing panel data regression models with spatial error correlation, *Journal of Econometrics* **117**: 123–150.

Baltagi, B. H., Song, S. H. and Kwon, J. H. (2009). Testing for heteroskedasticity and spatial correlation in a random effects panel data model, *Computational Statistics & Data Analysis* **53**: 2897–2922.

Baltagi, B. H. and Yang, Z. (2013). Standardized LM tests for spatial error dependence in linear or panel regressions, *Econometrics Journal* **16**: 103–134.

Bera, A. K., Doğan, O., Taşpınar, S. and Leiluo, Y. (2019). Robust LM tests for spatial dynamic panel data models, *Regional Science and Urban Economics* **76**: 47–66.

Blume, L. E., Brock, W. A., Durlauf, S. N. and Jayaraman, R. (2015). Linear social interactions models, *Journal of Political Economy* **123**: 444–496.

Born, B. and Breitung, J. (2011). Simple regression-based tests for spatial dependence, *Econometrics Journal* **14**: 330–342.

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

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

Carrell, S. E., Sacerdote, B. I. and West, J. E. (2013). From natural variation to optimal policy? The importance of endogenous peer group formation, *Econometrica* **81**: 855–882.

Cliff, A. and Ord, J. (1973). *Spatial Autecorrelation*, Pion, London.

Cliff, A. and Ord, J. (1981). *Spatial Processes, Models and Applications*, Pion, London.

Davezies, L., D'Haultfoeuille, X. and Fougère, D. (2009). Identification of peer effects using group size variation, *Econometrics Journal* **12**: 397–413.

de Paula, Á. (2017). Econometrics of network models, Vol. 1 of *Advances in Economics and Econometrics: Eleventh World Congress*, Cambridge University Press, pp. 268–323.

Debarsy, N. and Ertur, C. (2010). Testing for spatial autocorrelation in a fixed effects panel data model, *Regional Science and Urban Economics* **40**: 453–470.

Drukker, D. M., Egger, P. H. and Prucha, I. R. (2023). Simultaneous equations models with higher-order spatial or social network interactions, *Econometric Theory* **39**: 1154âĂŞ1201.

Durbin, J. and Watson, G. S. (1950). Testing for serial correlation in least squares regression: I, *Biometrika* **37**: 409–428.

Durbin, J. and Watson, G. S. (1951). Testing for serial correlation in least squares regression. II, *Biometrika* **38**: 159–177.

Hall, P. and Heyde, C. C. (1981). Rates of convergence in the martingale central limit theorem, *The Annals of Probability* **9**: 395 – 404.

He, M. and Lin, K.-P. (2015). Testing spatial effects and random effects in a nested panel data model, *Economic Letters* **135**: 85–91.

Hillier, G. H. and Martellosio, F. (2018). Exact likelihood inference in group interaction network models, *Econometric Theory* **34**: 383–415.

Holm, S. (1979). A simple sequentially rejective multiple test procedure, *Scandinavian Journal of Statistics* **6**: 65–70.

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

Kelejian, H. H. and Prucha, I. R. (2004). Estimation of simultaneous systems of spatially interrelated cross sectional equations, *Journal of Econometrics* **118**: 27–50.

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

Kelly, M. (2019). The standard errors of persistence. Working paper, University College Dublin.

King, M. L. (1980). Robust tests for spherical symmetry and their application to least squares regression, *The Annals of Statistics* **8**: 1265–1271.

King, M. L. (1981). A small sample property of the Cliff-Ord test for spatial correlation, *Journal of the Royal Statistical Society: B* **43**: 263–264.

Kolaczyk, J. (2009). *Statistical Analysis of Network Data: Methods and Models*, Springer, New York.

Kuersteiner, G. M. and Prucha, I. R. (2020). Dynamic spatial panel models: Networks, common shocks, and sequential exogeneity, *Econometrica* **88**: 2109–2146.

Lee, L. F. (2007). Identification and estimation of econometric models with group interactions, contextual factors and fixed effects, *Journal of Econometrics* **140**: 333–374.

Lee, L. F., Liu, X. and Lin, X. (2010). Specification and estimation of social interaction models with network structures, *The Econometrics Journal* **13**: 145–176.

Lee, L. F. and Yu, J. (2014). Efficient GMM estimation of spatial dynamic panel data models with fixed effects, *Journal of Econometrics* **180**: 174–197.

Leeb, H. and Pötscher, B. M. (2003). The finite-sample distribution of post-model-selection estimators and uniform versus nonuniform approximations, *Econometric Theory* **19**: 100–142.

Liu, X. and Prucha, I. R. (2018). A robust test for network generated dependence, *Journal of Econometrics* **207**: 92–113.

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

Moran, P. (1950). Notes on continuous stochastic phenomena, *Biometrika* **37**: 17–23.

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

Pesaran, M. H. (2004). General diagnostic test for cross section dependence in panels. IZA discussion paper No. 1240.

Pesaran, M. H., Ullah, A. and Yamagata, T. (2008). A bias-adjusted LM test of error cross section independence, *The Econometrics Journal* **11**: 105–127.

Pötscher, B. M. and Prucha, I. R. (1997). *Dynamic Nonlinear Econometric Models: Asymptotic Theory*, Springer.

Robinson, P. M. (2008). Correlation testing in time series, spatial and cross-sectional data, *Journal of Econometrics* **147**: 5–16.

Taşpınar, S., Doğan, O. and Bera, A. K. (2017). GMM gradient tests for spatial dynamic panel data models, *Regional Science and Urban Economics* **65**: 65–88.

Yang, Z. (2018). Unified m-estimation of fixed-effects spatial dynamic models with short panels, *Journal of Econometrics* **205**: 423–447.

Yang, Z. (2021). Joint tests for dynamic and spatial effects in short panels with fixed effects and heteroskedasticity, *Empirical Economics* **60**: 51–92.

Appendices

A Moments of Functions of Transformed Disturbances

As before, let $u_t^+ = \sum_{\tau=1}^T \pi_{t\tau} u_\tau$ denote the Helmert transformed disturbances. Recall that $\sum_{\tau=1}^T \pi_{t\tau} = 0$ and thus $u_t^+ = \epsilon_t^+$, and that the transformation is orthonormal in that $\sum_{\tau=1}^T \pi_{t\tau}^2 = 1$, and $\sum_{\tau=1}^T \pi_{t\tau}\pi_{s\tau} = 0$ for $t \neq s$. Furthermore, let $u_t^\times = \sum_{\tau=1}^T \varpi_{t\tau} u_\tau$ denote some generically transformed disturbances with $\sum_{\tau=1}^T \varpi_{t\tau} = 0$. Observe that $u_t^\times = \epsilon_t^\times$, but the transformation may not be orthonormal. The results given below are formulated for transformations of ϵ_t, but since $u_t^+ = \epsilon_t^+$ and $u_t^\times = \epsilon_t^\times$ the same result also holds for transformations of u_t.

Assumption A.1. *Let $\epsilon = (\epsilon_1', \cdots, \epsilon_T')'$ with $\epsilon_t = (\epsilon_{1t}, \cdots, \epsilon_{nt})'$ denoting vectors of innovations, where the elements $\{\epsilon_{it} : i = 1, \cdots, n, t = 1, \cdots T\}$ are i.i.d. with $\mathrm{E}(\epsilon_{it}) = 0$, $\mathrm{E}(\epsilon_{it}^2) = \sigma^2$, $\mathrm{E}(\epsilon_{it}^3) = \mu_3$ and $\mathrm{E}(\epsilon_{it}^4) = \mu_4$ finite.*

Lemma A.1. *Suppose Assumption A.1 holds. Let A and B be non-stochastic $n \times n$ matrices, and let a and b be non-stochastic $n \times 1$ vectors. For $t, s = 1, \cdots, T-1$, let*

$$V_t^A = \epsilon_t^{\times\prime} A \epsilon_t^\times + a'\epsilon_t^\times, \qquad V_s^B = \epsilon_s^{\times\prime} B \epsilon_s^\times + b'\epsilon_s^\times,$$

then $\mathrm{E}(V_t^A) = \sigma^2 \sum_{\tau=1}^T \varpi_{t\tau}^2 \mathrm{tr}(A)$ and

$$\begin{aligned} \mathrm{Cov}(V_t^A, V_s^B) &= \sigma^4[\mathrm{tr}(AB + AB')](\sum_{\tau=1}^T \varpi_{t\tau}\varpi_{s\tau})^2 + \sigma^2 a'b \sum_{\tau=1}^T \varpi_{t\tau}\varpi_{s\tau} \\ &\quad + \mu_3[\sum_{\tau=1}^T \varpi_{t\tau}\varpi_{s\tau}^2 a' \mathrm{vec}_D(B) + \sum_{\tau=1}^T \varpi_{s\tau}\varpi_{t\tau}^2 b' \mathrm{vec}_D(A)] \\ &\quad + (\mu_4 - 3\sigma^4) \sum_{\tau=1}^T \varpi_{t\tau}^2 \varpi_{s\tau}^2 \mathrm{vec}_D(A)' \mathrm{vec}_D(B). \end{aligned} \tag{A.1}$$

Furthermore, for $\epsilon_t^{\times} = \epsilon_t^{+}$*, that is for Helmert transformed disturbances, and if additionally* $\mathrm{vec}_D(A) = \mathrm{vec}_D(B) = 0$*, we have* $\mathrm{E}(V_t^A) = 0$ *and*

$$\mathrm{Cov}(V_t^A, V_s^B) = \begin{cases} \sigma^4 \mathrm{tr}(AB + AB') + \sigma^2 a'b & \textit{for} \quad t = s \\ 0 & \textit{for} \quad t \neq s. \end{cases} . \tag{A.2}$$

Remark: The results in (A.1) can be readily used to establish that $\mathrm{E}(\epsilon_{it}^{+}) = 0$, $\mathrm{E}(\epsilon_{it}^{+})^2 = \sigma^2$, $\mathrm{E}(\epsilon_{it}^{+})^3 = \mu_3 \sum_{\tau=1}^{T} \pi_{t\tau}^3$, and $\mathrm{E}(\epsilon_{it}^{+})^4 = \mu_4 \sum_{\tau=1}^{T} \pi_{t\tau}^4 + 3\sigma^4(1 - \sum_{\tau=1}^{T} \pi_{t\tau}^4)$. Furthermore, $\mathrm{E}(\epsilon_{it}^{+}\epsilon_{is}^{+}) = 0$ for $t \neq s$, and $\mathrm{E}(\epsilon_{it}^{+}\epsilon_{jt}^{+}) = 0$ for $i \neq j$ since $(\epsilon_{it})_{t=1}^{T}$ and $(\epsilon_{jt})_{t=1}^{T}$ are independent for $i \neq j$.

Proof of Lemma A.1. [20] Observe that

$$\begin{aligned} \epsilon_t^{\times\prime} A \epsilon_t^{\times} &= \sum_{\varsigma=1}^{T} \sum_{\tau=1}^{T} \varpi_{t\varsigma} \varpi_{t\tau} \epsilon_{\varsigma}' A \epsilon_{\tau} = \epsilon' C \epsilon, \\ a' \epsilon_t^{\times} &= \sum_{\tau=1}^{T} a' \varpi_{t\tau} \epsilon_{\tau} = c' \epsilon, \end{aligned}$$

where $C = [C_{\varsigma\tau}]_{\varsigma,\tau=1,\cdots,T}$ and $c = (c_1', \cdots, c_T')'$, with $C_{\varsigma\tau} = \varpi_{t\varsigma}\varpi_{t\tau} A$ and $c_\tau = \varpi_{t\tau} a$. Similarly, $\epsilon_s^{\times\prime} B \epsilon_s^{\times} = \epsilon' D \epsilon$ and $b' \epsilon_s^{\times} = d' \epsilon$, where $D = [D_{\varsigma\tau}]_{\varsigma,\tau=1,\cdots,T}$ with $D_{\varsigma\tau} = \varpi_{s\varsigma}\varpi_{s\tau} B$ and $d = (d_1', \cdots, d_T')'$ with $d_\tau = \varpi_{s\tau} b$. Using Lemma A.1 in Kelejian and Prucha (2010), it follows that

$$\mathrm{E}(V_t^A) = \mathrm{E}(\epsilon' C \epsilon + c' \epsilon) = \sigma^2 \mathrm{tr}(C) = \sigma^2 \sum_{\tau=1}^{T} \varpi_{t\tau}^2 \mathrm{tr}(A), \tag{A.3}$$

[20] In proving the lemma we utilize results in Kelejian and Prucha (2010). Alternatively, the lemma could be established by specializing results in Kuersteiner and Prucha (2020).

which proves the first claim, and

$$\text{Cov}(V_t^A, V_s^B) = \text{Cov}[(\epsilon' C\epsilon + c'\epsilon)(\epsilon' D\epsilon + d'\epsilon)] \qquad \text{(A.4)}$$

$$= \sigma^4[\text{tr}(CD) + \text{tr}(CD')] + \sigma^2 c'd + \mu_3[c'\text{vec}_D(D) + d'\text{vec}_D(C)] + (\mu_4 - 3\sigma^4)\text{vec}_D(C)'\text{vec}_D(D).$$

Observe that

$$\text{tr}(CD) = \sum_{\varsigma=1}^{T}\sum_{\tau=1}^{T} \text{tr}(C_{\varsigma\tau}D_{\tau\varsigma}) = \sum_{\varsigma=1}^{T}\sum_{\tau=1}^{T} \text{tr}(\varpi_{t\varsigma}\varpi_{t\tau}A\varpi_{s\varsigma}\varpi_{s\tau}B) = \left[\sum_{\tau=1}^{T}\varpi_{t\tau}\varpi_{s\tau}\right]^2 \text{tr}(AB),$$

and analogously $\text{tr}(CD') = \left[\sum_{\tau=1}^{T}\varpi_{t\tau}\varpi_{s\tau}\right]^2 \text{tr}(AB')$. Additionally, observe that $\text{vec}_D(C) = [\varpi_{t1}^2\text{vec}_D(A)', \cdots, \varpi_{tT}^2\text{vec}_D(A)']'$, $\text{vec}_D(D) = [\varpi_{s1}^2\text{vec}_D(B)', \cdots, \varpi_{sT}^2\text{vec}_D(B)']'$, $c = [\varpi_{t1}a, \cdots, \varpi_{tT}a]$, and $d = [\varpi_{s1}b, \cdots, \varpi_{sT}b]$, and thus

$$\begin{aligned}
c'd &= \sum_{\tau=1}^{T}\varpi_{t\tau}\varpi_{s\tau}a'b = a'b\sum_{\tau=1}^{T}\varpi_{t\tau}\varpi_{s\tau}, \\
c'\text{vec}_D(D) &= \sum_{\tau=1}^{T}\varpi_{t\tau}\varpi_{s\tau}^2 a'\text{vec}_D(B), \\
d'\text{vec}_D(C) &= \sum_{\tau=1}^{T}\varpi_{s\tau}\varpi_{t\tau}^2 b'\text{vec}_D(A), \\
\text{vec}_D(C)'\text{vec}_D(D) &= \sum_{\tau=1}^{T}\varpi_{t\tau}^2\varpi_{s\tau}^2\text{vec}_D(A)'\text{vec}_D(B).
\end{aligned}$$

Substitution of these expressions into (A.4) completes the proof of the second claim. The remaining claims follow immediately because of the orthonormality of the weights of the Helmert transformation. □

Lemma A.2. *Suppose Assumption A.1 holds. Let* $A = [A_{st}]_{s,t=1,\cdots,T-1}$, $B = [B_{st}]_{s,t=1,\cdots,T-1}$, $a = (a_1', \cdots, a_{T-1}')'$, *and* $b = (b_1', \cdots, b_{T-1}')'$, *where* A_{st} *and* B_{st} *are non-stochastic* $n \times n$

matrices, and a_t *and* b_t *are non-stochastic* $n \times 1$ *vectors. Now let*

$$V^A = \sum_{s=1}^{T-1} \sum_{t=1}^{T-1} \epsilon_s^{+\prime} A_{st} \epsilon_t^+ + \sum_{t=1}^{T-1} a_t' \epsilon_t^+, \qquad V^B = \sum_{s=1}^{T-1} \sum_{t=1}^{T-1} \epsilon_s^{+\prime} B_{st} \epsilon_t^+ + \sum_{t=1}^{T-1} b_t' \epsilon_t^+,$$

Then $\mathrm{E}(V^A) = \sigma^2 \mathrm{tr}(A)$ *and*

$$\begin{aligned} \mathrm{Cov}(V^A, V^B) &= \sigma^4 \mathrm{tr}(AB + AB') + \sigma^2 a'b \\ &\quad + \mu_3 (c' \mathrm{vec}_D(D) + d' \mathrm{vec}_D(C)) + (\mu_4 - 3\sigma^4) \mathrm{vec}_D(C)' \mathrm{vec}_D(D), \end{aligned}$$

where $C = [C_{\varsigma\tau}]$*,* $D = [D_{\varsigma\tau}]$*,* $c = (c_1', \cdots, c_T')'$ *and* $d = (d_1', \cdots, d_T')'$*, with* $C_{\varsigma\tau} = \sum_{s=1}^{T-1} \sum_{t=1}^{T-1} \pi_{s\varsigma} \pi_{t\tau} A_{st}$*,* $D_{\varsigma\tau} = \sum_{s=1}^{T-1} \sum_{t=1}^{T-1} \pi_{s\varsigma} \pi_{t\tau} B_{st}$*,* $c_\tau = \sum_{t=1}^{T-1} \pi_{t\tau} a_t$ *and* $d_\tau = \sum_{t=1}^{T-1} \pi_{t\tau} b_t$*. Clearly if* $\mathrm{vec}_D(A) = \mathrm{vec}_D(B) = 0$*, then* $\mathrm{vec}_D(C) = \mathrm{vec}_D(D) = 0$*, and in this case the terms involving the third and fourth moments of the elements of* ϵ *are zero.*

Proof of Lemma A.2. Observe that

$$\sum_{s=1}^{T-1} \sum_{t=1}^{T-1} \epsilon_s^{+\prime} A_{st} \epsilon_t^+ = \sum_{s=1}^{T-1} \sum_{t=1}^{T-1} \sum_{\varsigma=1}^{T} \sum_{\tau=1}^{T} \pi_{s\varsigma} \pi_{t\tau} \epsilon_\varsigma' A_{st} \epsilon_\tau = \sum_{\varsigma=1}^{T} \sum_{\tau=1}^{T} \epsilon_\varsigma' C_{\varsigma\tau} \epsilon_\tau = \epsilon' C \epsilon$$

and

$$\sum_{t=1}^{T-1} a_t' \epsilon_t^+ = \sum_{t=1}^{T-1} \sum_{\tau=1}^{T} a_t' \pi_{t\tau} \epsilon_\tau = \sum_{\tau=1}^{T} c_\tau' \epsilon_\tau = c' \epsilon,$$

and similarly $\sum_{s=1}^{T-1} \sum_{t=1}^{T-1} \epsilon_s^{+\prime} B_{st} \epsilon_t^+ = \epsilon' D \epsilon$ and $\sum_{t=1}^{T-1} b_t' \epsilon_t^+ = d' \epsilon$. Then by Lemma A.1 in Kelejian and Prucha (2010)

$$\mathrm{E}(V^A) = \sigma^2 \mathrm{tr}(C), \qquad \mathrm{E}(V^B) = \sigma^2 \mathrm{tr}(D), \tag{A.5}$$

and

$$\operatorname{Cov}(V^A, V^B) = \sigma^4[\operatorname{tr}(CD) + \operatorname{tr}(CD')] + \sigma^2 c'd + \mu_3(c'\operatorname{vec}_D(D) + d'\operatorname{vec}_D(C)) + (\mu_4 - 3\sigma^4)\operatorname{vec}_D(C)'\operatorname{vec}_D(D). \quad \text{(A.6)}$$

As $\sum_{\tau=1}^{T} \pi_{s\tau}\pi_{t\tau} = 0$ for $s \neq t$ and $\sum_{\tau=1}^{T} \pi_{t\tau}^2 = 1$, we have

$$\begin{aligned}
\operatorname{tr}(C) &= \sum_{\tau=1}^{T} \operatorname{tr}(\sum_{s=1}^{T-1}\sum_{t=1}^{T-1} \pi_{s\tau}\pi_{t\tau} A_{st}) = \sum_{s=1}^{T-1}\sum_{t=1}^{T-1} \operatorname{tr}[(\sum_{\tau=1}^{T} \pi_{s\tau}\pi_{t\tau}) A_{st}] = \sum_{t=1}^{T-1} \operatorname{tr}(A_{tt}) = \operatorname{tr}(A), \\
\operatorname{tr}(CD) &= \sum_{\varsigma=1}^{T}\sum_{\tau=1}^{T} \operatorname{tr}(C_{\varsigma\tau} D_{\tau\varsigma}) = \sum_{\varsigma=1}^{T}\sum_{\tau=1}^{T} \operatorname{tr}(\sum_{s=1}^{T-1}\sum_{t=1}^{T-1} \pi_{s\varsigma}\pi_{t\tau} A_{st} \sum_{u=1}^{T-1}\sum_{v=1}^{T-1} \pi_{u\varsigma}\pi_{v\tau} B_{vu}) \\
&= \sum_{s=1}^{T-1}\sum_{t=1}^{T-1}\sum_{u=1}^{T-1}\sum_{v=1}^{T-1} \operatorname{tr}[(\sum_{\varsigma=1}^{T} \pi_{s\varsigma}\pi_{u\varsigma})(\sum_{\tau=1}^{T} \pi_{t\tau}\pi_{v\tau}) A_{st} B_{vu}] = \sum_{s=1}^{T-1}\sum_{t=1}^{T-1} \operatorname{tr}(A_{st} B_{ts}) = \operatorname{tr}(AB),
\end{aligned}$$

and

$$c'd = \sum_{\tau=1}^{T} c_\tau' d_\tau = \sum_{\tau=1}^{T}\sum_{s=1}^{T-1}\sum_{t=1}^{T-1} \pi_{s\tau}\pi_{t\tau} a_s' b_t = \sum_{s=1}^{T-1}\sum_{t=1}^{T-1} (\sum_{\tau=1}^{T} \pi_{s\tau}\pi_{t\tau}) a_s' b_t = \sum_{t=1}^{T-1} a_t' b_t = a'b.$$

Similarly, $\operatorname{tr}(D) = \operatorname{tr}(B)$ and $\operatorname{tr}(CD') = \operatorname{tr}(AB')$. The claims of the lemma are now readily verified upon substitution of the above expressions into (A.5) and (A.6). □

Assumption A.2. *Let* $\xi = (\xi_1', \cdots, \xi_G')'$, *with* $\xi_g = (\xi_{1,g}', \cdots, \xi_{T,g}')'$ *and* $\xi_{t,g} = (\xi_{1t,g}, \cdots, \xi_{nt,g})'$ *denoting some vectors of basic innovations, where the elements* $\{\xi_{it,g} : i = 1, \cdots, n, t = 1, \cdots T, g = 1, \cdots, G\}$ *are i.i.d. with* $\operatorname{E}(\xi_{it,g}) = 0$, $\operatorname{E}(\xi_{it,g}^2) = 1$, $\operatorname{E}(\xi_{it,g}^3) = \mu_3^\xi$ *and* $\operatorname{E}(\xi_{it,g}^4) = \mu_4^\xi$ *finite.*

Lemma A.3. *Suppose Assumption A.2 holds. Let* $A = \operatorname{diag}_{t=1}^{T-1}\{A_t\}$, $B = \operatorname{diag}_{t=1}^{T-1}\{B_t\}$, $a = (a_1', \cdots, a_{T-1}')'$, $b = (b_1', \cdots, b_{T-1}')'$ *where* A_t *and* B_t *be non-stochastic* $n \times n$ *matrices, and* a_t *and* b_t *are non-stochastic* $n \times 1$ *vectors. Let* $P = [p_{gq}]$ *be a non-stochastic* $G \times G$

matrix and $p_{g\cdot}$ *denote the gth row of* P. *For* $g = 1, \cdots, G$, *let* $\epsilon_g = (\epsilon'_{1,g}, \cdots, \epsilon'_{T,g})$ *with* $\epsilon_{t,g} = (\epsilon_{1t,g}, \cdots, \epsilon_{nt,g})'$ *be generated as*

$$\epsilon_g = \sum_{q=1}^{G} p_{gq}\xi_q = (p_{g\cdot} \otimes I_{nT})\xi.$$

Observe that $\mathrm{E}(\epsilon_g) = 0$ *and* $\mathrm{Cov}(\epsilon_g, \epsilon_h) = \sigma_{gh} I_{nT}$ *with* $\sigma_{gh} = \sum_{q=1}^{G} p_{gq}p_{hq}$. *Let* g_1, g_2, g_3, g_4 *be distinct elements of* $\{1, \cdots, G\}$ *and let*

$$\begin{aligned} V^A &= (\epsilon_{g_1}^{+\prime} A \epsilon_{g_2}^{+} + \epsilon_{g_1}^{+\prime} a) = \sum_{t=1}^{T-1} (\epsilon_{t,g_1}^{+\prime} A_t \epsilon_{t,g_2}^{+} + \epsilon_{t,g_1}^{+\prime} a_t), \\ V^B &= (\epsilon_{g_3}^{+\prime} B \epsilon_{g_4}^{+} + \epsilon_{g_3}^{+\prime} b) = \sum_{t=1}^{T-1} (\epsilon_{t,g_3}^{+\prime} B_t \epsilon_{t,g_4}^{+} + \epsilon_{t,g_3}^{+\prime} b_t). \end{aligned}$$

Then

$$\begin{aligned} V^A &= \epsilon'_{g_1} C \epsilon_{g_2} + \epsilon'_{g_1} c = \xi' C_* \xi + \xi' c_*, \\ V^B &= \epsilon'_{g_3} D \epsilon_{g_4} + \epsilon'_{g_3} d = \xi' D_* \xi + \xi' d_*, \end{aligned}$$

where $C = [C_{\varsigma\tau}]_{\varsigma,\tau=1,\cdots,T}$ *with* $C_{\varsigma\tau} = \sum_{t=1}^{T-1} \pi_{t\varsigma}\pi_{t\tau} A_t$, *where* $D = [D_{\varsigma\tau}]_{\varsigma,\tau=1,\cdots,T}$ *with* $D_{\varsigma\tau} = \sum_{t=1}^{T-1} \pi_{t\varsigma}\pi_{t\tau} B_t$, *where* $c = (c'_1, \cdots, c'_T)'$ *with* $c_\tau = \sum_{t=1}^{T-1} \pi_{t\tau} a_t$, *where* $d = (d'_1, \cdots, d'_T)'$ *with* $d_\tau = \sum_{t=1}^{T-1} \pi_{t\tau} b_t$, *and where* $C_* = p'_{g_1\cdot} p_{g_2\cdot} \otimes C$, $D_* = p'_{g_3\cdot} p_{g_4\cdot} \otimes D$, $c_* = p'_{g_1\cdot} \otimes c$, *and* $d_* = p'_{g_3\cdot} \otimes d$. *Furthermore* $\mathrm{E}(V^A) = \sigma_{g_1 g_2}\mathrm{tr}(A)$ *and*

$$\begin{aligned} \mathrm{Cov}(V^A, V^B) &= \sigma_{g_1g_4}\sigma_{g_2g_3}\mathrm{tr}(AB) + \sigma_{g_1g_3}\sigma_{g_2g_4}\mathrm{tr}(AB') + \sigma_{g_1g_3} a'b \\ &\quad + \mu_3^{\xi}[p_{g_1\cdot}\mathrm{vec}_D(p'_{g_3\cdot}p_{g_4\cdot}) \otimes c'\mathrm{vec}_D(D) + p_{g_3\cdot}\mathrm{vec}_D(p'_{g_1\cdot}p_{g_2\cdot}) \otimes d'\mathrm{vec}_D(C)] \\ &\quad + (\mu_4^{\xi} - 3)\mathrm{vec}_D(p'_{g_1\cdot}p_{g_2\cdot})'\mathrm{vec}_D(p'_{g_3\cdot}p_{g_4\cdot}) \otimes \mathrm{vec}_D(C)'\mathrm{vec}_D(D). \end{aligned}$$

Clearly if $\text{vec}_D(A_t) = \text{vec}_D(B_t) = 0$*, then* $\text{vec}_D(C) = \text{vec}_D(D) = 0$*, and in this case the terms involving the third and fourth moments of the elements of* ξ *are zero.*

Remark: Since the lemma does not restrict the elements of P, the setup allows for some or all of the ϵ_g to be the same.

Proof of Lemma A.3. W.o.l.o.g. we consider $g_1 = 1, g_2 = 2, g_3 = 3$ and $g_4 = 4$. Then $\epsilon_g = (p_{g\cdot} \otimes I_{nT})\xi$ and

$$\begin{aligned}
\sum_{t=1}^{T-1} \epsilon_{t,1}^{+\prime} A_t \epsilon_{t,2}^{+} &= \sum_{t=1}^{T-1}\sum_{\varsigma=1}^{T}\sum_{\tau=1}^{T} \pi_{t\varsigma}\pi_{t\tau}\epsilon_{\varsigma,1}' A_t \epsilon_{\tau,2} = \sum_{\varsigma=1}^{T}\sum_{\tau=1}^{T} \epsilon_{\varsigma,1}' C_{\varsigma\tau}\epsilon_{\tau,2} = \epsilon_1' C \epsilon_2 = \xi' C_* \xi, \\
\sum_{t=1}^{T-1} \epsilon_{t,1}^{+\prime} a_t &= \sum_{t=1}^{T-1}\sum_{\tau=1}^{T} \pi_{t\tau}\epsilon_{\tau,1}' a_t = \sum_{\tau=1}^{T} \epsilon_{\tau,1}' \sum_{t=1}^{T-1} \pi_{t\tau} a_t = \epsilon_1' c = \xi' c_*,
\end{aligned}$$

where $C_* = (p_{1\cdot}' \otimes I_{nT})C(p_{2\cdot} \otimes I_{nT}) = p_{1\cdot}'p_{2\cdot} \otimes C$ and $c_* = (p_{1\cdot}' \otimes I_{nT})c = p_{1\cdot}' \otimes c$. Similarly, $\sum_{t=1}^{T-1} \epsilon_{t,3}^{+\prime} B_t \epsilon_{t,4}^{+} = \xi' D_* \xi$, where $D_* = (p_{3\cdot}' \otimes I_{nT})D(p_{4\cdot} \otimes I_{nT}) = p_{3\cdot}'p_{4\cdot} \otimes D$, and $\sum_{t=1}^{T-1} \epsilon_{t,3}^{+\prime} b_t = \xi' d_*$ where $d_* = (p_{3\cdot}' \otimes I_{nT})d = p_{3\cdot}' \otimes d$.

Using Lemma A.1 in Kelejian and Prucha (2010), it follows that

$$\begin{aligned}
\text{E}(V^A) &= \text{E}(\xi' C_* \xi + \xi' c_*) = \text{tr}[C(p_{2\cdot}p_{1\cdot}' \otimes I_{nT})] = \sigma_{12}\text{tr}(C) = \sigma_{12}\sum_{\tau=1}^{T} \text{tr}(C_{\tau\tau}) \\
&= \sigma_{12}\sum_{\tau=1}^{T} \text{tr}(\sum_{t=1}^{T-1} \pi_{t\tau}^2 A_t) = \sigma_{12}\sum_{t=1}^{T-1} \text{tr}(A_t \sum_{\tau=1}^{T} \pi_{t\tau}^2) = \sigma_{12}\sum_{t=1}^{T-1} \text{tr}(A_t) = \sigma_{12}\text{tr}(A),
\end{aligned}$$

which proves the first claim, and

$$\begin{aligned}
\text{Cov}(V^A, V^B) &= \text{Cov}(\xi' C_* \xi + \xi' c_*, \xi' D_* \xi + \xi' d_*) = [\text{tr}(C_* D_*) + \text{tr}(C_* D_*')] + c_*' d_* \qquad \text{(A.7)} \\
&\quad + \mu_3^{\xi}[c_*' \text{vec}_D(D_*) + d_*' \text{vec}_D(C_*)] + (\mu_4^{\xi} - 3)\text{vec}_D(C_*)' \text{vec}_D(D_*).
\end{aligned}$$

Observe that

$$
\begin{aligned}
\mathrm{tr}(C_* D_*) &= \mathrm{tr}[(p'_{1\cdot} \otimes I_{nT}) C (p_{2\cdot} \otimes I_{nT})(p'_{3\cdot} \otimes I_{nT}) D (p_{4\cdot} \otimes I_{nT})] \qquad (A.8)\\
&= \sigma_{14}\sigma_{23}\mathrm{tr}(CD) = \sigma_{14}\sigma_{23}\sum_{\varsigma=1}^{T}\sum_{\tau=1}^{T}\mathrm{tr}(C_{\varsigma\tau}D_{\tau\varsigma})\\
&= \sigma_{14}\sigma_{23}\sum_{\varsigma=1}^{T}\sum_{\tau=1}^{T}\mathrm{tr}(\sum_{t=1}^{T-1}\sum_{s=1}^{T-1}\pi_{t\varsigma}\pi_{t\tau}\pi_{s\varsigma}\pi_{s\tau}A_t B_s)\\
&= \sigma_{14}\sigma_{23}\sum_{t=1}^{T-1}\sum_{s=1}^{T-1}\mathrm{tr}(A_t B_s)\sum_{\varsigma=1}^{T}\pi_{t\varsigma}\pi_{s\varsigma}\sum_{\tau=1}^{T}\pi_{t\tau}\pi_{s\tau}\\
&= \sigma_{14}\sigma_{23}\sum_{t=1}^{T-1}\mathrm{tr}(A_t B_t) = \sigma_{14}\sigma_{23}\mathrm{tr}(AB)
\end{aligned}
$$

in light of the orthonormality of the weights of the Helmert transformation, and analogously $\mathrm{tr}(C_* D'_*) = \sigma_{13}\sigma_{24}\mathrm{tr}(AB')$. Next, observe that

$$
\begin{aligned}
c'_* d_* &= c'(p_{1\cdot} p'_{3\cdot} \otimes I_{nT}) d = \sigma_{13}\sum_{\tau=1}^{T} c'_\tau d_\tau = \sigma_{13}\sum_{\tau=1}^{T}\sum_{t=1}^{T-1}\sum_{s=1}^{T-1}\pi_{t\tau}\pi_{s\tau}a'_t b_s \qquad (A.9)\\
&= \sigma_{13}\sum_{t=1}^{T-1}\sum_{s=1}^{T-1}a'_t b_s\sum_{\tau=1}^{T}\pi_{t\tau}\pi_{s\tau} = \sigma_{13}\sum_{t=1}^{T-1}a'_t b_t = \sigma_{13}a'b
\end{aligned}
$$

in light of the orthonormality of the weights of the Helmert transformation. Additionally observe that $\mathrm{vec}_D(C_*) = \mathrm{vec}_D(p'_{1\cdot} p_{2\cdot} \otimes C) = \mathrm{vec}_D(p'_{1\cdot} p_{2\cdot}) \otimes \mathrm{vec}_D(C)$ and similarly $\mathrm{vec}_D(D_*) = \mathrm{vec}_D(p'_{3\cdot} p_{4\cdot}) \otimes \mathrm{vec}_D(D)$, and thus

$$
\begin{aligned}
c'_*\mathrm{vec}_D(D_*) &= p_{1\cdot}\mathrm{vec}_D(p'_{3\cdot} p_{4\cdot}) \otimes c'\mathrm{vec}_D(D), \qquad (A.10)\\
d'_*\mathrm{vec}_D(C_*) &= p_{3\cdot}\mathrm{vec}_D(p'_{1\cdot} p_{2\cdot}) \otimes d'\mathrm{vec}_D(C),\\
\mathrm{vec}_D(C_*)'\mathrm{vec}_D(D_*) &= \mathrm{vec}_D(p'_{1\cdot} p_{2\cdot})'\mathrm{vec}_D(p'_{3\cdot} p_{4\cdot}) \otimes \mathrm{vec}_D(C)'\mathrm{vec}_D(D).
\end{aligned}
$$

Substitution of expressions (A.8)-(A.10) into (A.7) completes the proof of the second claim.

□

Lemma A.4. *Consider the first-stage regression given by (12). Suppose* $\text{plim}_{n\to\infty} n^{-1} H^{+\prime} \underline{X}^{+} \Gamma$ *and* $\lim_{n\to\infty} n^{-1} \bar{H}_r^{+\prime} \underline{X}^{+} \Gamma$*, for* $r = 1, \cdots, q$*, are finite with full column rank. Suppose the elements* ϵ_{it} *of* ϵ_t *and* $e_{it,g}$ *of* E_t *are generated as*

$$\epsilon_{it} = \sum_{l=1}^{G} p_{0l} \xi_{it,l} \quad \textit{and} \quad e_{it,g} = \sum_{l=1}^{G} p_{gl} \xi_{it,l} \tag{A.11}$$

where the basic innovations $\xi_{it,l}$ *are defined in Assumption A.2. Let* $A = \text{diag}_{t=1}^{T-1}\{A_t\}$*, where* A_t *is non-stochastic and the row and column sums of* A_t *are uniformly bounded in absolute value. Then, (i)* $\text{plim}_{n\to\infty} n^{-1} H^{+\prime} Z^{+}$ *and* $\text{plim}_{n\to\infty} n^{-1} \bar{H}_r^{+\prime} Z^{+}$*, for* $r = 1, \cdots, q$*, are finite with full column rank, (ii)* $n^{-1} Z^{+\prime} A Z^{+} = O_p(1)$ *and (iii)* $n^{-1} Z^{+\prime} A \epsilon^{+} = n^{-1} \text{E}(Z^{+\prime} A \epsilon^{+}) + o_p(1)$*, where* $n^{-1} \text{E}(Z^{+\prime} A \epsilon^{+}) = O(1)$ *and* $\text{E}(Z^{+\prime} A \epsilon^{+}) = 0$ *when* $\text{tr}(A) = 0$*.*

Proof. Let $p_{j.} = [p_{j1}, \ldots, p_{jG}]$, $j = 0, 1, \ldots$, then in light of Assumption A.2 and (A.11) we have

$$\begin{aligned} \sigma_\varepsilon^2 &= \text{Var}(\epsilon_{it}) = p_{0.} p_{0.}', \\ \sigma_g^2 &= \text{Var}(e_{it,g}) = p_{g.} p_{g.}', \quad g = 1, 2, \ldots \end{aligned}$$

From Lemma A.3 and its proof it is readily seen that

$$\begin{aligned} \epsilon &= [\epsilon_1', \ldots, \epsilon_T']' = (p_{0.} \otimes I_{nT}) \xi, \\ e_g &= [e_{1,g}', \ldots, e_{T,g}']' = (p_{g.} \otimes I_{nT}) \xi, \quad g = 1, 2, \ldots \end{aligned}$$

From (12) it follows that the columns of Z_t^{+} can then be written as

$$z_{t,j}^{+} = \underline{X}_t^{+} \gamma_j + e_{t,j}^{+}. \tag{A.12}$$

Hence, (i) follows directly from the assumption that $\text{plim}_{n\to\infty} n^{-1} H^{+\prime} \underline{X}^{+}\Gamma$ and $\lim_{n\to\infty} n^{-1}\bar{H}_r^{+\prime}\underline{X}^{+}\Gamma$, for $r = 1, \cdots, q$, are finite with full column rank.

To prove (ii), it suffices to show that $n^{-1} z_{t,j}^{+\prime} A_t z_{t,l}^{+\prime} = O_p(1)$. From (A.12) we see that $\text{Var}(z_{it,j}^{+}) = \text{Var}(e_{it,j}^{+}) = \sigma_j^2$. The claim now follows immediately from arguments analogous to those of Remark A.1 in Kelejian and Prucha (2004).

To prove (iii), it suffices to prove the claims for an arbitrary column. Focusing on the first column we have

$$n^{-1}\sum_{t=1}^{T-1} z_{t,1}^{+\prime} A_t \epsilon_t^{+} = n^{-1}\sum_{t=1}^{T-1} a_t^{\prime}\epsilon_t^{+} + n^{-1}\sum_{t=1}^{T-1} e_{t,1}^{+} A_t \epsilon_t^{+}.$$

with $a_t^{\prime} = \gamma_1^{\prime}\underline{X}_t^{+\prime} A_t$. Observing that $\sum_{t=1}^{T-1} z_{t,1}^{+\prime} A_t \epsilon_t^{+}$ is an instance of the linear quadratic forms considered by Lemma A.3, it follows immediately from that lemma that

$$\text{E}(n^{-1}\sum_{t=1}^{T-1} z_{t,1}^{+\prime} A_t \epsilon_t^{+}) = \sigma_{01}\sum_{t=1}^{T-1} n^{-1}\text{tr}(A_t).$$

Since the elements of A_t are uniformly bounded, we have $\text{E}(n^{-1}\sum_{t=1}^{T-1} z_{t,1}^{+\prime} A_t \epsilon_t^{+}) = O(1)$ and $\text{E}(n^{-1}\sum_{t=1}^{T-1} z_{t,1}^{+\prime} A_t \epsilon_t^{+}) = 0$ if $\text{tr}(A_t) = 0$. This proves the claims regarding $n^{-1}Z^{+\prime}A\epsilon^{+}$. $\square$

B Proofs of Theorems

Proof of Theorem 1. Under H_0^u and Assumptions 1 and 4,

$$
\begin{aligned}
n^{1/2}(\widehat{\delta}-\delta_0) &= (n^{-1}\widehat{Z}^{+\prime}\widehat{Z}^{+})^{-1}n^{-1/2}\widehat{Z}^{+\prime}\epsilon^{+} \\
&= (Q_{HZ}^{\prime}Q_{HH}^{-1}Q_{HZ})^{-1}Q_{HZ}^{\prime}Q_{HH}^{-1}(n^{-1/2}H^{+\prime}\epsilon^{+}) + o_p(1) = O_p(1),
\end{aligned}
\tag{B.1}
$$

where $Q_{HZ} = \text{plim}_{n\to\infty} n^{-1}H^{+\prime}Z^{+}$ and $Q_{HH} = \lim_{n\to\infty} n^{-1}H^{+\prime}H^{+}$. As $\widehat{u}^{+} = \epsilon^{+} - Z^{+}(\widehat{\delta}-\delta_0)$, we have

$$
\begin{aligned}
n^{-1/2}\widehat{u}^{+\prime}W_r^{*}\widehat{u}^{+} &= n^{-1/2}\epsilon^{+\prime}\mathring{W}_r^{*}\epsilon^{+} - 2(n^{-1}Z^{+\prime}\mathring{W}_r^{*}\epsilon^{+})^{\prime}n^{1/2}(\widehat{\delta}-\delta_0) \\
&\quad +(\widehat{\delta}-\delta_0)^{\prime}(n^{-1}Z^{+\prime}\mathring{W}_r^{*}Z^{+})n^{1/2}(\widehat{\delta}-\delta_0).
\end{aligned}
\tag{B.2}
$$

Therefore, in light of (B.1), (B.2), and Assumption 3, for the rth element of $\widehat{V}_Q$, $\widehat{V}_{r,Q} \equiv \widehat{u}^{+\prime}W_r^{*}\widehat{u}^{+}$, we have

$$
n^{-1/2}\widehat{V}_{r,Q} = n^{-1/2}V_{r,Q} + o_p(1), \tag{B.3}
$$

where

$$
V_{r,Q} = \epsilon^{+\prime}\mathring{W}_r^{*}\epsilon^{+} + a_r^{\prime}\epsilon^{+} = \epsilon^{\prime}C_r\epsilon + c_r^{\prime}\epsilon,
$$

with

$$
a_r = -2H^{+}Q_{HH}^{-1}Q_{HZ}(Q_{HZ}^{\prime}Q_{HH}^{-1}Q_{HZ})^{-1}[n^{-1}\text{E}(Z^{+\prime}\mathring{W}_r^{*}\epsilon^{+})],
$$

$C_r = (\Pi^{\prime}\otimes I_n)\mathring{W}_r^{*}(\Pi\otimes I_n)$, and $c_r = (\Pi^{\prime}\otimes I_n)a_r$. Clearly, the row and column sums of $\Pi\otimes I_n$ are uniformly bounded in absolute value. By Assumption 2, the row and column sums of $\mathring{W}_r^{*}$, and thus those of C_r, are uniformly bounded in absolute value. By Assumptions 3 and 4, the elements of H and $n^{-1}\text{E}(Z^{+\prime}\mathring{W}_r^{*}\epsilon^{+})$ are uniformly bounded in absolute value. This in turn implies that the elements of a_r and c_r are uniformly bounded in absolute value.

Together with Assumption 1 for the elements of ϵ, this verifies that the linear quadratic forms $V_{r,Q}$ satisfy the conditions A.1-A.3 postulated by the CLT given as Theorem A.1 in Kelejian and Prucha (2010).

Let $V_Q = (V_{1,Q}, \cdots, V_{q,Q})'$. Then by Lemma A.2 we have $\mathrm{E}(V_Q) = 0$ and the (r,s)th element of it's VC matrix $\Psi_Q = \mathrm{E}(V_Q V_Q')$ is given by

$$\psi_{rs,Q} = \mathrm{E}(V_{r,Q} V_{s,Q}) = \phi_{rs,Q} + \sigma_0^2 a_r' a_s, \tag{B.4}$$

where $\phi_{rs,Q} = 2\sigma_0^4 \mathrm{tr}(\mathring{W}_r^* \mathring{W}_s^*)$. In light of the above discussion, $n^{-1}\phi_{rs,Q} = O(1)$ and $n^{-1}\psi_{rs,Q} = O(1)$. Since by assumption the smallest eigenvalues of $n^{-1}\Psi_Q$ are bounded away from zero, it follows from Theorem A.1 in Kelejian and Prucha (2010) that

$$\Psi_Q^{-1/2} V_Q \xrightarrow{d} N(0, I_q). \tag{B.5}$$

The (r,s)th element of $\widehat{\Psi}_Q$ is given by

$$\widehat{\psi}_{rs,Q} = \widehat{\phi}_{rs,Q} + \widehat{\sigma}^2 \widehat{a}_r' \widehat{a}_s, \tag{B.6}$$

where $\widehat{\phi}_{rs,Q} = 2\widehat{\sigma}^4 \mathrm{tr}(\mathring{W}_r^* \mathring{W}_s^*)$ and $\widehat{a}_r = -2\widehat{Z}^+(\widehat{Z}^{+\prime}\widehat{Z}^+)^{-1}(Z^+ - \widehat{Z}^+)'\mathring{W}_s^* \widehat{u}^+$. By Assumption 3,

$$\begin{aligned}(T-1)\widehat{\sigma}^2 &= n^{-1}\widehat{u}^{+\prime}\widehat{u}^+ = n^{-1}\epsilon^{+\prime}\epsilon^+ - 2(n^{-1}Z^{+\prime}\epsilon^+)'(\widehat{\delta} - \delta_0) + (\widehat{\delta} - \delta_0)'(n^{-1}Z^{+\prime}Z^+)(\widehat{\delta} - \delta_0) \\ &= n^{-1}\epsilon^{+\prime}\epsilon^+ + o_p(1).\end{aligned}$$

In light of the remark after Lemma A.1 and Tschebychev's inequality, we have $[n(T -$

$1)]^{-1}\epsilon^{+\prime}\epsilon^{+} = \sigma_0^2 + o_p(1)$ and $\widehat{\sigma}^2 = \sigma_0^2 + o_p(1)$. Observing that

$$n^{-1}Z^{+\prime}\mathring{W}_r^{*}\widehat{u}^{+} = n^{-1}Z^{+\prime}\mathring{W}_r^{*}\epsilon^{+} - n^{-1}Z^{+\prime}\mathring{W}_r^{*}Z^{+\prime}(\widehat{\delta}-\delta_0) = n^{-1}\mathrm{E}(Z^{+\prime}\mathring{W}_r^{*}\epsilon^{+}) + o_p(1),$$

and

$$n^{-1}\widehat{Z}^{+\prime}\mathring{W}_r^{*}\widehat{u}^{+} = n^{-1}Z^{+\prime}P_{H^+}\mathring{W}_r^{*}\epsilon^{+} - n^{-1}Z^{+\prime}P_{H^+}\mathring{W}_r^{*}Z^{+\prime}(\widehat{\delta}-\delta_0) = o_p(1),$$

where $P_{H^+} = H^{+}(H^{+\prime}H^{+})^{-1}H^{+\prime}$, it is now readily seen from (B.4) and (B.6) and the above results that $n^{-1}\widehat{\psi}_{rs,Q} - n^{-1}\psi_{rs,Q} = o_p(1)$. Furthermore, since $n^{-1}\Psi_Q = O(1)$ and the smallest eigenvalue of $n^{-1}\Psi_Q$ is bounded away from zero by assumption, it follows from Lemma F.1 in Poetscher and Prucha (1997) that $n^{1/2}\widehat{\Psi}_Q^{-1/2} = n^{1/2}\Psi_Q^{-1/2} + o_p(1)$. Consequently, in light of (B.3) and (B.5),

$$\widehat{\Psi}_Q^{-1/2}\widehat{V}_Q = \Psi_Q^{-1/2}\widehat{V}_Q + o_p(1) = \Psi_Q^{-1/2}V_Q + o_p(1) \xrightarrow{d} N(0, I_q).$$

The claim now follows from the continuous mapping theorem. □

Proof of Theorem 2. Under H_0^y and Assumptions 1 and 4',

$$\begin{aligned} n^{1/2}(\widehat{\delta}-\delta_0) &= (n^{-1}\widehat{Z}^{+\prime}\widehat{Z}^{+})^{-1}n^{-1/2}\widehat{Z}^{+\prime}\epsilon^{+} \qquad \text{(B.7)} \\ &= (Q'_{HZ}Q_{HH}^{-1}Q_{HZ})^{-1}Q'_{HZ}Q_{HH}^{-1}(n^{-1/2}H^{+\prime}\epsilon^{+}) + o_p(1) = O_p(1), \end{aligned}$$

where $Q_{HZ} = \mathrm{plim}_{n\to\infty} n^{-1}H^{+\prime}Z^{+}$ and $Q_{HH} = \lim_{n\to\infty} n^{-1}H^{+\prime}H^{+}$. As $\widehat{u}^{+} = \epsilon^{+} - Z^{+}(\widehat{\delta}-$

δ_0), we have

$$n^{-1/2}\bar{H}_r^{+\prime}\widehat{u}^{+} = n^{-1/2}\bar{H}_r^{+\prime}\epsilon^{+} - n^{-1/2}\bar{H}_r^{+\prime}Z^{+}(\widehat{\delta}-\delta_0), \tag{B.8}$$
$$n^{-1/2}\widehat{u}^{+\prime}W_r^{*}\widehat{u}^{+} = n^{-1/2}\epsilon^{+\prime}\mathring{W}_r^{*}\epsilon^{+} - 2(n^{-1}Z^{+\prime}\mathring{W}_r^{*}\epsilon^{+})^{\prime}n^{1/2}(\widehat{\delta}-\delta_0) + (\widehat{\delta}-\delta_0)^{\prime}(n^{-1}Z^{+\prime}\mathring{W}_r^{*}Z^{+})n^{1/2}(\widehat{\delta}-\delta_0).$$

Observing that $\mathrm{tr}(\mathring{W}_r^{*}) = 0$, it follows by Assumption 3' that $n^{-1}Z^{+\prime}\mathring{W}_r^{*}\epsilon^{+} = o_p(1)$. Let $Q_{\bar{H}_r Z} = \mathrm{plim}_{n\to\infty} n^{-1}\bar{H}_r^{+\prime}Z^{+}$. Then, in light of (B.7), (B.8), and Assumptions 3' and 4', we have

$$n^{-1/2}\widehat{V}_{r,L} \equiv n^{-1/2}\bar{H}_r^{+\prime}\widehat{u}^{+} = n^{-1/2}V_{r,L} + o_p(1), \tag{B.9}$$
$$n^{-1/2}\widehat{V}_{r,Q} \equiv n^{-1/2}\widehat{u}^{+\prime}W_r^{*}\widehat{u}^{+} = n^{-1/2}V_{r,Q} + o_p(1),$$

where

$$V_{r,L} = a_r^{\prime}\epsilon^{+} = c_r^{\prime}\epsilon, \tag{B.10}$$
$$V_{r,Q} = \epsilon^{+\prime}\mathring{W}_r^{*}\epsilon^{+} = \epsilon^{\prime}C_r\epsilon,$$

with

$$a_r = \bar{H}_r^{+} - H^{+}Q_{HH}^{-1}Q_{HZ}(Q_{HZ}^{\prime}Q_{HH}^{-1}Q_{HZ})^{-1}Q_{\bar{H}_r Z}^{\prime},$$

$c_r = (\Pi^{\prime}\otimes I_n)a_r$, and $C_r = (\Pi^{\prime}\otimes I_n)W_r^{*}(\Pi\otimes I_n)$. Clearly, the row and column sums of $\Pi\otimes I_n$ are uniformly bounded in absolute value. By Assumption 2, the row and column sums of $\mathring{W}_{r,n}^{*}$, and thus those of C_r, are uniformly bounded in absolute value. By Assumptions 2 and 4', the elements of H_n and $\bar{H}_{r,n}^{+}$ are uniformly bounded in absolute value. This in turn implies that the elements of a_r and c_r are uniformly bounded in absolute value. Together with Assumption 1 for the elements of ϵ, this verifies that the linear quadratic forms $V_{r,L}$ and $V_{r,Q}$ satisfy the conditions A.1-A.3 postulated by the CLT given as Theorem A.1 in Kelejian and Prucha (2010).

Let $V = [V_L', V_Q']'$ with $V_L = [V_{1,L}', \cdots, V_{q,L}']'$ and $V_Q = [V_{1,Q}', \cdots, V_{q,Q}']'$. Then by Lemma A.2 we have $\mathrm{E}(V) = 0$, and it's VC matrix $\Phi = \mathrm{E}(VV')$ is given by

$$\Phi = \begin{bmatrix} \Phi_L & 0 \\ 0 & \Phi_Q \end{bmatrix} \tag{B.11}$$

with the (r, s)th submatrix of Φ_L given by

$$\phi_{rs,L} = \mathrm{Cov}(V_{r,L}, V_{s,L}) = \sigma_0^2 a_r' a_s,$$

and the (r, s)th element of Φ_Q given by

$$\phi_{rs,Q} = \mathrm{Cov}(V_{r,Q}, V_{s,Q}) = 2\sigma_0^4 \mathrm{tr}(\mathring{W}_r^* \mathring{W}_s^*).$$

In light of the above discussion, $n^{-1}\phi_{rs,L} = O(1)$ and $n^{-1}\phi_{rs,Q} = O(1)$. Since by assumption the smallest eigenvalues of $n^{-1}\Phi$ are bounded away from zero, it follows from Theorem A.1 in Kelejian and Prucha (2010) that

$$\Phi^{-1/2} V \xrightarrow{d} N(0, I_q). \tag{B.12}$$

Let $\widehat{a}_r = M_{\widehat{Z}Z} \bar{H}_r^+$, where $M_{\widehat{Z}Z} = I_{n(T-1)} - \widehat{Z}^+(\widehat{Z}^{+\prime}\widehat{Z}^+)^{-1}Z^{+\prime}$. Then the (r, s)th submatrix of $\widehat{\Phi}_L$ is given by $\widehat{\phi}_{rs,L} = \widehat{\sigma}^2 \widehat{a}_r' \widehat{a}_s$, and the (r, s)th element of $\widehat{\Phi}_Q$ is given by $\widehat{\phi}_{rs,Q} = 2\widehat{\sigma}^4 \mathrm{tr}(\mathring{W}_r^* \mathring{W}_s^*)$. By analogous arguments as in the proof of Theorem 1 it is readily seen that $n^{-1}\widehat{\phi}_{rs,L} - n^{-1}\phi_{rs,L} = o_p(1)$ and $n^{-1}\widehat{\phi}_{rs,Q} - n^{-1}\phi_{rs,Q} = o_p(1)$. Furthermore, since $n^{-1}\Phi = O(1)$ and the smallest eigenvalue of $n^{-1}\Phi$ is bounded away from zero by assumption, it follows from Lemma F.1 in Pötscher and Prucha (1997) that $n^{1/2}\widehat{\Phi}^{-1/2} = n^{1/2}\Phi^{-1/2} + o_p(1)$. Consequently, in light of (B.9) and (B.12),

$$\widehat{\Phi}^{-1/2}\widehat{V} = \Phi^{-1/2}\widehat{V} + o_p(1) = \Phi^{-1/2}V + o_p(1) \xrightarrow{d} N(0, I_{(K_H+1)q}).$$

The claim now follows from the continuous mapping theorem. $\square$

Table 1. Rejection Rates for $I_u^2(2)$ Tests with W_1 and W_2 (Small T)

ρ_1	ρ_2	Helmert weighting	Time-average weighting	Initial-period weighting
		n = 250, T = 5		
0	0	0.0507	0.0508	0.0493
.2	0	0.9379	0.6577	0.2385
.4	0	1.0000	0.9997	0.8232
0	.2	0.9391	0.6698	0.2560
0	.4	1.0000	0.9999	0.8510
.2	.2	0.9998	0.9815	0.7358
.4	.4	1.0000	1.0000	1.0000
		n = 500, T = 5		
0	0	0.0486	0.0507	0.0506
.2	0	0.9989	0.9114	0.4284
.4	0	1.0000	1.0000	0.9818
0	.2	0.9982	0.9121	0.4170
0	.4	1.0000	1.0000	0.9815
.2	.2	1.0000	0.9999	0.9461
.4	.4	1.0000	1.0000	1.0000

Nominal size is 0.05

Table 2. Rejection Rates for $I_u^2(2)$ Tests with W_1 and W_2 (Large T)

ρ_1	ρ_2	Helmert weighting	Time-average weighting	Initial-period weighting
		n = 250, T = 10		
0	0	0.0493	0.0493	0.0497
.2	0	0.9999	0.8042	0.2574
.4	0	1.0000	1.0000	0.8634
0	.2	0.9999	0.7995	0.2590
0	.4	1.0000	1.0000	0.8873
.2	.2	1.0000	0.9991	0.8509
.4	.4	1.0000	1.0000	1.0000
		n = 250, T = 20		
0	0	0.0500	0.0517	0.0513
.2	0	1.0000	0.9248	0.3332
.4	0	1.0000	1.0000	0.9546
0	.2	1.0000	0.9283	0.3422
0	.4	1.0000	1.0000	0.9589
.2	.2	1.0000	1.0000	0.9710
.4	.4	1.0000	1.0000	1.0000

Nominal size is 0.05

Table 3. Rejection Rates for $I_y^2(2)$ Tests with W_1 and W_2 (Small T)

λ_1	λ_2	ρ_1	ρ_2	γ_1	γ_2	Helmert weighting	Time-average weighting	Initial-period weighting
							$n = 250, T = 5$	
0	0	0	0	0	0	0.0506	0.0501	0.0497
.1	0	0	0	0	0	0.7405	0.6580	0.5967
.2	0	0	0	0	0	1.0000	0.9998	0.9995
0	.1	0	0	0	0	0.7302	0.6469	0.5827
0	.2	0	0	0	0	1.0000	0.9997	0.9994
.1	.1	0	0	0	0	0.9910	0.9803	0.9638
.2	.2	0	0	0	0	1.0000	1.0000	1.0000
0	0	.2	0	0	0	0.8521	0.5141	0.1820
0	0	.4	0	0	0	1.0000	0.9984	0.7233
0	0	0	.2	0	0	0.8565	0.5246	0.1946
0	0	0	.4	0	0	1.0000	0.9992	0.7654
0	0	.2	.2	0	0	0.9984	0.9521	0.6191
0	0	.4	.4	0	0	1.0000	1.0000	1.0000
.1	.1	.2	.2	0	0	1.0000	1.0000	0.9980
.2	.2	.4	.4	0	0	1.0000	1.0000	1.0000
0	0	0	0	.1	0	0.4898	0.4900	0.4880
0	0	0	0	.2	0	0.9931	0.9934	0.9933
0	0	0	0	0	.1	0.4748	0.4761	0.4748
0	0	0	0	0	.2	0.9917	0.9913	0.9916
0	0	0	0	.1	.1	0.8606	0.8616	0.8611
0	0	0	0	.2	.2	1.0000	1.0000	1.0000
							$n = 500, T = 5$	
0	0	0	0	0	0	0.0503	0.0508	0.0509
.1	0	0	0	0	0	0.9659	0.9295	0.8934
.2	0	0	0	0	0	1.0000	1.0000	1.0000
0	.1	0	0	0	0	0.9712	0.9431	0.9123
0	.2	0	0	0	0	1.0000	1.0000	1.0000
.1	.1	0	0	0	0	1.0000	1.0000	0.9997
.2	.2	0	0	0	0	1.0000	1.0000	1.0000
0	0	.2	0	0	0	0.9937	0.8155	0.3125
0	0	.4	0	0	0	1.0000	1.0000	0.9532
0	0	0	.2	0	0	0.9898	0.8164	0.3208
0	0	0	.4	0	0	1.0000	1.0000	0.9627
0	0	.2	.2	0	0	1.0000	0.9993	0.8934
0	0	.4	.4	0	0	1.0000	1.0000	1.0000
.1	.1	.2	.2	0	0	1.0000	1.0000	1.0000
.2	.2	.4	.4	0	0	1.0000	1.0000	1.0000
0	0	0	0	.1	0	0.8335	0.8331	0.8346
0	0	0	0	.2	0	1.0000	1.0000	1.0000
0	0	0	0	0	.1	0.8639	0.8627	0.8619
0	0	0	0	0	.2	1.0000	1.0000	1.0000
0	0	0	0	.1	.1	0.9956	0.9957	0.9957
0	0	0	0	.2	.2	1.0000	1.0000	1.0000

Nominal size is 0.05

Table 4. Rejection Rates for $I_y^2(2)$ Tests with W_1 and W_2 (Large T)

λ_1	λ_2	ρ_1	ρ_2	γ_1	γ_2	Helmert weighting	Time-average weighting	Initial-period weighting
						$n = 250, T = 10$		
0	0	0	0	0	0	0.0488	0.0498	0.0498
.1	0	0	0	0	0	0.9829	0.9349	0.9078
.2	0	0	0	0	0	1.0000	1.0000	1.0000
0	.1	0	0	0	0	0.9890	0.9581	0.9407
0	.2	0	0	0	0	1.0000	1.0000	1.0000
.1	.1	0	0	0	0	1.0000	1.0000	0.9998
.2	.2	0	0	0	0	1.0000	1.0000	1.0000
0	0	.2	0	0	0	0.9990	0.6746	0.1930
0	0	.4	0	0	0	1.0000	1.0000	0.7763
0	0	0	.2	0	0	0.9989	0.6716	0.1933
0	0	0	.4	0	0	1.0000	1.0000	0.8031
0	0	.2	.2	0	0	1.0000	0.9968	0.7534
0	0	.4	.4	0	0	1.0000	1.0000	1.0000
.1	.1	.2	.2	0	0	1.0000	1.0000	1.0000
.2	.2	.4	.4	0	0	1.0000	1.0000	1.0000
0	0	0	0	.1	0	0.8672	0.8674	0.8675
0	0	0	0	.2	0	1.0000	1.0000	1.0000
0	0	0	0	0	.1	0.8989	0.9000	0.8979
0	0	0	0	0	.2	1.0000	1.0000	1.0000
0	0	0	0	.1	.1	0.9978	0.9981	0.9981
0	0	0	0	.2	.2	1.0000	1.0000	1.0000
						$n = 250, T = 20$		
0	0	0	0	0	0	0.0501	0.0506	0.0522
.1	0	0	0	0	0	1.0000	1.0000	0.9999
.2	0	0	0	0	0	1.0000	1.0000	1.0000
0	.1	0	0	0	0	1.0000	1.0000	0.9999
0	.2	0	0	0	0	1.0000	1.0000	1.0000
.1	.1	0	0	0	0	1.0000	1.0000	1.0000
.2	.2	0	0	0	0	1.0000	1.0000	1.0000
0	0	.2	0	0	0	1.0000	0.8429	0.2427
0	0	.4	0	0	0	1.0000	1.0000	0.9013
0	0	0	.2	0	0	1.0000	0.8482	0.2533
0	0	0	.4	0	0	1.0000	1.0000	0.9156
0	0	.2	.2	0	0	1.0000	1.0000	0.9300
0	0	.4	.4	0	0	1.0000	1.0000	1.0000
.1	.1	.2	.2	0	0	1.0000	1.0000	1.0000
.2	.2	.4	.4	0	0	1.0000	1.0000	1.0000
0	0	0	0	.1	0	0.9995	0.9996	0.9995
0	0	0	0	.2	0	1.0000	1.0000	1.0000
0	0	0	0	0	.1	0.9992	0.9991	0.9991
0	0	0	0	0	.2	1.0000	1.0000	1.0000
0	0	0	0	.1	.1	1.0000	1.0000	1.0000
0	0	0	0	.2	.2	1.0000	1.0000	1.0000

Nominal size is 0.05

Table 5. Rejection Rates for $I_u^2(q)$ Tests with only Exogenous Covariates

ρ_1	ρ_2	$I_u^2(1)$ Test with W_1	$I_u^2(1)$ Test with W_2	$I_u^2(2)$ Test with W_1, W_2
			n = 250, T = 5	
0	0	0.0488	0.0502	0.0507
.2	0	0.9654	0.1036	0.9379
.4	0	1.0000	0.3328	1.0000
0	.2	0.0995	0.9664	0.9391
0	.4	0.2956	1.0000	1.0000
.2	.2	0.9940	0.9949	0.9998
.4	.4	1.0000	1.0000	1.0000
			n = 500, T = 5	
0	0	0.0487	0.0506	0.0486
.2	0	0.9996	0.1444	0.9989
.4	0	1.0000	0.5361	1.0000
0	.2	0.1344	0.9994	0.9982
0	.4	0.4861	1.0000	1.0000
.2	.2	1.0000	0.9999	1.0000
.4	.4	1.0000	1.0000	1.0000

Nominal size is 0.05. The DGP is Defined by Equation (28).

Table 6. Rejection Rates for $I_u^2(q)$ Tests with an Endogenous Covariate y_t^0

ρ_1	ρ_2	$I_u^2(1)$ Test with W_1	$I_u^2(1)$ Test with W_2	$I_u^2(2)$ Test with W_1, W_2
			n = 250, T = 5	
0	0	0.0521	0.0515	0.0527
.2	0	0.9659	0.1009	0.9397
.4	0	1.0000	0.3094	1.0000
0	.2	0.0975	0.9659	0.9395
0	.4	0.2871	1.0000	1.0000
.2	.2	0.9938	0.9944	0.9997
.4	.4	1.0000	1.0000	1.0000
			n = 500, T = 5	
0	0	0.0495	0.0502	0.0515
.2	0	0.9994	0.1461	0.9984
.4	0	1.0000	0.5573	1.0000
0	.2	0.1443	0.9996	0.9986
0	.4	0.5468	1.0000	1.0000
.2	.2	1.0000	1.0000	1.0000
.4	.4	1.0000	1.0000	1.0000

Nominal size is 0.05. The DGP is Defined by Equation (29).

Table 7. Rejection Rates for $I_u^2(q)$ Tests with Spatial Lags of y_t

ρ_1	ρ_2	$I_u^2(1)$ Test with W_1	$I_u^2(1)$ Test with W_2	$I_u^2(2)$ Test with W_1, W_2
			$\mathrm{n} = 250, \mathrm{T} = 5$	
0	0	0.0501	0.0499	0.0498
.2	0	0.8933	0.0972	0.8274
.4	0	1.0000	0.2948	1.0000
0	.2	0.0796	0.9635	0.9343
0	.4	0.1964	1.0000	1.0000
.2	.2	0.9673	0.9942	0.9990
.4	.4	1.0000	1.0000	1.0000
			$\mathrm{n} = 500, \mathrm{T} = 5$	
0	0	0.0493	0.0484	0.0497
.2	0	0.9947	0.1470	0.9873
.4	0	1.0000	0.5566	1.0000
0	.2	0.1145	0.9995	0.9985
0	.4	0.4097	1.0000	1.0000
.2	.2	0.9996	1.0000	1.0000
.4	.4	1.0000	1.0000	1.0000

Nominal size is 0.05. The DGP is Defined by Equation (30).

Table 8. Rejection Rates for $I_u^2(q)$ Tests with Spatial Lags of y_t and X_t

ρ_1	ρ_2	$I_u^2(1)$ Test with W_1	$I_u^2(1)$ Test with W_2	$I_u^2(2)$ Test with W_1, W_2
			$\mathrm{n} = 250, \mathrm{T} = 5$	
0	0	0.0495	0.0484	0.0487
.2	0	0.3965	0.0838	0.2522
.4	0	0.9592	0.2192	0.8594
0	.2	0.0494	0.9473	0.9245
0	.4	0.0529	1.0000	1.0000
.2	.2	0.4736	0.9880	0.9722
.4	.4	0.9169	0.9820	0.9951
			$\mathrm{n} = 500, \mathrm{T} = 5$	
0	0	0.0482	0.0483	0.0477
.2	0	0.8198	0.1194	0.7063
.4	0	1.0000	0.4299	0.9998
0	.2	0.0791	0.9993	0.9982
0	.4	0.2452	1.0000	1.0000
.2	.2	0.9427	1.0000	1.0000
.4	.4	1.0000	1.0000	1.0000

Nominal size is 0.05. The DGP is Defined by Equation (31).

Table 9. Rejection Rates for $I_y^2(q)$ Tests with only Exogenous Covariates under the Null

λ_1	λ_2	ρ_1	ρ_2	γ_1	γ_2	$I_y^2(1)$ Test w/ W_1	$I_y^2(1)$ Test w/ W_2	$I_y^2(2)$ Test w/ W_1, W_2
							$n = 250, T = 5$	
0	0	0	0	0	0	0.0488	0.0512	0.0506
.1	0	0	0	0	0	0.8383	0.0641	0.7405
.2	0	0	0	0	0	1.0000	0.1044	1.0000
0	.1	0	0	0	0	0.0614	0.8295	0.7302
0	.2	0	0	0	0	0.0988	1.0000	1.0000
.1	.1	0	0	0	0	0.9307	0.9323	0.9910
.2	.2	0	0	0	0	1.0000	1.0000	1.0000
0	0	.2	0	0	0	0.9148	0.0864	0.8521
0	0	.4	0	0	0	1.0000	0.2538	1.0000
0	0	0	.2	0	0	0.0814	0.9182	0.8565
0	0	0	.4	0	0	0.2185	1.0000	1.0000
0	0	.2	.2	0	0	0.9818	0.9832	0.9984
0	0	.4	.4	0	0	1.0000	1.0000	1.0000
.1	.1	.2	.2	0	0	1.0000	1.0000	1.0000
.2	.2	.4	.4	0	0	1.0000	1.0000	1.0000
0	0	0	0	.1	0	0.6160	0.0530	0.4898
0	0	0	0	.2	0	0.9981	0.0609	0.9931
0	0	0	0	0	.1	0.0513	0.5989	0.4748
0	0	0	0	0	.2	0.0582	0.9976	0.9917
0	0	0	0	.1	.1	0.6890	0.6736	0.8606
0	0	0	0	.2	.2	0.9995	0.9993	1.0000
							$n = 500, T = 5$	
0	0	0	0	0	0	0.0501	0.0486	0.0503
.1	0	0	0	0	0	0.9874	0.0634	0.9659
.2	0	0	0	0	0	1.0000	0.1109	1.0000
0	.1	0	0	0	0	0.0684	0.9898	0.9712
0	.2	0	0	0	0	0.1486	1.0000	1.0000
.1	.1	0	0	0	0	0.9975	0.9980	1.0000
.2	.2	0	0	0	0	1.0000	1.0000	1.0000
0	0	.2	0	0	0	0.9980	0.1075	0.9937
0	0	.4	0	0	0	1.0000	0.4119	1.0000
0	0	0	.2	0	0	0.1007	0.9966	0.9898
0	0	0	.4	0	0	0.3655	1.0000	1.0000
0	0	.2	.2	0	0	0.9999	0.9999	1.0000
0	0	.4	.4	0	0	1.0000	1.0000	1.0000
.1	.1	.2	.2	0	0	1.0000	1.0000	1.0000
.2	.2	.4	.4	0	0	1.0000	1.0000	1.0000
0	0	0	0	.1	0	0.9116	0.0489	0.8335
0	0	0	0	.2	0	1.0000	0.0506	1.0000
0	0	0	0	0	.1	0.0517	0.9322	0.8639
0	0	0	0	0	.2	0.0580	1.0000	1.0000
0	0	0	0	.1	.1	0.9313	0.9489	0.9956
0	0	0	0	.2	.2	1.0000	1.0000	1.0000

Nominal size is 0.05. The DGP is Defined by Equation (32).

Table 10. Rejection Rates for $I_y^2(q)$ Tests with an Endogenous Covariate y_t^0 under the Null

λ_1	λ_2	ρ_1	ρ_2	γ_1	γ_2	$I_y^2(1)$ Test w/ W_1	$I_y^2(1)$ Test w/ W_2	$I_y^2(2)$ Test w/ W_1, W_2
							$n = 250, T = 5$	
0	0	0	0	0	0	0.0515	0.0501	0.0507
.1	0	0	0	0	0	0.9490	0.0717	0.9008
.2	0	0	0	0	0	1.0000	0.1533	1.0000
0	.1	0	0	0	0	0.0721	0.9208	0.8574
0	.2	0	0	0	0	0.1456	1.0000	1.0000
.1	.1	0	0	0	0	0.9838	0.9729	0.9988
.2	.2	0	0	0	0	1.0000	1.0000	1.0000
0	0	.2	0	0	0	0.9170	0.0857	0.8574
0	0	.4	0	0	0	1.0000	0.2449	1.0000
0	0	0	.2	0	0	0.0826	0.9150	0.8548
0	0	0	.4	0	0	0.2104	1.0000	1.0000
0	0	.2	.2	0	0	0.9808	0.9821	0.9982
0	0	.4	.4	0	0	1.0000	1.0000	1.0000
.1	.1	.2	.2	0	0	1.0000	1.0000	1.0000
.2	.2	.4	.4	0	0	1.0000	1.0000	1.0000
0	0	0	0	.1	0	0.7306	0.0544	0.6146
0	0	0	0	.2	0	0.9998	0.0657	0.9988
0	0	0	0	0	.1	0.0572	0.6531	0.5338
0	0	0	0	0	.2	0.0670	0.9985	0.9952
0	0	0	0	.1	.1	0.7846	0.7099	0.9119
0	0	0	0	.2	.2	0.9999	0.9993	1.0000
							$n = 500, T = 5$	
0	0	0	0	0	0	0.0503	0.0512	0.0495
.1	0	0	0	0	0	0.9988	0.0803	0.9960
.2	0	0	0	0	0	1.0000	0.2114	1.0000
0	.1	0	0	0	0	0.0785	0.9989	0.9955
0	.2	0	0	0	0	0.1896	1.0000	1.0000
.1	.1	0	0	0	0	0.9999	0.9999	1.0000
.2	.2	0	0	0	0	1.0000	1.0000	1.0000
0	0	.2	0	0	0	0.9971	0.1086	0.9920
0	0	.4	0	0	0	1.0000	0.4264	1.0000
0	0	0	.2	0	0	0.1080	0.9974	0.9913
0	0	0	.4	0	0	0.4197	1.0000	1.0000
0	0	.2	.2	0	0	0.9999	1.0000	1.0000
0	0	.4	.4	0	0	1.0000	1.0000	1.0000
.1	.1	.2	.2	0	0	1.0000	1.0000	1.0000
.2	.2	.4	.4	0	0	1.0000	1.0000	1.0000
0	0	0	0	.1	0	0.9564	0.0554	0.9074
0	0	0	0	.2	0	1.0000	0.0687	1.0000
0	0	0	0	0	.1	0.0538	0.9538	0.9044
0	0	0	0	0	.2	0.0617	1.0000	1.0000
0	0	0	0	.1	.1	0.9625	0.9667	0.9987
0	0	0	0	.2	.2	1.0000	1.0000	1.0000

Nominal size is 0.05. The DGP is Defined by Equation (33).

Online Appendices to "On Testing for Spatial or Social Network Dependence in Panel Data Allowing for Network Variability"

by Xiaodong Liu and Ingmar R. Prucha

C Proofs of Propositions

Proof of Proposition 1. Let $G_t = -\mathrm{E}[\frac{\partial g_t(\theta)}{\partial \theta}|_{\theta_0}]$ and $\Omega_t = \mathrm{E}[g_t(\theta_0)g_t(\theta_0)']$, then under the maintained assumptions, in particular, $\mathrm{E}(\epsilon_t|\mu) = 0$ and $\mathrm{E}(\mu) = 0$,

$$G_t = \begin{bmatrix} 0 & H'(R_t X_t)^+ \\ G_{t,1} & 0 \\ \vdots & \vdots \\ G_{t,q} & 0 \end{bmatrix}, \qquad \Omega_t = \begin{bmatrix} \sigma_0^2 H'H & 0 \\ 0 & \Psi_Q \end{bmatrix}, \qquad \Psi_Q = \begin{bmatrix} \Omega_{11} & \cdots & \Omega_{1q} \\ \vdots & \ddots & \vdots \\ \Omega_{q1} & \cdots & \Omega_{qq} \end{bmatrix},$$

where $H = [X_1, \cdots, X_T]$, and

$$\begin{aligned} G_{t,r} &= 2\sigma_0^2 \begin{bmatrix} \sum_{\tau=1}^T \pi_{t\tau}^2 \mathrm{tr}(\mathring{W}_{1,r} W_{\tau,1} R_\tau^{-1}) & \cdots & \sum_{\tau=1}^T \pi_{t\tau}^2 \mathrm{tr}(\mathring{W}_{1,r} W_{\tau,q} R_\tau^{-1}) \\ \vdots & \ddots & \vdots \\ \sum_{\tau=1}^T \pi_{t\tau}^2 \mathrm{tr}(\mathring{W}_{T,r} W_{\tau,1} R_\tau^{-1}) & \cdots & \sum_{\tau=1}^T \pi_{t\tau}^2 \mathrm{tr}(\mathring{W}_{T,r} W_{\tau,q} R_\tau^{-1}) \end{bmatrix}, \\ \Omega_{rs} &= 2\sigma_0^4 \begin{bmatrix} \mathrm{tr}(\mathring{W}_{1,r} \mathring{W}_{1,s}) & \cdots & \mathrm{tr}(\mathring{W}_{1,r} \mathring{W}_{T,s}) \\ \vdots & \ddots & \vdots \\ \mathrm{tr}(\mathring{W}_{T,r} \mathring{W}_{1,s}) & \cdots & \mathrm{tr}(\mathring{W}_{T,r} \mathring{W}_{T,s}) \end{bmatrix}, \end{aligned}$$

for $r, s = 1, \cdots, q$. Evaluated at the restricted estimates $\widehat{\theta} = (0, \widehat{\beta}')'$ and $\widehat{\sigma}^2$, the LM test statistic is given by

$$\begin{aligned} \text{GMM-LM}_u &= \widehat{g}'\widehat{\Omega}^{-1}\widehat{G}(\widehat{G}'\widehat{\Omega}^{-1}\widehat{G})^{-1}\widehat{G}'\widehat{\Omega}^{-1}\widehat{g} \\ &= \sum_{t=1}^{T-1} \widehat{g}_t'\widehat{\Omega}_t^{-1}\widehat{G}_t \left[\sum_{t=1}^{T-1} \widehat{G}_t'\widehat{\Omega}_t^{-1}\widehat{G}_t\right]^{-1} \sum_{t=1}^{T-1} \widehat{G}_t'\widehat{\Omega}_t^{-1}\widehat{g}_t, \end{aligned}$$

with $\widehat{g}_t = [H, W_{1,1}\widehat{u}_t^+, \cdots, W_{T,1}\widehat{u}_t^+, \cdots, W_{1,q}\widehat{u}_t^+, \cdots, W_{T,q}\widehat{u}_t^+]'\widehat{u}_t^+$, and

$$\widehat{G}_t = \begin{bmatrix} 0 & H'X_t^+ \\ \widehat{G}_{t,1} & 0 \\ \vdots & \vdots \\ \widehat{G}_{t,q} & 0 \end{bmatrix}, \qquad \widehat{\Omega}_t = \begin{bmatrix} \widehat{\sigma}^2 H'H & 0 \\ 0 & \widehat{\Psi}_Q \end{bmatrix}, \qquad \widehat{\Psi}_Q = \begin{bmatrix} \widehat{\Omega}_{11} & \cdots & \widehat{\Omega}_{1q} \\ \vdots & \ddots & \vdots \\ \widehat{\Omega}_{q1} & \cdots & \widehat{\Omega}_{qq} \end{bmatrix},$$

where, recalling that $W_{t,r}^* = \sum_{\tau=1}^T \pi_{t\tau}^2 W_{\tau,r}$,

$$\begin{aligned} \widehat{G}_{t,r} &= 2\widehat{\sigma}^2 \begin{bmatrix} \text{tr}(\mathring{W}_{1,r}W_{t,1}^*) & \cdots & \text{tr}(\mathring{W}_{1,r}W_{t,q}^*) \\ \vdots & \ddots & \vdots \\ \text{tr}(\mathring{W}_{T,r}W_{t,1}^*) & \cdots & \text{tr}(\mathring{W}_{T,r}W_{t,q}^*) \end{bmatrix}, \\ \widehat{\Omega}_{rs} &= 2\widehat{\sigma}^4 \begin{bmatrix} \text{tr}(\mathring{W}_{1,r}\mathring{W}_{1,s}) & \cdots & \text{tr}(\mathring{W}_{1,r}\mathring{W}_{T,s}) \\ \vdots & \ddots & \vdots \\ \text{tr}(\mathring{W}_{T,r}\mathring{W}_{1,s}) & \cdots & \text{tr}(\mathring{W}_{T,r}\mathring{W}_{T,s}) \end{bmatrix}, \end{aligned}$$

for $r, s = 1, \cdots, q$. Let $\bar{\pi}_t = [\pi_{t1}^2, \cdots, \pi_{tT}^2]$, then

$$\widehat{\Omega}_{rs}\bar{\pi}_t' = 2\widehat{\sigma}^4 \begin{bmatrix} \text{tr}\left[\mathring{W}_{1,r}(\sum_{\tau=1}^T \mathring{W}_{\tau,s}\pi_{t\tau}^2)\right] \\ \vdots \\ \text{tr}\left[\mathring{W}_{T,r}(\sum_{\tau=1}^T \mathring{W}_{\tau,s}\pi_{t\tau}^2)\right] \end{bmatrix} = 2\widehat{\sigma}^4 \begin{bmatrix} \text{tr}(\mathring{W}_{1,r}W_{t,s}^*) \\ \vdots \\ \text{tr}(\mathring{W}_{T,r}W_{t,s}^*) \end{bmatrix}.$$

In light of this, $\widehat{G}_{t,r} = \widehat{\sigma}^{-2}[\widehat{\Omega}_{r1}, \cdots, \widehat{\Omega}_{rq}](I_q \otimes \bar{\pi}_t')$ for $r = 1, \cdots, q$, and

$$\begin{aligned}
\widehat{G}_t &= \begin{bmatrix} 0 & H'X_t^+ \\ \widehat{\sigma}^{-2}\widehat{\Psi}_Q(I_q \otimes \bar{\pi}_t') & 0 \end{bmatrix}, \\
\widehat{G}_t'\widehat{\Omega}_t^{-1} &= \begin{bmatrix} 0 & \widehat{\sigma}^{-2}(I_q \otimes \bar{\pi}_t) \\ \widehat{\sigma}^{-2}X_t^{+\prime}H(H'H)^{-1} & 0 \end{bmatrix}.
\end{aligned}$$

Let $\widehat{\Phi}_{Q,t} = (I_q \otimes \bar{\pi}_t)\widehat{\Psi}_Q(I_q \otimes \bar{\pi}_t')$. Observing that $H(H'H)^{-1}H'X_t^+ = X_t^+$ and $\sum_{t=1}^{T-1} X_t^{+\prime}\widehat{u}_t^+ = 0$, we have

$$\widehat{G}_t'\widehat{\Omega}_t^{-1}\widehat{G}_t = \begin{bmatrix} \widehat{\sigma}^{-4}\widehat{\Phi}_{Q,t} & 0 \\ 0 & \widehat{\sigma}^{-2}X_t^{+\prime}X_t^+ \end{bmatrix},$$

and

$$\sum_{t=1}^{T-1} \widehat{G}_t'\widehat{\Omega}_t^{-1}\widehat{g}_t = \widehat{\sigma}^{-2}\sum_{t=1}^{T-1} \begin{bmatrix} \widehat{u}_t^{+\prime}W_{t,1}^*\widehat{u}_t^+ \\ \vdots \\ \widehat{u}_t^{+\prime}W_{t,q}^*\widehat{u}_t^+ \\ X_t^{+\prime}H(H'H)^{-1}H'\widehat{u}_t^+ \end{bmatrix} = \widehat{\sigma}^{-2}\sum_{t=1}^{T-1} \begin{bmatrix} \widehat{u}_t^{+\prime}W_{t,1}^*\widehat{u}_t^+ \\ \vdots \\ \widehat{u}_t^{+\prime}W_{t,q}^*\widehat{u}_t^+ \\ 0 \end{bmatrix} = \widehat{\sigma}^{-2}\begin{bmatrix} \widehat{V}_Q \\ 0 \end{bmatrix}.$$

Observing that $\bar{\pi}_t \widehat{\Omega}_{rs} \bar{\pi}_t' = 2\widehat{\sigma}^4 \text{tr}\left[(\sum_{\tau=1}^{T} \pi_{t\tau}^2 \mathring{W}_{\tau,r}) W_{t,s}^*\right] = 2\widehat{\sigma}^4 \text{tr}(W_{t,r}^* W_{t,s}^*)$, it follows further that

$$\begin{aligned}\sum_{t=1}^{T-1} \widehat{G}_t' \widehat{\Omega}_t^{-1} \widehat{G}_t &= \sum_{t=1}^{T-1} \begin{bmatrix} 2\text{tr}(\mathring{W}_{t,1}^* \mathring{W}_{t,1}^*) & \cdots & 2\text{tr}(\mathring{W}_{t,1}^* \mathring{W}_{t,q}^*) & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 2\text{tr}(\mathring{W}_{t,q}^* \mathring{W}_{t,1}^*) & \cdots & 2\text{tr}(\mathring{W}_{t,q}^* \mathring{W}_{t,q}^*) & 0 \\ 0 & \cdots & 0 & \widehat{\sigma}^{-2} X_t^{+\prime} X_t^{+} \end{bmatrix} \\ &= \widehat{\sigma}^{-4} \begin{bmatrix} \widehat{\Phi}_Q & 0 \\ 0 & \widehat{\sigma}^2 \sum_{t=1}^{T-1} X_t^{+\prime} X_t^{+} \end{bmatrix}.\end{aligned}$$

Therefore, GMM-LM$_u = \widehat{V}_Q' \widehat{\Phi}_Q^{-1} \widehat{V}_Q = \mathcal{I}_u^2(q)$. □

Proof of Proposition 2. The first derivatives of the log-likelihood function are

$$\begin{aligned}\frac{\partial \ln L}{\partial \rho_r} &= -(T-1)\text{tr}(W_{1,r} R_1(\rho)^{-1}) + \frac{1}{\sigma^2} \sum_{t=1}^{T-1} (y_t^+ - X_t^+ \beta)' W_{1,r}' R_1(\rho)(y_t^+ - X_t^+ \beta), \\ \frac{\partial \ln L}{\partial \beta'} &= \frac{1}{\sigma^2} \sum_{t=1}^{T-1} X_t^{+\prime} R_1(\rho)' R_1(\rho)(y_t^+ - X_t^+ \beta), \\ \frac{\partial \ln L}{\partial \sigma^2} &= -\frac{n(T-1)}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{t=1}^{T-1} (y_t^+ - X_t^+ \beta)' R_1(\rho)' R_1(\rho)(u_t^+ - X_t^+ \beta),\end{aligned}$$

and some second derivatives are

$$\begin{aligned}\frac{\partial^2 \ln L}{\partial \rho_r \partial \rho_s} &= -(T-1)\text{tr}(W_{1,r} R_1(\rho)^{-1} W_{1,s} R_1(\rho)^{-1}) - \frac{1}{\sigma^2} \sum_{t=1}^{T-1} (y_t^+ - X_t^+ \beta)' W_{1,r}' W_{1,s}(y_t^+ - X_t^+ \beta), \\ \frac{\partial^2 \ln L}{\partial \rho_r \partial \beta'} &= -\frac{1}{\sigma^2} \sum_{t=1}^{T-1} X_t^{+\prime} [W_{1,r}' R_1(\rho) + R_1(\rho)' W_{1,r}](y_t^+ - X_t^+ \beta), \\ \frac{\partial^2 \ln L}{\partial \rho_r \partial \sigma^2} &= -\frac{1}{\sigma^4} \sum_{t=1}^{T-1} (y_t^+ - X_t^+ \beta)' W_{1,r}' R_1(\rho)(y_t^+ - X_t^+ \beta).\end{aligned}$$

Evaluated at the true parameters, the expected values of the second derivatives are

$$
\begin{aligned}
\mathrm{E}(\frac{\partial^2 \ln L}{\partial \rho_r \partial \rho_s}|_{\theta_0,\sigma_0^2}) &= -(T-1)[\mathrm{tr}(W_{1,r}R_1^{-1}W_{1,s}R_1^{-1}) + \mathrm{tr}(R_1'^{-1}W_{1,r}'W_{1,s}R_1^{-1})], \\
\mathrm{E}(\frac{\partial^2 \ln L}{\partial \rho_r \partial \beta'}|_{\theta_0,\sigma_0^2}) &= 0, \\
\mathrm{E}(\frac{\partial^2 \ln L}{\partial \rho_r \partial \sigma^2}|_{\theta_0,\sigma_0^2}) &= -\frac{1}{\sigma_0^2}(T-1)\mathrm{tr}(W_{1,r}R_1^{-1}).
\end{aligned}
$$

Evaluated at the restricted estimators under H_0^u,

$$
\begin{aligned}
\frac{\partial \ln L}{\partial \rho_r}|_{\widehat{\theta},\widehat{\sigma}^2} &= \frac{1}{\widehat{\sigma}^2}\sum_{t=1}^{T-1} \widehat{u}_t^{+\prime} W_{1,r}' \widehat{u}_t^{+}, \\
\frac{\partial \ln L}{\partial \beta'}|_{\widehat{\theta},\widehat{\sigma}^2} &= \frac{1}{\widehat{\sigma}^2}\sum_{t=1}^{T-1} X_t^{+\prime} \widehat{u}_t^{+} = 0, \\
\frac{\partial \ln L}{\partial \sigma^2}|_{\widehat{\theta},\widehat{\sigma}^2} &= -\frac{n(T-1)}{2\widehat{\sigma}^2} + \frac{1}{2\widehat{\sigma}^4}\sum_{t=1}^{T-1} \widehat{u}_t^{+\prime}\widehat{u}_t^{+} = 0,
\end{aligned}
$$

and

$$
\begin{aligned}
[\mathrm{E}(\frac{\partial^2 \ln L}{\partial \rho_r \partial \rho_s}|_{\theta_0,\sigma_0^2})]_{\widehat{\theta},\widehat{\sigma}^2} &= -(T-1)[\mathrm{tr}(W_{1,r}W_{1,s}) + \mathrm{tr}(W_{1,r}'W_{1,s})] = -2(T-1)\mathrm{tr}(\mathring{W}_{1,r}\mathring{W}_{1,s}), \\
[\mathrm{E}(\frac{\partial^2 \ln L}{\partial \rho_r \partial \sigma^2}|_{\theta_0,\sigma_0^2})]_{\widehat{\theta},\widehat{\sigma}^2} &= -\frac{1}{\widehat{\sigma}^2}(T-1)\mathrm{tr}(W_{1,r}) = 0.
\end{aligned}
$$

Consequently,

$$
\begin{aligned}
\text{ML-LM}_u &= \begin{bmatrix} \frac{\partial \ln L}{\partial \rho'} \\ \frac{\partial \ln L}{\partial \beta'} \\ \frac{\partial \ln L}{\partial \sigma^2} \end{bmatrix}_{\widehat{\theta},\widehat{\sigma}^2}' \left[-\mathrm{E}\left(\begin{matrix} \frac{\partial^2 \ln L}{\partial \rho \partial \rho'} & \frac{\partial^2 \ln L}{\partial \beta \partial \rho'} & \frac{\partial^2 \ln L}{\partial \sigma^2 \partial \rho'} \\ \frac{\partial^2 \ln L}{\partial \rho \partial \beta'} & \frac{\partial^2 \ln L}{\partial \beta \partial \beta'} & \frac{\partial^2 \ln L}{\partial \sigma^2 \partial \beta'} \\ \frac{\partial^2 \ln L}{\partial \rho \partial \sigma^2} & \frac{\partial^2 \ln L}{\partial \beta \partial \sigma^2} & \frac{\partial^2 \ln L}{(\partial \sigma^2)^2} \end{matrix} \Bigg|_{\theta_0,\sigma_0^2} \right) \right]_{\widehat{\theta},\widehat{\sigma}^2}^{-1} \begin{bmatrix} \frac{\partial \ln L}{\partial \rho'} \\ \frac{\partial \ln L}{\partial \beta'} \\ \frac{\partial \ln L}{\partial \sigma^2} \end{bmatrix}_{\widehat{\theta},\widehat{\sigma}^2} \\
&= \left[\frac{\partial \ln L}{\partial \rho} \right]_{\widehat{\theta},\widehat{\sigma}^2} \left[-\mathrm{E}\left(\frac{\partial^2 \ln L}{\partial \rho \partial \rho'}\Bigg|_{\theta_0,\sigma_0^2} \right) \right]_{\widehat{\theta},\widehat{\sigma}^2}^{-1} \left[\frac{\partial \ln L}{\partial \rho'} \right]_{\widehat{\theta},\widehat{\sigma}^2} = \mathcal{I}_u^2(q).
\end{aligned}
$$

The latter equality is readily seen to hold observing that $W_{t,r} = W_{1,r}$ for $t = 1, \cdots, T$ implies $W^*_{t,r} = \sum_{\tau=t}^{T} \pi^2_{t\tau} W_{1,r} = W_{1,r}$ and $\widehat{u}^{+\prime} W^*_r \widehat{u}^{+} = \sum_{t=1}^{T-1} \widehat{u}_t^{+\prime} W_{1,r} \widehat{u}_t^{+}$, and $\mathrm{tr}(\mathring{W}^*_r \mathring{W}^*_s) = (T-1)\mathrm{tr}(\mathring{W}_{1,r}\mathring{W}_{1,s})$. $\square$

Proof of Proposition 3. Recalling $\bar{X}_{t,r} = W_{t,r}X_t$, under the maintained assumptions, in particular, $\mathrm{E}(\epsilon_t|\mu) = 0$ and $\mathrm{E}(\mu) = 0$,

$$G_t = -\mathrm{E}[\frac{\partial g_t(\vartheta)}{\partial \vartheta}|_{\vartheta_0}] = \begin{bmatrix} G_t^{\lambda} & 0 & H'(R_t\bar{X}_{t,1})^+ & \cdots & H'(R_t\bar{X}_{t,q})^+ & H'(R_tX_t)^+ \\ G_{t,1}^{\lambda} & G_{t,1}^{\rho} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ G_{t,q}^{\lambda} & G_{t,q}^{\rho} & 0 & \cdots & 0 & 0 \end{bmatrix},$$

$$\Omega_t = \mathrm{E}[g_t(\vartheta_0)g_t(\vartheta_0)'] = \begin{bmatrix} \sigma_0^2 H'H & 0 \\ 0 & \Psi_Q \end{bmatrix}, \qquad \Psi_Q = \begin{bmatrix} \Omega_{11} & \cdots & \Omega_{1q} \\ \vdots & \ddots & \vdots \\ \Omega_{q1} & \cdots & \Omega_{qq} \end{bmatrix}$$

where $H = [X_1, \cdots, X_T, \bar{X}_{1,1}, \cdots, \bar{X}_{T,1}, \cdots, \bar{X}_{1,q}, \cdots, \bar{X}_{T,q}]$,

$$G_t^{\lambda} = \left[H'[R_tW_{t,1}S_t^{-1}(X_t\beta + \sum_{r=1}^{q}\bar{X}_{t,r}\gamma_r)]^+, \cdots, H'[R_tW_{t,q}S_t^{-1}(X_t\beta + \sum_{r=1}^{q}\bar{X}_{t,r}\gamma_r)]^+ \right],$$

$$G_{t,r}^{\lambda} = 2\sigma_0^2 \begin{bmatrix} \sum_{\tau=1}^{T}\pi^2_{t\tau}\mathrm{tr}(\mathring{W}_{1,r}R_\tau W_{\tau,1}S_\tau^{-1}R_\tau^{-1}) & \cdots & \sum_{\tau=1}^{T}\pi^2_{t\tau}\mathrm{tr}(\mathring{W}_{1,r}R_\tau W_{\tau,q}S_\tau^{-1}R_\tau^{-1}) \\ \vdots & \ddots & \vdots \\ \sum_{\tau=1}^{T}\pi^2_{t\tau}\mathrm{tr}(\mathring{W}_{T,r}R_\tau W_{\tau,1}S_\tau^{-1}R_\tau^{-1}) & \cdots & \sum_{\tau=1}^{T}\pi^2_{t\tau}\mathrm{tr}(\mathring{W}_{T,r}R_\tau W_{\tau,q}S_\tau^{-1}R_\tau^{-1}) \end{bmatrix},$$

$$G_{t,r}^{\rho} = 2\sigma_0^2 \begin{bmatrix} \sum_{\tau=1}^{T}\pi^2_{t\tau}\mathrm{tr}(\mathring{W}_{1,r}W_{\tau,1}R_\tau^{-1}) & \cdots & \sum_{\tau=1}^{T}\pi^2_{t\tau}\mathrm{tr}(\mathring{W}_{1,r}W_{\tau,q}R_\tau^{-1}) \\ \vdots & \ddots & \vdots \\ \sum_{\tau=1}^{T}\pi^2_{t\tau}\mathrm{tr}(\mathring{W}_{T,r}W_{\tau,1}R_\tau^{-1}) & \cdots & \sum_{\tau=1}^{T}\pi^2_{t\tau}\mathrm{tr}(\mathring{W}_{T,r}W_{\tau,q}R_\tau^{-1}) \end{bmatrix},$$

and, as in the proof of Proposition 1,

$$\Omega_{rs} = 2\sigma_0^4 \begin{bmatrix} \mathrm{tr}(\mathring{W}_{1,r}\mathring{W}_{1,s}) & \cdots & \mathrm{tr}(\mathring{W}_{1,r}\mathring{W}_{T,s}) \\ \vdots & \ddots & \vdots \\ \mathrm{tr}(\mathring{W}_{T,r}\mathring{W}_{1,s}) & \cdots & \mathrm{tr}(\mathring{W}_{T,r}\mathring{W}_{T,s}) \end{bmatrix},$$

for $r, s = 1, \cdots, q$. Evaluated at the restricted estimates $\widehat{\vartheta} = (0, \widehat{\beta}')'$ and $\widehat{\sigma}^2$, the LM test statistic is given by

$$\begin{aligned} \text{GMM-LM}_y &= \widehat{g}'\widehat{\Omega}^{-1}\widehat{G}(\widehat{G}'\widehat{\Omega}^{-1}\widehat{G})^{-}\widehat{G}'\widehat{\Omega}^{-1}\widehat{g} \\ &= \sum_{t=1}^{T-1} \widehat{g}_t'\widehat{\Omega}_t^{-1}\widehat{G}_t \left[\sum_{t=1}^{T-1} \widehat{G}_t'\widehat{\Omega}_t^{-1}\widehat{G}_t\right]^{-} \sum_{t=1}^{T-1} \widehat{G}_t'\widehat{\Omega}_t^{-1}\widehat{g}_t, \end{aligned}$$

with $\widehat{g}_t = [H, W_{1,1}\widehat{u}_t^+, \cdots, W_{T,1}\widehat{u}_t^+, \cdots, W_{1,q}\widehat{u}_t^+, \cdots, W_{T,q}\widehat{u}_t^+]'\widehat{u}_t^+$, and

$$\begin{aligned} \widehat{G}_t &= \begin{bmatrix} \widehat{G}_t^{\lambda} & 0 & H'\bar{X}_{t,1}^+ & \cdots & H'\bar{X}_{t,q}^+ & H'X_t^+ \\ \widehat{G}_{t,1}^{\lambda} & \widehat{G}_{t,1}^{\rho} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \widehat{G}_{t,q}^{\lambda} & \widehat{G}_{t,q}^{\rho} & 0 & \cdots & 0 & 0 \end{bmatrix}, \\ \widehat{\Omega}_t &= \begin{bmatrix} \widehat{\sigma}^2 H'H & 0 \\ 0 & \widehat{\Psi}_Q \end{bmatrix}, \qquad \widehat{\Psi}_Q = \begin{bmatrix} \widehat{\Omega}_{11} & \cdots & \widehat{\Omega}_{1q} \\ \vdots & \ddots & \vdots \\ \widehat{\Omega}_{q1} & \cdots & \widehat{\Omega}_{qq} \end{bmatrix}, \end{aligned}$$

where $\widehat{G}_t^{\lambda} = [H'\bar{X}_{t,1}^{+}\widehat{\beta}, \cdots, H'\bar{X}_{t,q}^{+}\widehat{\beta}]$ and, recalling that $W_{t,r}^{*} = \sum_{\tau=1}^{T} \pi_{t\tau}^2 W_{\tau,r}$,

$$\begin{aligned} \widehat{G}_{t,r}^{\lambda} &= \widehat{G}_{t,r}^{\rho} = 2\widehat{\sigma}^2 \begin{bmatrix} \mathrm{tr}(\mathring{W}_{1,r}W_{t,1}^{*}) & \cdots & \mathrm{tr}(\mathring{W}_{1,r}W_{t,q}^{*}) \\ \vdots & \ddots & \vdots \\ \mathrm{tr}(\mathring{W}_{T,r}W_{t,1}^{*}) & \cdots & \mathrm{tr}(\mathring{W}_{T,r}W_{t,q}^{*}) \end{bmatrix}, \\ \widehat{\Omega}_{rs} &= 2\widehat{\sigma}^4 \begin{bmatrix} \mathrm{tr}(\mathring{W}_{1,r}\mathring{W}_{1,s}) & \cdots & \mathrm{tr}(\mathring{W}_{1,r}\mathring{W}_{T,s}) \\ \vdots & \ddots & \vdots \\ \mathrm{tr}(\mathring{W}_{T,r}\mathring{W}_{1,s}) & \cdots & \mathrm{tr}(\mathring{W}_{T,r}\mathring{W}_{T,s}) \end{bmatrix}, \end{aligned}$$

for $r, s = 1, \cdots, q$. Let $\bar{\pi}_t = [\pi_{t1}^2, \cdots, \pi_{tT}^2]$, and, as was shown in the proof of Proposition 1, $\widehat{\Omega}_{rs}\bar{\pi}_t' = 2\widehat{\sigma}^4[\mathrm{tr}(\mathring{W}_{1,r}W_{t,s}^{*}), \ldots, \mathrm{tr}(\mathring{W}_{T,r}W_{t,s}^{*})]'$ and $\bar{\pi}_t\widehat{\Omega}_{rs}\bar{\pi}_t' = 2\widehat{\sigma}^4\mathrm{tr}(W_{t,r}^{*}W_{t,s}^{*})$. In light of this, $\widehat{G}_{t,r}^{\lambda} = \widehat{G}_{t,r}^{\rho} = \widehat{\sigma}^{-2}[\widehat{\Omega}_{r1}, \cdots, \widehat{\Omega}_{rq}](I_q \otimes \bar{\pi}_t')$ for $r = 1, \cdots, q$, and adopting the notation $\bar{X}_{[t]}^{+} = [\bar{X}_{t,1}^{+}, \cdots, \bar{X}_{t,q}^{+}]$

$$\begin{aligned} \widehat{G}_t &= \begin{bmatrix} H'\bar{X}_{[t]}^{+}(I_q \otimes \widehat{\beta}) & 0 & H'\bar{X}_{[t]}^{+} & H'X_t^{+} \\ \widehat{\sigma}^{-2}\widehat{\Psi}_Q(I_q \otimes \bar{\pi}_t') & \widehat{\sigma}^{-2}\widehat{\Psi}_Q(I_q \otimes \bar{\pi}_t') & 0 & 0 \end{bmatrix}, \\ \widehat{G}_t'\widehat{\Omega}_t^{-1} &= \widehat{\sigma}^{-2} \begin{bmatrix} (I_q \otimes \widehat{\beta}')\bar{X}_{[t]}^{+\prime}H(H'H)^{-1} & (I_q \otimes \bar{\pi}_t) \\ 0 & (I_q \otimes \bar{\pi}_t) \\ \bar{X}_{[t]}^{+\prime}H(H'H)^{-1} & 0 \\ X_t^{+\prime}H(H'H)^{-1} & 0 \end{bmatrix}. \end{aligned}$$

Let $\widehat{\Phi}_{Q,t} = (I_q \otimes \bar{\pi}_t)\widehat{\Psi}_Q(I_q \otimes \bar{\pi}_t')$. Observing that $H(H'H)^{-1}H'\bar{X}_{t,r}^{+} = \bar{X}_{t,r}^{+}$, $H(H'H)^{-1}H'X_t^{+} =$

X_t^+ and $\sum_{t=1}^{T-1} X_t^{+\prime}\widehat{u}_t^+ = 0$, we have

$$\widehat{G}_t'\widehat{\Omega}_t^{-1}\widehat{G}_t = \widehat{\sigma}^{-4}\begin{bmatrix} \widehat{\Phi}_{Q,t} + \widehat{\sigma}^2(I_q\otimes\widehat{\beta}')\bar{X}_{[t]}^{+\prime}\bar{X}_{[t]}^{+\prime}(I_q\otimes\widehat{\beta}) & \widehat{\Phi}_{Q,t} & \widehat{\sigma}^2(I_q\otimes\widehat{\beta}')\bar{X}_{[t]}^{+\prime}\bar{X}_{[t]}^{+} & \widehat{\sigma}^2(I_q\otimes\widehat{\beta}')\bar{X}_{[t]}^{+\prime}X_t^{+} \\ \widehat{\Phi}_{Q,t} & \widehat{\Phi}_{Q,t} & 0 & 0 \\ \widehat{\sigma}^2\bar{X}_{[t]}^{+\prime}\bar{X}_{[t]}^{+}(I_q\otimes\widehat{\beta}) & 0 & \widehat{\sigma}^2\bar{X}_{[t]}^{+\prime}\bar{X}_{[t]}^{+} & \widehat{\sigma}^2\bar{X}_{[t]}^{+\prime}X_t^{+} \\ \widehat{\sigma}^2X_t^{+\prime}\bar{X}_{[t]}^{+}(I_q\otimes\widehat{\beta}) & 0 & \widehat{\sigma}^2X_t^{+\prime}\bar{X}_{[t]}^{+} & \widehat{\sigma}^2X_t^{+\prime}X_t^{+} \end{bmatrix},$$

and

$$\sum_{t=1}^{T-1}\widehat{G}_t'\widehat{\Omega}_t^{-1}\widehat{g}_t = \widehat{\sigma}^{-2}\begin{bmatrix} \widehat{V}_Q + (I_q\otimes\widehat{\beta})'\widehat{V}_L \\ \widehat{V}_Q \\ \widehat{V}_L \\ 0 \end{bmatrix}.$$

Observing that $\sum_{t=1}^{T-1}\widehat{\Phi}_{Q,t} = \widehat{\Phi}_Q$ and

$$M_{11} = \sum_{t=1}^{T-1}\bar{X}_{[t]}^{+\prime}\bar{X}_{[t]}^{+} = \begin{bmatrix} \bar{X}_1^{+\prime}\bar{X}_1^{+} & \cdots & \bar{X}_1^{+\prime}\bar{X}_q^{+} \\ \vdots & \ddots & \vdots \\ \bar{X}_q^{+\prime}\bar{X}_1^{+} & \cdots & \bar{X}_q^{+\prime}\bar{X}_q^{+} \end{bmatrix}, \qquad M_{21} = \sum_{t=1}^{T-1}X_t^{+\prime}\bar{X}_{[t]}^{+} = \left[X^{+\prime}\bar{X}_1^{+}, \cdots, X^{+\prime}\bar{X}_q^{+}\right],$$

we have

$$\sum_{t=1}^{T-1}\widehat{G}_t'\widehat{\Omega}_t^{-1}\widehat{G}_t = \widehat{\sigma}^{-4}\begin{bmatrix} \widehat{\Phi}_Q + \widehat{\sigma}^2(I_q\otimes\widehat{\beta})'M_{11}(I_q\otimes\widehat{\beta}) & \widehat{\Phi}_Q & \widehat{\sigma}^2(I_q\otimes\widehat{\beta})'M_{11} & \widehat{\sigma}^2(I_q\otimes\widehat{\beta})'M_{21}' \\ \widehat{\Phi}_Q & \widehat{\Phi}_Q & 0 & 0 \\ \widehat{\sigma}^2M_{11}(I_q\otimes\widehat{\beta}) & 0 & \widehat{\sigma}^2M_{11} & \widehat{\sigma}^2M_{21}' \\ \widehat{\sigma}^2M_{21}(I_q\otimes\widehat{\beta}) & 0 & \widehat{\sigma}^2M_{21} & \widehat{\sigma}^2X^{+\prime}X^{+} \end{bmatrix}.$$

Let

$$\widehat{V}_* = \begin{bmatrix} \widehat{V}_Q \\ \widehat{V}_L \\ 0 \end{bmatrix}, \qquad \widehat{\Phi}_* = \begin{bmatrix} \widehat{\Phi}_Q & 0 & 0 \\ 0 & \widehat{\sigma}^2 M_{11} & \widehat{\sigma}^2 M_{21}' \\ 0 & \widehat{\sigma}^2 M_{21} & \widehat{\sigma}^2 X^{+\prime} X^+ \end{bmatrix}, \qquad \Gamma = \begin{bmatrix} I_q \\ I_q \otimes \widehat{\beta} \\ 0 \end{bmatrix}.$$

Then,

$$\sum_{t=1}^{T-1} \widehat{G}_t' \widehat{\Omega}_t^{-1} \widehat{g}_t = \widehat{\sigma}^{-2} \begin{bmatrix} \Gamma' \widehat{V}_* \\ \widehat{V}_* \end{bmatrix}, \qquad \sum_{t=1}^{T-1} \widehat{G}_t' \widehat{\Omega}_t^{-1} \widehat{G}_t = \widehat{\sigma}^{-4} \begin{bmatrix} \Gamma' \widehat{\Phi}_* \Gamma & \Gamma' \widehat{\Phi}_* \\ \widehat{\Phi}_* \Gamma & \widehat{\Phi}_* \end{bmatrix}.$$

It follows by a similar argument as in the proof of Proposition 1 in Liu and Prucha (2018)

$$\begin{aligned}
\text{GMM-LM}_y &= \widehat{V}_*' \widehat{\Phi}_*^{-1} \widehat{V}_* \\
&= \begin{bmatrix} \widehat{V}_Q \\ \widehat{V}_L \\ 0 \end{bmatrix}' \begin{bmatrix} \widehat{\Phi}_Q & 0 & 0 \\ 0 & \widehat{\sigma}^2 M_{11} & \widehat{\sigma}^2 M_{21}' \\ 0 & \widehat{\sigma}^2 M_{21} & \widehat{\sigma}^2 X^{+\prime} X^+ \end{bmatrix}^{-1} \begin{bmatrix} \widehat{V}_Q \\ \widehat{V}_L \\ 0 \end{bmatrix} \\
&= \begin{bmatrix} \widehat{V}_Q \\ \widehat{V}_L \end{bmatrix}' \begin{bmatrix} \widehat{\Phi}_Q & 0 \\ 0 & \widehat{\sigma}^2 M_{11} - \widehat{\sigma}^2 M_{21}' (X^{+\prime} X^+)^{-1} M_{21} \end{bmatrix}^{-1} \begin{bmatrix} \widehat{V}_Q \\ \widehat{V}_L \end{bmatrix} \\
&= \mathcal{I}_y^2(q).
\end{aligned}$$

□

Proof of Proposition 4. Let $\epsilon_t^+(\vartheta) = R_1(\rho)[S_1(\lambda) y_t^+ - X_t^+ \beta - \sum_{r=1}^q W_{1,r} X_t^+ \gamma_r]$. The first derivatives of the log-likelihood function are

$$\begin{aligned}
\frac{\partial \ln L}{\partial \lambda_r} &= -(T-1)\mathrm{tr}(W_{1,r}S_1(\lambda)^{-1}) + \frac{1}{\sigma^2}\sum_{t=1}^{T-1}\epsilon_t^+(\vartheta)'R_1(\rho)W_{1,r}y_t^+ \\
\frac{\partial \ln L}{\partial \rho_r} &= -(T-1)\mathrm{tr}(W_{1,r}R_1(\rho)^{-1}) + \frac{1}{\sigma^2}\sum_{t=1}^{T-1}\epsilon_t^+(\vartheta)'W_{1,r}[S_1(\lambda)y_t^+ - X_t^+\beta - \sum_{r=1}^{q}W_{1,r}X_t^+\gamma_r] \\
\frac{\partial \ln L}{\partial \gamma_r'} &= \frac{1}{\sigma^2}\sum_{t=1}^{T-1}X_t^{+\prime}W_{1,r}'R_1(\rho)'\epsilon_t^+(\vartheta) \\
\frac{\partial \ln L}{\partial \beta'} &= \frac{1}{\sigma^2}\sum_{t=1}^{T-1}X_t^{+\prime}R_1(\rho)'\epsilon_t^+(\vartheta) \\
\frac{\partial \ln L}{\partial \sigma^2} &= -\frac{n(T-1)}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{t=1}^{T-1}\epsilon_t^+(\vartheta)'\epsilon_t^+(\vartheta).
\end{aligned}$$

Evaluated at the true parameters, the expected values of the second-order derivatives are

$$\begin{aligned}
\mathrm{E}(\frac{\partial^2 \ln L}{\partial \lambda_r \partial \lambda_s}|_{\vartheta_0,\sigma_0^2}) &= -(T-1)\mathrm{tr}(W_{1,r}S_1^{-1}W_{1,s}S_1^{-1}) - (T-1)\mathrm{tr}[(R_1W_{1,s}S_1^{-1}R_1^{-1})'(R_1W_{1,r}S_1^{-1}R_1^{-1})] \\
&\quad -\frac{1}{\sigma^2}\sum_{t=1}^{T-1}(R_1W_{1,s}S_1^{-1}F_t^+)'(R_1W_{1,r}S_1^{-1}F_t^+), \\
\mathrm{E}(\frac{\partial^2 \ln L}{\partial \lambda_r \partial \rho_s}|_{\vartheta_0,\sigma_0^2}) &= -(T-1)\mathrm{tr}(W_{1,s}W_{1,r}S_1^{-1}R_1^{-1}) - (T-1)\mathrm{tr}[(R_1W_{1,r}S_1^{-1}R_1^{-1})'W_{1,s}R_1^{-1}], \\
\mathrm{E}(\frac{\partial^2 \ln L}{\partial \lambda_r \partial \gamma_s'}|_{\vartheta_0,\sigma_0^2}) &= -\frac{1}{\sigma^2}\sum_{t=1}^{T-1}X_t^{+\prime}W_{1,s}'R_1'R_1W_{1,r}S_1^{-1}F_t^+, \\
\mathrm{E}(\frac{\partial^2 \ln L}{\partial \lambda_r \partial \beta'}|_{\vartheta_0,\sigma_0^2}) &= -\frac{1}{\sigma^2}\sum_{t=1}^{T-1}X_t^{+\prime}R_1'R_1W_{1,r}S_1^{-1}F_t^+, \\
\mathrm{E}(\frac{\partial^2 \ln L}{\partial \lambda_r \partial \sigma^2}|_{\vartheta_0,\sigma_0^2}) &= -\frac{1}{\sigma^2}(T-1)\mathrm{tr}(R_1W_{1,r}S_1^{-1}R_1^{-1}),
\end{aligned}$$

$$
\begin{aligned}
\mathrm{E}(\frac{\partial^2 \ln L}{\partial\rho_r\partial\rho_s}|_{\vartheta_0,\sigma_0^2}) &= -(T-1)\mathrm{tr}(W_{1,r}R_1^{-1}W_{1,s}R_1^{-1}) - (T-1)\mathrm{tr}[(W_{1,s}R_1^{-1})'W_{1,r}R_1^{-1}],\\
\mathrm{E}(\frac{\partial^2 \ln L}{\partial\rho_r\partial\gamma_s'}|_{\vartheta_0,\sigma_0^2}) &= \mathrm{E}(\frac{\partial^2 \ln L}{\partial\rho_r\partial\beta'}|_{\vartheta_0,\sigma_0^2}) = 0,\\
\mathrm{E}(\frac{\partial^2 \ln L}{\partial\rho_r\partial\sigma^2}|_{\vartheta_0,\sigma_0^2}) &= -\frac{1}{\sigma^2}(T-1)\mathrm{tr}(W_{1,r}R_1^{-1}),
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{E}(\frac{\partial^2 \ln L}{\partial\gamma_s\partial\gamma_r'}|_{\vartheta_0,\sigma_0^2}) &= -\frac{1}{\sigma^2}\sum_{t=1}^{T-1} X_t^{+\prime}W_{1,r}'R_1'R_1W_{1,s}X_t^{+},\\
\mathrm{E}(\frac{\partial^2 \ln L}{\partial\beta\partial\gamma_r'}|_{\vartheta_0,\sigma_0^2}) &= -\frac{1}{\sigma^2}\sum_{t=1}^{T-1} X_t^{+\prime}W_{1,r}'R_1'R_1X_t^{+},\\
\mathrm{E}(\frac{\partial^2 \ln L}{\partial\beta\partial\beta'}|_{\vartheta_0,\sigma_0^2}) &= -\frac{1}{\sigma^2}\sum_{t=1}^{T-1} X_t^{+\prime}R_1'R_1X_t^{+},\\
\mathrm{E}(\frac{\partial^2 \ln L}{\partial\sigma^2\partial\gamma_r'}|_{\vartheta_0,\sigma_0^2}) &= \mathrm{E}(\frac{\partial^2 \ln L}{\partial\sigma^2\partial\beta'}|_{\vartheta_0,\sigma_0^2}) = 0,\\
\mathrm{E}(\frac{\partial^2 \ln L}{\partial(\sigma^2)^2}|_{\vartheta_0,\sigma_0^2}) &= -\frac{n(T-1)}{2\sigma^4},
\end{aligned}
$$

where $F_t^+ \equiv X_t^+\beta_0 + \sum_{r=1}^{q} W_r X_t^+\gamma_{r0}$. Evaluated at the restricted estimators under H_0^y, we have $\frac{\partial \ln L}{\partial\lambda_r}|_{\hat\vartheta,\hat\sigma^2} = \hat\sigma^{-2}\sum_{t=1}^{T-1}\hat u_t^{+\prime}W_{1,r}y_t^+$, $\frac{\partial \ln L}{\partial\rho_r}|_{\hat\vartheta,\hat\sigma^2} = \hat\sigma^{-2}\sum_{t=1}^{T-1}\hat u_t^{+\prime}W_{1,r}'\hat u_t^+$, $\frac{\partial \ln L}{\partial\gamma_r'}|_{\hat\vartheta,\hat\sigma^2} =$

$\widehat{\sigma}^{-2}\sum_{t=1}^{T-1} X_t^{+\prime} W_{1,r}' \widehat{u}_t^+$, $\frac{\partial \ln L}{\partial \beta'}|_{\widehat{\vartheta},\widehat{\sigma}^2} = 0$, $\frac{\partial \ln L}{\partial \sigma^2}|_{\widehat{\vartheta},\widehat{\sigma}^2} = 0$,

$$\begin{aligned}
[\mathrm{E}(\frac{\partial^2 \ln L}{\partial \lambda_r \partial \lambda_s}|_{\vartheta_0,\sigma_0^2})]_{\widehat{\vartheta},\widehat{\sigma}^2} &= -2(T-1)\mathrm{tr}(\mathring{W}_{1,r}\mathring{W}_{1,s}) - \frac{1}{\widehat{\sigma}^2}\sum_{t=1}^{T-1} \widehat{\beta}' X_t^{+\prime} W_{1,s}' W_{1,r} X_t^+ \widehat{\beta}, \\
[\mathrm{E}(\frac{\partial^2 \ln L}{\partial \lambda_r \partial \rho_s}|_{\vartheta_0,\sigma_0^2})]_{\widehat{\vartheta},\widehat{\sigma}^2} &= [\mathrm{E}(\frac{\partial^2 \ln L}{\partial \rho_r \partial \rho_s}|_{\vartheta_0,\sigma_0^2})]_{\widehat{\vartheta},\widehat{\sigma}^2} = -2(T-1)\mathrm{tr}(\mathring{W}_{1,r}\mathring{W}_{1,s}), \\
[\mathrm{E}(\frac{\partial^2 \ln L}{\partial \lambda_r \partial \gamma_s'}|_{\vartheta_0,\sigma_0^2})]_{\widehat{\vartheta},\widehat{\sigma}^2} &= -\frac{1}{\widehat{\sigma}^2}\sum_{t=1}^{T-1} X_t^{+\prime} W_{1,s}' W_{1,r} X_t^+ \widehat{\beta}, \\
[\mathrm{E}(\frac{\partial^2 \ln L}{\partial \lambda_r \partial \beta'}|_{\vartheta_0,\sigma_0^2})]_{\widehat{\vartheta},\widehat{\sigma}^2} &= -\frac{1}{\widehat{\sigma}^2}\sum_{t=1}^{T-1} X_t^{+\prime} W_{1,r} X_t^+ \widehat{\beta}, \\
[\mathrm{E}(\frac{\partial^2 \ln L}{\partial \gamma_s \partial \gamma_r'}|_{\vartheta_0,\sigma_0^2})]_{\widehat{\vartheta},\widehat{\sigma}^2} &= -\frac{1}{\widehat{\sigma}^2}\sum_{t=1}^{T-1} X_t^{+\prime} W_{1,r}' W_{1,s} X_t^+, \\
[\mathrm{E}(\frac{\partial^2 \ln L}{\partial \beta \partial \gamma_r'}|_{\vartheta_0,\sigma_0^2})]_{\widehat{\vartheta},\widehat{\sigma}^2} &= -\frac{1}{\widehat{\sigma}^2}\sum_{t=1}^{T-1} X_t^{+\prime} W_{1,r}' X_t^+, \\
[\mathrm{E}(\frac{\partial^2 \ln L}{\partial \beta \partial \beta'}|_{\vartheta_0,\sigma_0^2})]_{\widehat{\vartheta},\widehat{\sigma}^2} &= -\frac{1}{\widehat{\sigma}^2}\sum_{t=1}^{T-1} X_t^{+\prime} X_t^+, \\
[\mathrm{E}(\frac{\partial^2 \ln L}{\partial (\sigma^2)^2}|_{\vartheta_0,\sigma_0^2})]_{\widehat{\vartheta},\widehat{\sigma}^2} &= -\frac{n(T-1)}{2(\widehat{\sigma}^2)^2},
\end{aligned}$$

and $[\mathrm{E}(\frac{\partial^2 \ln L}{\partial \lambda_r \partial \sigma^2}|_{\vartheta_0,\sigma_0^2})]_{\widehat{\vartheta},\widehat{\sigma}^2} = [\mathrm{E}(\frac{\partial^2 \ln L}{\partial \rho_r \partial \gamma_{k,s}}|_{\vartheta_0,\sigma_0^2})]_{\widehat{\vartheta},\widehat{\sigma}^2} = [\mathrm{E}(\frac{\partial^2 \ln L}{\partial \rho_r \partial \beta_k}|_{\vartheta_0,\sigma_0^2})]_{\widehat{\vartheta},\widehat{\sigma}^2} = [\mathrm{E}(\frac{\partial^2 \ln L}{\partial \rho_r \partial \sigma^2}|_{\vartheta_0,\sigma_0^2})]_{\widehat{\vartheta},\widehat{\sigma}^2} = [\mathrm{E}(\frac{\partial^2 \ln L}{\partial \sigma^2 \partial \gamma_{k,r}}|_{\vartheta_0,\sigma_0^2})]_{\widehat{\vartheta},\widehat{\sigma}^2} = [\mathrm{E}(\frac{\partial^2 \ln L}{\partial \sigma^2 \partial \beta_k}|_{\vartheta_0,\sigma_0^2})]_{\widehat{\vartheta},\widehat{\sigma}^2} = 0$. Hence,

$$\begin{aligned}
\text{ML-LM}_y &= \begin{bmatrix} \frac{\partial \ln L}{\partial \vartheta'} \\ \frac{\partial \ln L}{\partial \sigma^2} \end{bmatrix}'_{\widehat{\vartheta},\widehat{\sigma}^2} \left[-\mathrm{E}\left(\begin{matrix} \frac{\partial^2 \ln L}{\partial \vartheta \partial \vartheta'} & \frac{\partial^2 \ln L}{\partial \sigma^2 \partial \vartheta'} \\ \frac{\partial^2 \ln L}{\partial \vartheta \partial \sigma^2} & \frac{\partial^2 \ln L}{(\partial \sigma^2)^2} \end{matrix} \Bigg|_{\vartheta_0,\sigma_0^2} \right) \right]^-_{\widehat{\vartheta},\widehat{\sigma}^2} \begin{bmatrix} \frac{\partial \ln L}{\partial \vartheta'} \\ \frac{\partial \ln L}{\partial \sigma^2} \end{bmatrix}_{\widehat{\vartheta},\widehat{\sigma}^2} \\
&= \begin{bmatrix} \frac{\partial \ln L}{\partial \vartheta'}|_{\widehat{\vartheta},\widehat{\sigma}^2} \\ 0 \end{bmatrix}' \left[-\mathrm{E}\left(\begin{matrix} \frac{\partial^2 \ln L}{\partial \vartheta \partial \vartheta'} & 0 \\ 0 & \frac{\partial^2 \ln L}{(\partial \sigma^2)^2} \end{matrix} \Bigg|_{\vartheta_0,\sigma_0^2} \right) \right]^-_{\widehat{\vartheta},\widehat{\sigma}^2} \begin{bmatrix} \frac{\partial \ln L}{\partial \vartheta'}|_{\widehat{\vartheta},\widehat{\sigma}^2} \\ 0 \end{bmatrix}_{\widehat{\vartheta},\widehat{\sigma}^2} \\
&= \left[\frac{\partial \ln L}{\partial \vartheta'}|_{\widehat{\vartheta},\widehat{\sigma}^2} \right]' \left[-\mathrm{E}\left(\frac{\partial^2 \ln L}{\partial \vartheta \partial \vartheta'}\Bigg|_{\vartheta_0,\sigma_0^2} \right) \right]^-_{\widehat{\vartheta},\widehat{\sigma}^2} \left[\frac{\partial \ln L}{\partial \vartheta'}|_{\widehat{\vartheta},\widehat{\sigma}^2} \right].
\end{aligned}$$

Let

$$\widehat{V}_* = \widehat{\sigma}^2 \begin{bmatrix} \frac{\partial \ln L}{\partial \rho'}\big|_{\widehat{\vartheta},\widehat{\sigma}^2} \\ \frac{\partial \ln L}{\partial \gamma_1'}\big|_{\widehat{\vartheta},\widehat{\sigma}^2} \\ \vdots \\ \frac{\partial \ln L}{\partial \gamma_q'}\big|_{\widehat{\vartheta},\widehat{\sigma}^2} \\ 0 \end{bmatrix}, \qquad \widehat{\Phi}_* = -\widehat{\sigma}^4 \left[\mathrm{E} \left(\begin{matrix} \frac{\partial^2 \ln L}{\partial \rho \partial \rho'} & 0 & \cdots & 0 & 0 \\ 0 & \frac{\partial^2 \ln L}{\partial \gamma_1 \partial \gamma_1'} & \cdots & \frac{\partial^2 \ln L}{\partial \gamma_q \partial \gamma_1'} & \frac{\partial^2 \ln L}{\partial \beta \partial \gamma_1'} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \frac{\partial^2 \ln L}{\partial \gamma_1 \partial \gamma_q'} & \cdots & \frac{\partial^2 \ln L}{\partial \gamma_q \partial \gamma_q'} & \frac{\partial^2 \ln L}{\partial \beta \partial \gamma_q'} \\ 0 & \frac{\partial^2 \ln L}{\partial \gamma_1 \partial \beta'} & \cdots & \frac{\partial^2 \ln L}{\partial \gamma_q \partial \beta'} & \frac{\partial^2 \ln L}{\partial \beta \partial \beta'} \end{matrix} \Bigg|_{\vartheta_0,\sigma_0^2} \right) \right]_{\widehat{\vartheta},\widehat{\sigma}^2}.$$

Since $W_{t,r} = W_{1,r}$ for all t, we have

$$\widehat{V}_* = \begin{bmatrix} \widehat{V}_Q \\ \widehat{V}_L \\ 0 \end{bmatrix}, \qquad \widehat{\Phi}_* = \begin{bmatrix} \widehat{\Phi}_Q & 0 & 0 \\ 0 & \widehat{\sigma}^2 M_{11} & \widehat{\sigma}^2 M_{21}' \\ 0 & \widehat{\sigma}^2 M_{21} & \widehat{\sigma}^2 X^{+\prime} X^{+} \end{bmatrix},$$

where

$$M_{11} = \begin{bmatrix} \bar{X}_1^{+\prime} \bar{X}_1^{+} & \cdots & \bar{X}_1^{+\prime} \bar{X}_q^{+} \\ \vdots & \ddots & \vdots \\ \bar{X}_q^{+\prime} \bar{X}_1^{+} & \cdots & \bar{X}_q^{+\prime} \bar{X}_q^{+} \end{bmatrix}, \qquad M_{21} = \left[X^{+\prime} \bar{X}_1^{+}, \cdots, X^{+\prime} \bar{X}_q^{+} \right].$$

Then, for $\Gamma = [I_q, I_q \otimes \widehat{\beta}', 0]'$, we have

$$\frac{\partial \ln L}{\partial \vartheta'}\big|_{\widehat{\vartheta},\widehat{\sigma}^2} = \begin{bmatrix} \Gamma' \widehat{V}_* \\ \widehat{V}_* \end{bmatrix}, \qquad \left[-\mathrm{E} \left(\frac{\partial^2 \ln L}{\partial \vartheta \partial \vartheta'} \Big|_{\vartheta_0,\sigma_0^2} \right) \right]_{\widehat{\vartheta},\widehat{\sigma}^2} = \begin{bmatrix} \Gamma' \widehat{\Phi}_* \Gamma & \Gamma' \widehat{\Phi}_* \\ \widehat{\Phi}_* \Gamma & \widehat{\Phi}_* \end{bmatrix}.$$

It follows by a similar argument as in the proof of Proposition 1 in Liu and Prucha (2018)

$$
\begin{aligned}
\text{ML-LM}_y &= \widehat{V}_*'\widehat{\Phi}_*^{-1}\widehat{V}_* \\
&= \begin{bmatrix} \widehat{V}_Q \\ \widehat{V}_L \\ 0 \end{bmatrix}' \begin{bmatrix} \widehat{\Phi}_Q & 0 & 0 \\ 0 & \widehat{\sigma}^2 M_{11} & \widehat{\sigma}^2 M_{21}' \\ 0 & \widehat{\sigma}^2 M_{21} & \widehat{\sigma}^2 X^{+\prime} X^{+} \end{bmatrix}^{-1} \begin{bmatrix} \widehat{V}_Q \\ \widehat{V}_L \\ 0 \end{bmatrix} \\
&= \begin{bmatrix} \widehat{V}_Q \\ \widehat{V}_L \end{bmatrix}' \begin{bmatrix} \widehat{\Phi}_Q & 0 \\ 0 & \widehat{\sigma}^2 M_{11} - \widehat{\sigma}^2 M_{21}'(X^{+\prime}X^{+})^{-1} M_{21} \end{bmatrix}^{-1} \begin{bmatrix} \widehat{V}_Q \\ \widehat{V}_L \end{bmatrix} \\
&= \mathcal{I}_y^2(q).
\end{aligned}
$$

□

D Assumptions: Additional Discussions

D.1 Sufficient Conditions for Assumption 3

Assumption 3 is a high level assumption. In the following we explore lower level sufficient conditions. In preparation of this exploration we first collect some useful results in a lemma. Parts (i) and (ii) of the lemma below restate results given, e.g., in Remark A.1 of Kelejian and Prucha (2010) for the convenience of the reader. Parts (iii) and (iv) extend parts of Remark A.1 to accommodate fixed effects. The reason is that while the Helmert transformation removes fixed effects from $\mu + \epsilon_t$, it does not remove them from regressors that are spatial lags of y_t, if the weight matrices vary over time. Proofs for the lemmata given in this subsection are given in a subsequent subsection.

Lemma D.1. *Let A_n and B_n be nonstochastic $n \times n$ matrices whose row and column sums of the absolute elements are bounded uniformly by finite constants K_A and K_B, let a_n and b_n be some nonstochastic $n \times 1$ vectors whose elements are bounded uniformly in absolute value by some finite constants K_a and K_b. Then:*

(i) *The row and column sums of the absolute elements of $A_n B_n$ are bounded uniformly by $K_A K_B$.*

(ii) *The elements of $A_n a_n$ are bounded uniformly in absolute value by the constant $K_A K_a$ and $n^{-1} b_n' A_n a_n$ is uniformly bounded in absolute value by $K_A K_a K_b$.*

Furthermore, let $\xi_n = s_n + S_n \xi_n^*$ where the s_n are nonstochastic $n \times 1$ vectors whose elements are bounded uniformly in absolute value by some finite constant K_s, the S_n are nonstochastic $n \times n$ matrices where the elements of $\Sigma_n = S_n S_n'$ are uniformly bounded in absolute value by some finite constant K_σ, and the ξ_n^* are $n \times 1$ random vectors whose *elements are i.i.d.* $(0,1)$*, and hence* $\xi_n \sim (s_n, \Sigma_n)$*. Assume furthermore the elements of* ξ_n^* *have uniformly bounded finite* $4+\delta$ *moments for some* $\delta > 0$*. Then:*

(iii) $n^{-1} a_n' \xi_n$ *is* $O_p(1)$*, and furthermore if* $\xi_n = \sigma_\xi \xi_n^*$ *we have* $n^{-1/2} a_n' \xi_n$ *is* $O_p(1)$*.*

(iv) $n^{-1} \xi_n' A_n \xi_n$ *is* $O_p(1)$*, and furthermore for random vectors* ς_n *with* $\mathrm{E}(\varsigma_n \mid \xi_n) = 0$ *and* $\mathrm{E}(\varsigma_n \varsigma_n' \mid \xi_n) = \sigma_\varsigma^2 I_n$ *we have* $n^{-1} \xi_n' A_n \varsigma_n = o_p(1)$ *and* $n^{-1} \varsigma_n' A_n \varsigma_n = n^{-1} \mathrm{E}(\varsigma_n' A_n \varsigma_n) + o_p(1)$ *with* $n^{-1} \mathrm{E}(\varsigma_n' A_n \varsigma_n) = \sigma_\varsigma^2 n^{-1} \mathrm{tr}(A_n) = O(1)$*.*

As remarked in the text, the conditions on A postulated in Assumption 3 hold under the assumptions maintained for the weight matrices W_r for the leading application where $A = \mathring{W}_r^*$, and of course the conditions also hold for $A = I_{n(T-1)}$. We next postulate lower level assumptions on the regressors in Z_t. We then give a lemma that shows that, under those assumptions, the conditions of Assumption 3 that $n^{-1} Z^{+\prime} A Z^{+} = O_p(1)$ and $n^{-1} Z^{+\prime} A \epsilon^{+} = n^{-1} \mathrm{E}(Z^{+\prime} A \epsilon^{+}) + o_p(1)$, where $n^{-1} \mathrm{E}(Z^{+\prime} A \epsilon^{+}) = O(1)$, hold. We then show that in particular the regressors of a higher order Cliff-Ord network model satisfy the postulated lower level assumptions.

Assumption D.1. *The columns of* Z_t *are of the form*

$$z_{tk,n} = c_{tk,n} + C_{tk,n}(\mu_n + \epsilon_{t,n}) \tag{D.1}$$

where the $c_{tk,n}$ *are nonstochastic* $n \times 1$ *vectors whose elements are uniformly bounded in absolute value by a finite constant, say,* Δ_c*, the* C_{tk} *are nonstochastic* $n \times n$ *matrices whose row and column sums of the absolute elements are uniformly bounded by a finite constant, say,* Δ_C*. Furthermore the fixed effects are of the form* $\mu_n = \nu_n + V_n^{1/2} \xi_n^*$ *where the* ν_n *are nonstochastic* $n \times 1$ *vectors whose elements are uniformly bounded in absolute value by a* *finite constant* Δ_v*, the* $V_n = V_n^{1/2}V_n^{1/2\prime}$ *are nonstochastic* $n \times n$ *matrices whose elements are uniformly bounded in absolute value by a finite constant* Δ_V*, and the elements of the* $n \times 1$ *vectors* ξ_n^* *are i.i.d.* $(0,1)$ *with uniformly bounded finite* $4+\delta$ *moments for some* $\delta > 0$*. Furthermore* $E(\epsilon_{t,n} \mid \mu_n) = 0$*.*

The assumption allows for fairly general forms of fixed effects μ_n. Note that we only assume that the elements of ν_n and V_n are uniformly bounded in absolute value. We do not impose bounds on the row and column sums of V_n.

Lemma D.2. *Suppose the innovations satisfy Assumption 1,* $A_n = \text{diag}_{t=1}^{T-1}\{A_{t,n}\}$ *satisfies the conditions postulated in Assumption 3, and the columns of* Z_t *satisfy Assumption D.1. Then the remainder of Assumption 3 holds, i.e.,* $n^{-1}Z_n^{+\prime}A_nZ_n^+ = O_p(1)$ *and* $n^{-1}Z_n^{+\prime}A_n\epsilon_n^+ = n^{-1}\text{E}(Z_n^{+\prime}A_n\epsilon_n^+) + o_p(1)$*, where* $n^{-1}\text{E}(Z_n^{+\prime}A_n\epsilon_n^+) = O(1)$*.*

We now apply the above lemma illustratively to verify that under H_0^u all regressors of the higher-order spatial Cliff-Ord model (21) considered in the text satisfy the conditions postulated in Assumption 4 regarding the regressors. Under H_0^u model (21) is given by

$$y_t = \sum_{r=1}^{q} \lambda_{r0} W_{t,r} y_t + X_t\beta_0 + \sum_{r=1}^{q} W_{t,r} X_t \gamma_{r0} + u_t, \text{ and } u_t = \mu + \epsilon_t, \quad \text{for } t = 1, \cdots, T, \quad \text{(D.2)}$$

and the regressor matrix is given by $Z_t = [W_{t,1}y_t, \ldots, W_{t,q}y_t, X_t, W_{t,1}X_t, \ldots, W_{t,q}X_t]$.[21] Also assume, as common in the spatial literature, that $S_t = I_n - \sum_{r=1}^{q} \lambda_{r0} W_{t,r}$ is nonsingular and that the row and column sums of the absolute elements of S_t^{-1} are uniformly bounded by some finite constant, and that the fixed effects satisfy the conditions postulated in Assumption D.1. We now verify that the regressors in Z_t satisfy the conditions

[21] Under H_0^y the regressors constitute a subset of those considered under H_0^y, and thus the subsequent discussion also covers H_0^y.

postulated in Assumption D.1. Note that the elements of X_t are uniformly bounded in absolute value by Assumption 4. By Lemma D.1(ii) it follows further that the elements of all spatial lags $W_{t,k}X_t$ are uniformly bounded in absolute value. Hence all columns of $X_t, W_{t,1}X_t, \dots, W_{t,q}X_t$ are of the form $z_{tk} = c_{tk}$ and satisfy Assumption D.1 with $C_{tk} = 0$. Next consider the spatial lags of $W_{t,k}y_t$ in Z_t. Observe that

$$y_t = S_t^{-1}X_t\beta_0 + \sum_{r=1}^{q} S_t^{-1}W_{t,r}X_t\gamma_{r0} + S_t^{-1}(\mu + \epsilon_t),$$

and consequently

$$z_{tk} = W_{t,k}y_t = c_{tk} + C_{tk}(\mu + \epsilon_t)$$

with $c_{tk} = W_{t,k}S_t^{-1}X_t\beta_0 + \sum_{r=1}^{q} W_{t,k}S_t^{-1}W_{t,r}X_t\gamma_{r0}$ and $C_{tk} = W_{t,k}S_t^{-1}$. Under the maintained assumption the row and column sums of $W_{t,k}S_t^{-1}$ and $W_{t,k}S_t^{-1}W_{t,r}$ are uniformly bounded in absolute value by Lemma D.1(i), and thus C_{tk} satisfies Assumption D.1. Furthermore by Lemma D.1(ii) the elements of c_{tk} are uniformly bounded in absolute value, and thus also the c_{tk} satisfy Assumption D.1. Having verified the assumptions of postulated on $z_{z_{tk}}$ in Assumption D.1 it now follows from Lemma D.2 that $n^{-1}Z_n^{+\prime}A_nZ_n^{+}$ and $n^{-1}Z_n^{+\prime}A_n\epsilon_n^{+}$ satisfy the conditions postulated in Assumption 3.

D.2 Proofs of Lemmata

Proof of Lemma D.1. For parts (i) and (ii) see Remark A.1 in Kelejian and Prucha (2010). To verify the first claim of part (iii) observe that $\mathrm{E}[(n^{-1}a_n'\xi_n)^2] = n^{-2}(a_n's_n)^2 + n^{-2}a_n'\Sigma_na_n \leq K_a^2K_s^2 + K_\sigma K_s^2$. The claim now follows from Fuller (1976), Corollary 5.1.1.1. Analogously, to verify the second claim of part (iii) observe that $\mathrm{E}[(n^{-1/2}a_n'\xi_n)^2] = n^{-1}\sigma_\xi^2a_n'a_n \leq \sigma_\xi^2K_\alpha^2$.

To verify the first claim of part (iv) we maintain w.o.l.o.g. that A_n is symmetric, given that $\xi_n' A_n \xi_n = \xi_n'[(A_n + A_n')/2]\xi_n$. Observe that

$$n^{-1}\xi_n' A_n \xi_n = n^{-1}s_n' A_n s_n + n^{-1}s_n' A_n S_n \xi_n^* + n^{-1}\xi_n^{*\prime} S_n' A_n s_n + n^{-1}\xi_n^{*\prime} S_n' A_n S_n \xi_n^*.$$

It now follows immediately form part (ii) that $n^{-1}s_n' A_n s_n = O(1)$. Next observe that $\mathrm{E}[n^{-1}s_n' A_n S_n \xi_n^*] = 0$ and $\mathrm{Var}[n^{-1}s_n' A_n S_n \xi_n^*] = n^{-2}s_n' A_n \Sigma_n A_n s_n$. In light of part (ii) we have that the elements of $A_n s_n$ are uniformly bounded by $K_A K_s$, and thus $n^{-2}s_n' A_n \Sigma_n A_n s_n \leq K_A^2 K_s^2 K_\sigma$. It now follows from Fuller (1976), Corollary 5.1.1.2, that $n^{-1}s_n' A_n S_n \xi_n^*$ and analogously $n^{-1}\xi_n^{*\prime} S_n' A_n s_n$ are $O_p(1)$. Next observe that $n^{-1}\mathrm{E}[\xi_n^{*\prime} S_n' A S_n \xi_n^*] = n^{-1}\mathrm{tr}[A_n \Sigma_n]$. Let $a_{i.,n}$ and $\sigma_{.i,n}$ denote the i-th row and column of A_n and Σ_n, respectively. Then $\mathrm{tr}[A_n \Sigma_n] = \sum_{i=1}^n a_{i.,n}\sigma_{.i,n}$. Observing that $|a_{i.,n}\sigma_{.i,n}| \leq K_\sigma \sum_{j=1}^n |a_{ij,n}| \leq K_\sigma K_A$ we have $|n^{-1}\mathrm{E}[\xi_n^{*\prime} S_n' A_n S_n \xi_n^*]| \leq K_\sigma K_A$ and thus $n^{-1}\mathrm{E}[\xi_n^{*\prime} S_n' A_n S_n \xi_n^*] = O(1)$. Next observe that in light of, say, Kelejian and Prucha (2001) we have

$$\mathrm{Var}(n^{-1}\xi_n^{*\prime} S_n' A_n S_n \xi_n^*) = n^{-2}2\mathrm{tr}(D_n D_n) + n^{-2}\sum_{i=1}^n d_{ii,n}^2(\mathrm{E}\xi_{i,n}^{*4} - 3).$$

$D_n = (d_{ij,n}) = S_n' A_n S_n$. Recalling that $\Sigma_n = S_n S_n'$ we have $\mathrm{tr}(D_n D_n) = \mathrm{tr}(A_n \Sigma_n A_n \Sigma_n)$. Consider the (k,l)th element of $A_n \Sigma_n$ given by $a_{k.,n}\sigma_{.l,n}$, then by argumentation analogous to above $|a_{k.,n}\sigma_{.l,n}| \leq K_S K_A$. This in turn implies that all elements of $A_n \Sigma_n A_n \Sigma_n$ are bounded in absolute value by $nK_\sigma^2 K_A^2$. Hence $\left|n^{-2}\mathrm{tr}(D_n D_n)\right| \leq K_\sigma^2 K_A^2$. Next observe that $|\sigma_{ij,n}| = |\sum_{r=1}^n s_{ir,n}s_{jr,n}| \leq K_\sigma$, and as a special that $|\sigma_{ii,n}| = \sum_{r=1}^n s_{ir,n}^2 \leq K_\sigma$. Note that the latter implies that $s_{ir,n}^2 \leq K_\sigma$ and thus $|s_{ir,n}|/K_\sigma^{1/2} \leq 1$ and also that $\sum_{r=1}^n |s_{ir,n}| \leq K_\sigma \sum_{r=1}^n s_{ir,n}^2 \leq K_\sigma^2$. Observing that $d_{ii,n} = \sum_{l=1}^n \sum_{k=1}^n a_{kl,n}s_{li,n}s_{ki,n}$, it then follows that

$$\begin{aligned}
\sum_{i=1}^{n} d_{ii,n}^2 &\leq \sum_{i=1}^{n}\sum_{l=1}^{n}\sum_{k=1}^{n}\sum_{u=1}^{n}\sum_{v=1}^{n} |a_{kl,n}a_{uv,n}s_{ki,n}s_{li,n}s_{ui,n}s_{vi,n}| \\
&\leq K_\sigma^{3/2} \sum_{l=1}^{n}\sum_{k=1}^{n}\sum_{u=1}^{n}\sum_{v=1}^{n} |a_{kl,n}|\,|a_{uv,n}| \sum_{i=1}^{n} |s_{vi,n}| \\
&\leq K_\sigma^{2+3/2} \sum_{k=1}^{n}\sum_{u=1}^{n}\sum_{l=1}^{n} |a_{kl,n}| \sum_{v=1}^{n} |a_{uv,n}| \leq K_\sigma^{2+3/2} \sum_{k=1}^{n}\sum_{u=1}^{n} K_A^2 = n^2 K_\sigma^{2+3/2} K_A^2.
\end{aligned}$$

Since $\mathrm{E}\xi_{i,n}^{*4}$ is uniformly bounded, it follows that also $n^{-2}\sum_{i=1}^{n} d_{ii,n}^2(\mathrm{E}\xi_{i,n}^{*4} - 3)$ is uniformly bounded by a finite constant, and thus that $\mathrm{Var}(n^{-1}\xi_n^{*\prime} S_n' A_n S_n \xi_n^*) = O(1)$. Using again Corollary 5.1.1.2 of Fuller (1976) we have $n^{-1}\xi_n^{*\prime} S_n' A_n S_n \xi_n^* = O_p(1)$, which completes the proof of the first claim of part (iv).

To prove the second claim of part (iv) observe that $n^{-1}\mathrm{E}(\xi_n' A_n \varsigma_n) = 0$ and $\mathrm{Var}(n^{-1}\xi_n' A_n \varsigma_n) = n^{-2}\mathrm{E}[\mathrm{tr}(\xi_n' A_n \varsigma_n \varsigma_n' A_n \xi_n)] = n^{-2}\mathrm{tr}[A_n \mathrm{E}(\varsigma_n \varsigma_n' \mid \xi_n) A_n \xi_n \xi_n'] = \sigma_\varsigma^2 n^{-2}\mathrm{tr}[A_n A_n \Sigma_n]$ by iterated expectations. Under the maintained assumptions it follows from part(ii) off the lemma that the elements of $A_n A_n \Sigma_n$ are bounded uniformly in absolute value by $K_A^2 K_\sigma$and $\mathrm{Var}(n^{-1}\xi_n' A_n \varsigma_n) \leq \sigma_\varsigma^2 K_A^2 K_\sigma / n \to 0$. The claim now follows from Chebyshev's inequality.

The third claim of part (iv) follows immediately from Kelejian and Prucha (2001). □

Proof of Lemma D.2. Recall that $Z_n^+ = [Z_{1,n}^{+\prime}, \cdots, Z_{T-1,n}^{+\prime}]'$, $\epsilon_n^+ = [\epsilon_{1,n}^{+\prime}, \cdots, \epsilon_{T-1,n}^{+\prime}]'$, and thus $n^{-1}Z_n^{+\prime} A_n Z_n^+ = \sum_{t=1}^{T-1} n^{-1} Z_{t,n}^{+\prime} A_{t,n} Z_{t,n}^+$ and $n^{-1}Z_n^{+\prime} A_n \epsilon_n^+ = \sum_{t=1}^{T-1} n^{-1} Z_{t,n}^{+\prime} A_{t,n} \epsilon_{t,n}^+$. To prove the lemma it thus suffices to show that $n^{-1}Z_{t,n}^{+\prime} A_{t,n} Z_{t,n}^+ = O_p(1)$ and $n^{-1}Z_{t,n}^{+\prime} A_{t,n} \epsilon_{t,n}^+ = n^{-1}\mathrm{E}(Z_{t,n}^{+\prime} A_{t,n} \epsilon_{t,n}^+) + o_p(1)$, where $n^{-1}\mathrm{E}(Z_{t,n}^{+\prime} A_{t,n} \epsilon_{t,n}^+) = O(1)$. For ease of notation we drop subscripts n in the following. The (k,l)th element of $n^{-1}Z_t^{+\prime}A_tZ_t^+$ is given by

$$
\begin{aligned}
n^{-1}z_{tk}^{+\prime}A_tz_{tl}^{+} &= n^{-1}c_{tk}^{+\prime}A_tc_{tl}^{+} + n^{-1}c_{tk}^{+\prime}A_t[C_{tl}(\mu+\epsilon_t)]^{+} + n^{-1}[(\mu+\epsilon_t)'C_{tk}']^{+}A_tc_{tl}^{+} \\
&\quad + n^{-1}[(\mu+\epsilon_t)'C_{tk}']^{+}A_t[C_{tl}(\mu+\epsilon_t)]^{+}. \qquad \text{(D.3)}
\end{aligned}
$$

To prove the claim we consider each term on the r.h.s. separately. Observe that by Lemma D.1(ii) that the elements of A_tc_{sl} are uniformly bounded in absolute value by $\Delta_A\Delta_c$, and thus

$$
\begin{aligned}
\left|n^{-1}c_{tk}^{+\prime}A_tc_{tl}^{+}\right| &= n^{-1}\left|\sum_{\tau,s=1}^{T}\pi_{t\tau}\pi_{t\tau}c_{\tau k}'A_tc_{sl}\right| \leq n^{-1}\sum_{\tau,s=1}^{T}\pi_{t\tau}\pi_{t\tau}\left|c_{\tau k}'A_tc_{sl}\right| \\
&\leq \sum_{\tau,s=1}^{T}\pi_{t\tau}\pi_{t\tau}\Delta_c^2\Delta_A \leq T^2\Delta_c^2\Delta_A,
\end{aligned}
$$

observing that $|\pi_{t\tau}| \leq 1$ and T is finite. This shows that $n^{-1}c_{tk}^{+\prime}A_tc_{tl}^{+} = O(1)$.

Next observe that

$$
n^{-1}c_{tk}^{+\prime}A_t[C_{tl}(\mu+\epsilon_t)]^{+} = n^{-1}\sum_{\tau,s=1}^{T}\pi_{t\tau}\pi_{ts}c_{\tau k}'A_tC_{sl}\mu + n^{-1}\sum_{\tau,s=1}^{T}\pi_{t\tau}\pi_{ts}c_{\tau k}'A_tC_{sl}\epsilon_t.
$$

By Lemma D.1(i) and (ii) the elements of all vectors $c_{\tau k}'A_tC_{sl}$ are uniformly bounded in absolute value by $\Delta_c\Delta_C\Delta_A$. Observing that μ satisfies the conditions postulated for ξ in Lemma D.1 it follows immediately from the first part of Lemma D.1(iii) that all terms $n^{-1}c_{\tau k}'A_tC_{sl}\mu$ are $O_p(1)$. Observing that ϵ_t satisfies the conditions of ξ postulated for the second part of Lemma D.1(iii) it follows further that all terms $n^{-1}c_{\tau k}'A_tC_{sl}\epsilon_t$ are also $O_p(1)$. Since $|\pi_{t\tau}| \leq 1$ and T is finite this shows that $n^{-1}c_{tk}^{+\prime}A_t[C_{tl}(\mu+\epsilon_t)]^{+} = O_p(1)$ and analogously that $n^{-1}[(\mu+\epsilon_t)'C_{tk}']^{+}A_tc_{tl}^{+} = O_p(1)$.

Finally observe that

$$
\begin{aligned}
& n^{-1}[(\epsilon_t+\mu)'C_{tk}']^+ A_t [C_{tl}(\epsilon_t+\mu)]^+ \\
= \; & n^{-1}\sum_{\tau,s=1}^{T}\pi_{t\tau}\pi_{ts}\epsilon_t' C_{\tau k}' A_t C_{sl}\epsilon_t + 2n^{-1}\sum_{\tau,s=1}^{T}\pi_{t\tau}\pi_{ts}\mu' C_{\tau k}' A_t C_{sl}\epsilon_t + n^{-1}\sum_{\tau,s=1}^{T}\pi_{t\tau}\pi_{ts}\mu' C_{\tau k}' A_t C_{sl}\mu.
\end{aligned}
$$

By Lemma D.1(i) the row and column sums of the absolute elements of the matrices $C_{\tau k}' A_t C_{sl}$ are uniformly bounded by $\Delta_C^2\Delta_A$. It now follows immediately from the first part of Lemma D.1(iv) that the terms $n^{-1}\epsilon_t' C_{\tau k}' A_t C_{sl}\epsilon_t$ and $n^{-1}\mu' C_{\tau k}' A_t C_{sl}\mu$ are $O_p(1)$, observing respectively that both ϵ_t and μ satisfy the conditions postulated for ξ in the lemma. The terms $n^{-1}\mu' C_{\tau k}' A_t C_{sl}\epsilon_t$ are seen to be $o_p(1)$ from the second part of Lemma D.1(iv) upon associating ξ with μ and ς_n with ϵ_t. Since $|\pi_{t\tau}| \leq 1$ and T is finite. this shows that $n^{-1}[(\epsilon_t+\mu)'C_{tk}']^+ A_t [C_{tl}(\epsilon_t+\mu)]^+ = O_p(1)$. Having shown that each term on the r.h.s. of (D.3) is at most $O_p(1)$ it follows that $n^{-1}z_{tk}^{+\prime} A_t z_{tl}^+ = O_p(1)$, which completes the proof of the first claim.

To prove the second claim, observe that the kth element of $n^{-1}Z_t^{+\prime} A_t \epsilon_t^+$ is given by

$$
n^{-1}z_{tk}^{+\prime} A_t \epsilon_t^+ = n^{-1}c_{tk}^{+\prime} A_t \epsilon_t^+ + n^{-1}[\mu' C_{tk}']^+ A_t \epsilon_t^+ + n^{-1}[\epsilon_t' C_{tk}']^+ A_t \epsilon_t^+.
$$

To prove the claim we consider each term on the r.h.s. separately. Observe that

$$
n^{-1}c_{tk}^{+\prime} A_t \epsilon_t^+ = n^{-1}\sum_{\tau,s=1}^{T}\pi_{t\tau}\pi_{ts} c_{\tau k}' A_t \epsilon_s,
$$

and thus clearly $n^{-1}\mathrm{E}(c_{tk}^{+\prime} A_t \epsilon_t^+) = 0$. Recalling from the above discussion that the elements of $c_{\tau k}' A_t$ are uniformly bounded in absolute value it follows from Lemma D.1(iii) upon associating ξ with ϵ_s that each of the terms $n^{-1}c_{\tau k}' A_t \epsilon_s$ are $o_p(1)$ and thus that $n^{-1}c_{tk}^{+\prime} A_t \epsilon_t^+ = o_p(1)$. We have shown above $n^{-1}\mathrm{E}(\mu' C_{\tau k}' A_t C_{sl}\epsilon_t) = 0$ and $n^{-1}\mu' C_{\tau k}' A_t C_{sl}\epsilon_t = o_p(1)$, and thus $\mathrm{E}(n^{-1}[\mu' C'_{tk}]^+ A_t \epsilon_t^+) = 0$ and $n^{-1}[\mu' C'_{tk}]^+ A_t \epsilon_t^+ = o_p(1)$ is seen to hold as a special case, taking $C_{sl} = I_n$. Next consider

$$n^{-1}[\epsilon_t' C'_{tk}]^+ A_t \epsilon_t^+ = n^{-1} \sum_{\tau,s=1}^{T} \pi_{t\tau} \pi_{ts} \epsilon_t' C'_{\tau k} A_t \epsilon_s.$$

Let $B_t = [C'_{\tau k} A_t + A_t C_{\tau k}]$, then in light of Lemma D.1(i) the row and column sums of B_t are uniformly bounded in absolute value. It now follows from the third claim of Lemma D.1(iv) that $n^{-1} \epsilon_t' C'_{\tau k} A_t \epsilon_s = n^{-1} \mathrm{E}(\epsilon_t' C'_{\tau k} A_t \epsilon_s) + o_p(1)$ with $n^{-1} \mathrm{E}(\epsilon_t' C'_{\tau k} A_t \epsilon_s) = O(1)$, and thus

$$n^{-1} z_{tk}^{+\prime} A_t \epsilon_t^+ = n^{-1} \mathrm{E}(z_{tk}^{+\prime} A_t \epsilon_t^+) + o_p(1)$$

with

$$n^{-1} \mathrm{E}(z_{tk}^{+\prime} A_t \epsilon_t^+) = \sum_{\tau,s=1}^{T} \pi_{t\tau} \pi_{ts} n^{-1} \mathrm{E}(\epsilon_t' C'_{\tau k} A_t \epsilon_s) = O(1).$$

□

E Additional Monte Carlo Simulations

In this appendix, we first provide Monte Carlo simulation results on the performance of the proposed $\mathcal{I}^2(q)$ tests when q is large. For comparison we also report on the performance of the Holm procedure. We then provide Monte Carlo simulation results for situations where the weight matrices W_t are endogenous.

E.1 Performance of $\mathcal{I}^2(q)$ Tests When q is Large

To generate the weight matrices $W_{t,r}$ (for $r = 1, \cdots, q$) for these Monte Carlo simulations, we partition n individuals into equal-sized groups with 10 individuals in each group. Let $\xi_{t,1}, \cdots, \xi_{t,q}$ be $n \times 1$ random vectors generated from a multivariate normal distribution with zero mean, unit variance and pairwise covariance (between $\xi_{t,r}$ and $\xi_{t,s}$) given by ϕ. Let $D_{t,r}$ (for $r = 1, \cdots, q$) be an observed $n \times n$ zero-diagonal matrix of indicator variables with the (i, j)th element being one if and only if individuals i and j are in the same group and $\left|\xi_{it,r} - \xi_{jt,r}\right| \leq 1$, where $\xi_{it,r}$ denotes the ith element of $\xi_{t,r}$. The weight matrices $W_{t,r}$ (for $r = 1, \cdots, q$) are then obtained by row-sum normalizing $D_{t,r}$ so that each non-zero row of $W_{t,r}$ sums to one.

For the $\mathcal{I}_u^2(q)$ tests, y_t is generated as

$$y_t = X_t\beta + u_t, \tag{E.1}$$

and, for the $\mathcal{I}_y^2(q)$ tests, y_t is generated as

$$y_t = \lambda_1 W_{t,1} y_t + X_t\beta + W_{t,1} X_t \gamma_1 + u_t, \tag{E.2}$$

where in both cases

$$u_t = \rho_1 W_{t,1} u_t + \mu + \epsilon_t,$$

for $t = 1, \cdots, T$. The individual effects μ_i and innovations ϵ_{it} are generated independently from $N(0,1)$. The observations on the two exogenous variables in the $n \times 2$ matrix X_t are generated independently from a Uniform$[0,3]$. We set $\beta = (1,1)'$ in the data generating process.

[Insert Tables E.1-E.4 here]

Simulation results for $n = 500$ and $T = 5$ based on 50,000 repetitions are reported in Tables E.1-E.4. We note that for the considered data generating processes, $W_{t,1}$ correctly models the network topology, whereas $W_{t,2}, \cdots, W_{t,q}$ are misspecified weight matrices. The reported results indicate that the actual sizes of the $\mathcal{I}_u^2(q)$ and $\mathcal{I}_y^2(q)$ tests are close to the asymptotic nominal size of 0.05. We find the power of the $\mathcal{I}_u^2(q)$ and $\mathcal{I}_y^2(q)$ tests decreases as q increases but the decrease is mostly modest. We also find that the Holm test tends to under-reject the null hypothesis. The downward size distortion of the Holm test is more severe as q gets larger and the correlation between the weight matrices (captured by ϕ) increases.

E.2 Performance of $\mathcal{I}^2(q)$ Tests with Endogenous Weight Matrices

In the following we report on Monte Carlo simulations for scenarios where the weight matrix is endogenous. In line with our discussion in the Section 4, we consider two forms of endogeneity. The first case arises when the weight matrix is correlated with the individual effects μ, but not with the idiosyncratic disturbances ϵ_t. In this case we can still use the weight matrix W_t in forming our test statistics. The second case arises when the weight matrix is also correlated with the idiosyncratic disturbances ϵ_t. In this latter case we consider test statistics obtained by replacing the actual weight matrix with "projected" or "instrumented" weight matrices, which only depend on exogenous variables.

For the $\mathcal{I}_u^2(q)$ tests, y_t is as generated as

$$y_t = X_t\beta + u_t,$$

and, for the $\mathcal{I}_y^2(q)$ tests, y_t is generated as

$$y_t = \lambda W_t y_t + X_t\beta + W_t X_t \gamma + u_t,$$

where in both cases

$$u_t = \rho W_t u_t + \mu + \epsilon_t,$$

for $t = 1, \cdots, T$. The individual effects μ_i and innovations ϵ_{it} are generated independently from $N(0,1)$. The elements of the $n \times 2$ matrix $X_t = [x_{it,k}]$ are given by $x_{it,k} = \mu_i + \tilde{x}_{it,k}$, where $\tilde{x}_{it,k}$ is drawn independently from Uniform$[0,3]$. That is, we allow for correlation between the $x_{it,k}$ and the u_{it} through the individual effects. We set $\beta = (1,1)'$ in the data generating process.

To generate W_t, we partition n individuals into equal-sized groups with 10 individuals in each group. Suppose that for each individual we observe two characteristics generated as $\xi_{it,r} = \mu_i + \tilde{\xi}_{it,r}$, for $r = 1, 2$, where $\tilde{\xi}_{it,r}$ is drawn independently from $N(0,1)$. Let $D_{t,r} = [d_{ij,t,r}]$ be an $n \times n$ zero-diagonal matrix of indicator variables. When i and j are not in the same group, define $d_{ij,t,r} = 0$. When i and j are in the same group, define $d_{ij,t,r} = 1$ if $\xi_{it,r}$ and $\xi_{jt,r}$ are in the same quartile of their distribution.

Suppose links in W_t are formed if i and j are in the same group and

$$\bar{d}_{ij,t,1} + \mu_i + \mu_j + e_{ij,t} > 0, \tag{E.3}$$

where $\bar{d}_{ij,t,1}$ is the standardized $d_{ij,t,1}$. We consider two specifications of $e_{ij,t}$. In the first specification, the $e_{ij,t}$ are generated i.i.d. $N(0,1)$, independently from the idiosyncratic disturbances ϵ_{it}. In the second specification, $e_{ij,t} = (\epsilon_{it} + \epsilon_{jt})/\sqrt{2}$, thus allowing for dependence between the $e_{ij,t}$ and the idiosyncratic disturbances ϵ_{it}. The weight matrix W_t is the row-sum normalized adjacency matrix of the resulting network. For an exemplary interpretation, the design of the weight matrix W_t is motivated by a friendship network based on a simple homophily link formation model (E.3), where two individuals in the same group are more likely to form a link if they share some specific characteristic given by $\xi_{it,1}$. In that sense, $D_{t,1}$ is a more informative proxy for the actual adjacency matrix W_t compared to $D_{t,2}$. Although $d_{ij,t,2}$ does not show up in the link formation model (E.3), $D_{t,2}$ is not entirely uninformative because (i) it captures the group structure of the network and (ii) $\xi_{it,1}$ and $\xi_{it,2}$ are correlated due to the individual effects μ_i in their definitions.

[Insert Tables E.5-E.8 here]

Simulation results for $n \in \{250, 500\}$ and $T = 5$ based on 20,000 repetitions are reported in Tables E.5-E.8. For Tables E.5 and E.6, the $e_{ij,t}$ are generated independently from the idiosyncratic disturbances ϵ_t, and W_t is correlated with the error term u_t of the main regression only via the individual effects μ. As the Helmert transformation eliminates the individual effects in the error term u_t, W_t becomes uncorrelated with the Helmert transformed error terms. Hence, and as confirmed by the simulations, in this case the $\mathcal{I}^2(1)$ test based on W_t has the proper size. In contrast, for Tables E.7 and E.8, $e_{ij,t} = (\epsilon_{it} + \epsilon_{jt})/\sqrt{2}$, and W_t is correlated with the error term u_t of the main regression via both the individual effects μ and the idiosyncratic disturbances ϵ_t. Hence, the weight matrix W_t remains endogenous after the Helmert transformation. As a result, in this case, the $\mathcal{I}^2(1)$ tests based on W_t exhibit severe upward size distortions. Observing that the matrices $D_{t,1}$ and/or $D_{t,2}$ are uncorrelated with the Helmert transformed error term, they can be used as exogenous "proxies" for W_t to construct $\mathcal{I}^2(q)$ test statistics. We find that the $\mathcal{I}^2(q)$ tests based on $D_{t,1}$ and/or $D_{t,2}$ are properly sized and the power increases as the amount of cross sectional dependence increases. As expected, since $D_{t,1}$ enters explicitly into the link formation process (E.3), the $\mathcal{I}^2(1)$ test with $D_{t,1}$ outperforms that with $D_{t,2}$. The $\mathcal{I}^2(2)$ test with both $D_{t,1}$ and $D_{t,2}$ has less power than the $\mathcal{I}^2(1)$ test with $D_{t,1}$ but the power loss is modest. The results suggest that the $\mathcal{I}^2(2)$ test, incorporating both $D_{t,1}$ and $D_{t,2}$, offers researchers significant robustness when they are uncertain which of $D_{t,1}$ or $D_{t,2}$ is a better proxy for W_t.

Table E.1. Rejection Rates for $I_u^2(q)$ Tests with $W_1, \ldots, W_q$ ($\mathrm{n} = 500, \mathrm{T} = 5, \phi = 0$)

ρ_1	$q = 1$	$q = 2$	$q = 5$	$q = 10$
0	0.0504	0.0507	0.0506	0.0503
.2	0.8227	0.7459	0.6105	0.4823
.4	0.9999	0.9998	0.9992	0.9961
.6	1.0000	1.0000	1.0000	1.0000
.8	1.0000	1.0000	1.0000	1.0000

Nominal size is 0.05. The DGP is Defined by Equation (E1).

Table E.2. Rejection Rates for $I_y^2(q)$ Tests with $W_1, \ldots, W_q$ ($\mathrm{n} = 500, \mathrm{T} = 5, \phi = 0$)

λ_1	ρ_1	γ_1	$q = 1$	$q = 2$	$q = 5$	$q = 10$
0	0	0	0.0502	0.0499	0.0491	0.0506
.1	0	0	0.9890	0.9714	0.9031	0.7921
.2	0	0	1.0000	1.0000	1.0000	1.0000
0	.2	0	0.9996	0.9988	0.9923	0.9714
0	.4	0	1.0000	1.0000	1.0000	1.0000
.1	.2	0	1.0000	1.0000	1.0000	1.0000
.2	.4	0	1.0000	1.0000	1.0000	1.0000
0	0	.1	0.8881	0.8026	0.6211	0.4589
0	0	.2	1.0000	1.0000	0.9997	0.9978

Nominal size is 0.05. The DGP is Defined by Equation (E2).

Table E.3. Rejection Rates for $I_u^2(q)$ Tests with $W_1, \ldots, W_q$ ($\mathrm{n} = 500, \mathrm{T} = 5, \rho_1 = 0$)

	$q = 1$	$q = 2$	$q = 5$	$q = 10$
	$\phi = 0$			
$I_u^2(q)$ Test	0.0504	0.0507	0.0506	0.0503
Holm Test	0.0504	0.0475	0.0406	0.0360
	$\phi = 0.9$			
$I_u^2(q)$ Test	0.0504	0.0500	0.0505	0.0501
Holm Test	0.0504	0.0441	0.0309	0.0235

Nominal size is 0.05. The DGP is Defined by Equation (E1).

Table E.4. Rejection Rates for $I_y^2(q)$ Tests with $W_1, \ldots, W_q$ ($\mathrm{n} = 500, \mathrm{T} = 5, \lambda_1 = \rho_1 = \gamma_1 = 0$)

	$q = 1$	$q = 2$	$q = 5$	$q = 10$
	$\phi = 0$			
$I_y^2(q)$ Test	0.0502	0.0499	0.0491	0.0506
Holm Test	0.0502	0.0495	0.0457	0.0455
	$\phi = 0.9$			
$I_y^2(q)$ Test	0.0502	0.0513	0.0507	0.0507
Holm Test	0.0502	0.0473	0.0385	0.0317

Nominal size is 0.05. The DGP is Defined by Equation (E2).

Table E.5. Rejection Rates for $I_u^2(q)$ Tests when the Network Links are Correlated with μ_i but not ϵ_{it}

ρ	$I_u^2(1)$ Test with W	$I_u^2(1)$ Test with D_1	$I_u^2(1)$ Test with D_2	$I_u^2(2)$ Test with D_1, D_2
	$n = 250, T = 5$			
0	0.0532	0.0515	0.0502	0.0510
.2	0.9829	0.8435	0.5693	0.7994
.4	1.0000	1.0000	0.9984	1.0000
.6	1.0000	1.0000	1.0000	1.0000
.8	1.0000	1.0000	1.0000	1.0000
	$n = 500, T = 5$			
0	0.0508	0.0510	0.0480	0.0493
.2	0.9998	0.9834	0.8476	0.9740
.4	1.0000	1.0000	1.0000	1.0000
.6	1.0000	1.0000	1.0000	1.0000
.8	1.0000	1.0000	1.0000	1.0000

Nominal size is 0.05

Table E.6. Rejection Rates for $I_y^2(q)$ Tests when the Network Links are Correlated with μ_i but not ϵ_{it}

λ	ρ	γ	$I_y^2(1)$ Test with W	$I_y^2(1)$ Test with D_1	$I_y^2(1)$ Test with D_2	$I_y^2(2)$ Test with D_1, D_2
			$n = 250, T = 5$			
0	0	0	0.0517	0.0508	0.0495	0.0508
.2	0	0	1.0000	0.8824	0.5060	0.8306
.4	0	0	1.0000	1.0000	0.9971	1.0000
0	.2	0	0.9785	0.7281	0.4345	0.6649
0	.4	0	1.0000	1.0000	0.9948	1.0000
.2	.2	0	1.0000	0.9999	0.9691	0.9995
.4	.4	0	1.0000	1.0000	1.0000	1.0000
0	0	.5	1.0000	0.8248	0.1923	0.7360
0	0	1	1.0000	0.9994	0.5785	0.9975
			$n = 500, T = 5$			
0	0	0	0.0500	0.0512	0.0471	0.0477
.2	0	0	1.0000	0.9971	0.8229	0.9930
.4	0	0	1.0000	1.0000	1.0000	1.0000
0	.2	0	1.0000	0.9545	0.7298	0.9296
0	.4	0	1.0000	1.0000	1.0000	1.0000
.2	.2	0	1.0000	1.0000	1.0000	1.0000
.4	.4	0	1.0000	1.0000	1.0000	1.0000
0	0	.5	1.0000	0.9998	0.4330	0.9987
0	0	1	1.0000	1.0000	0.9434	1.0000

Nominal size is 0.05

Table E.7. Rejection Rates for $I_u^2(q)$ Tests when the Network Links are Correlated with both μ_i and ϵ_{it}

ρ	$I_u^2(1)$ Test with W	$I_u^2(1)$ Test with D_1	$I_u^2(1)$ Test with D_2	$I_u^2(2)$ Test with D_1, D_2
	n = 250, T = 5			
0	0.2829	0.0481	0.0474	0.0471
.2	0.7634	0.6413	0.4221	0.5906
.4	1.0000	0.9996	0.9908	0.9996
.6	1.0000	1.0000	1.0000	1.0000
.8	1.0000	1.0000	1.0000	1.0000
	n = 500, T = 5			
0	0.4310	0.0479	0.0487	0.0467
.2	0.9696	0.9013	0.6955	0.8758
.4	1.0000	1.0000	1.0000	1.0000
.6	1.0000	1.0000	1.0000	1.0000
.8	1.0000	1.0000	1.0000	1.0000

Nominal size is 0.05

Table E.8. Rejection Rates for $I_y^2(q)$ Tests when the Network Links are Correlated with both μ_i and ϵ_{it}

λ	ρ	γ	$I_y^2(1)$ Test with W	$I_y^2(1)$ Test with D_1	$I_y^2(1)$ Test with D_2	$I_y^2(2)$ Test with D_1, D_2
			n = 250, T = 5			
0	0	0	0.6063	0.0498	0.0461	0.0470
.2	0	0	1.0000	0.7390	0.4360	0.6840
.4	0	0	1.0000	1.0000	0.9735	1.0000
0	.2	0	0.9950	0.5101	0.3018	0.4508
0	.4	0	1.0000	0.9981	0.9723	0.9971
.2	.2	0	1.0000	0.9932	0.9246	0.9899
.4	.4	0	1.0000	1.0000	1.0000	1.0000
0	0	.5	1.0000	0.9442	0.3335	0.8961
0	0	1	1.0000	1.0000	0.8208	0.9998
			n = 500, T = 5			
0	0	0	0.7301	0.0502	0.0500	0.0492
.2	0	0	1.0000	0.9860	0.7250	0.9758
.4	0	0	1.0000	1.0000	1.0000	1.0000
0	.2	0	1.0000	0.8093	0.5485	0.7633
0	.4	0	1.0000	1.0000	0.9998	1.0000
.2	.2	0	1.0000	1.0000	0.9980	1.0000
.4	.4	0	1.0000	1.0000	1.0000	1.0000
0	0	.5	1.0000	1.0000	0.6439	1.0000
0	0	1	1.0000	1.0000	0.9940	1.0000

Nominal size is 0.05