Journal of Econometrics 207 (2018) 92–113

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

A robust test for network generated dependence ☆

Xiaodong Liu [a], Ingmar R. Prucha [b,*]

[a] Department of Economics, University of Colorado Boulder, Boulder, CO 80309, USA
[b] Department of Economics, University of Maryland, College Park, MD 20742, USA

ARTICLE INFO

Article history:
Received 30 June 2016
Received in revised form 25 April 2018
Accepted 21 May 2018
Available online 23 July 2018

JEL classification:
C12
C21

Keywords:
Test for network dependence
Generalized Moran $\mathcal{I}$ test
LM test
Laplace approximation
Network endogeneity
Network specification
Robustness

ABSTRACT

The paper introduces a robust testing procedure for network generated cross sectional dependence in the endogenous variables, exogenous variables and/or disturbances. Empirical researchers often face situations where they are unsure about how to model the proximity between cross sectional units in a network. The tests considered provide the empirical researcher an important degree of robustness in such situations. They generalize the Moran (1950) I test for dependence in spatial networks. The asymptotic properties of the tests are established under general conditions. The paper also discusses the use of the test statistics in situations where the network topology is endogenous.

1. Introduction

The paper introduces a robust test for network generated cross sectional dependence, and derives the statistical properties of the test. Empirical researchers often face situations where they are unsure about how to model the proximity between cross sectional units in a network. The test considered in this paper is aimed at providing to the empirical researcher an important degree of robustness in such situations.

As remarked by Kolaczyk (2009), "... 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. The importance of network dependencies has, in particular, been recognized early in the regional science, urban economics and geography literature. The focus of this literature is on spatial networks. An important class of spatial network models was introduced by Cliff and Ord (1973, 1981), where, as a formal modeling device, weight matrices are used to capture the existence and directional importance of links in a spatial network. It is important to note that in the Cliff–Ord type models the weights are only viewed as related to a measure of proximity between units,

☆ We thank the editor Oliver Linton, the associate editor, and three anonymous referees for helpful comments and suggestions. We would also like to thank the participants of the 26th EC2 Conference on Theory and Practice of Spatial Econometrics, Edinburgh, the Western Regional Association 56th Meeting Annual Meeting, Santa Fe, the XIth World Congress of the Spatial Econometrics Association, Singapore, and participants at seminars at the ETH Zuerich and the London School of Economics for their helpful comments.

* Corresponding author.
E-mail addresses: xiaodong.liu@colorado.edu (X. Liu), prucha@econ.umd.edu (I.R. Prucha).

https://doi.org/10.1016/j.jeconom.2018.05.005
0304-4076/

A Robust Test for Network Generated Dependence*

Xiaodong Liu† Ingmar R. Prucha‡

April 25, 2018
Forthcoming in *Journal of Econometrics*

Abstract

The paper introduces a robust testing procedure for network generated cross sectional dependence in the endogenous variables, exogenous variables and/or disturbances. Empirical researchers often face situations where they are unsure about how to model the proximity between cross sectional units in a network. The tests considered provide the empirical researcher an important degree of robustness in such situations. They generalize the Moran (1950) I test for dependence in spatial networks. The asymptotic properties of the tests are established under general conditions. The paper also discusses the use of the test statistics in situations where the network topology is endogenous.

JEL Classification: C12; C21

Key Words: Test for network dependence; generalized Moran $\mathcal{I}$ test, LM test, Laplace approximation, network endogeneity, network specification, robustness.

*We thank the editor Oliver Linton, the associate editor, and three anonymous referees for helpful comments and suggestions. We would also like to thank the participants of the 26th EC2 Conference on Theory and Practice of Spatial Econometrics, Edinburgh, the Western Regional Association 56th Meeting Annual Meeting, Santa Fe, the XIth World Congress of the Spatial Econometrics Association, Singapore, and participants at seminars at the ETH Zuerich and the London School of Economics for their helpful comments.

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

‡Corresponding author. Department of Economics, University of Maryland, College Park, MD 20742, USA. Tel.: 301-405-3499. Fax: 301-405-3542. E-mail: prucha@econ.umd.edu.

1 Introduction

The paper introduces a robust test for network generated cross sectional dependence, and derives the statistical properties of the test. Empirical researchers often face situations where they are unsure about how to model the proximity between cross sectional units in a network. The test considered in this paper is aimed at providing to the empirical researcher an important degree of robustness in such situations.

As remarked by Kolaczyk (2009), "... 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. The importance of network dependencies has, in particular, been recognized early in the regional science, urban economics and geography literature. The focus of this literature is on spatial networks. An important class of spatial network models was introduced by Cliff and Ord (1973, 1981), where, as a formal modeling device, weight matrices are used to capture the existence and directional importance of links in a spatial network. It is important to note that in the Cliff-Ord type models the weights are only viewed as related to a measure of proximity between units, but not necessarily to the geographic location of the units.[1] Thus by extending the notion of proximity from geographical proximity to economic proximity, technological proximity, social proximity, etc., these models are useful for a much wider class of applications with cross sectional interactions. This includes social interaction models as discussed in, e.g., Bramoullé, Djebbari and Fortin (2009), Lee, Liu and Lin (2010), Liu and Lee (2010), and Kuersteiner and Prucha (2015). For instance, a simple social interaction model can be specified as

$$y_i = \lambda \sum_{j=1}^n w_{ij} y_j + \beta x_i + \gamma \sum_{j=1}^n w_{ij} x_j + u_i \qquad \text{and} \qquad u_i = \varepsilon_i + \rho \sum_{j=1}^n w_{ij} \varepsilon_j, \tag{1}$$

where y_i, x_i and ε_i represent, respectively, the outcome, observed exogenous characteristic, and unobserved individual heterogeneity of cross sectional unit i. The weight w_{ij} captures social proximity of i and j in the network. Suppose $w_{ij} = n_i^{-1}$ if i and j are friends and $w_{ij} = 0$ otherwise, where n_i denotes the number of friends of i. Then, using the terminology in Manski (1993), $\sum_{j=1}^n w_{ij} y_j$ is the average outcome of i's friends with the coefficient λ representing *the endogenous peer effect*, $\sum_{j=1}^n w_{ij} x_j$ is the average of the observable characteristics of i's friends with the coefficient γ representing *the contextual effect*, and $\sum_{j=1}^n w_{ij} \varepsilon_j$ is the average of the unobservable characteristics of i's friends with the coefficient ρ representing *the correlated effect*. Of course, the above model also covers the simple group-average model as a special case with $w_{ij} = (n_g - 1)^{-1}$ if i and j belong to the same group of size n_g and $w_{ij} = 0$ otherwise; compare, e.g., Lee (2007), Davezies, D'Haultfoeuille and Fougère, (2009), and Carrell, Sacerdote and West (2013).

[1] This is in contrast to the literature on spatial random fields where units are indexed by location; see, e.g., Conley (1999) and Jenish and Prucha (2009, 2012) for contributions to the spatial econometrics literature.

In the spatial network literature one of the most widely used tests for cross sectional dependence is the Moran (1950) $\mathcal{I}$ test. This test statistic is formulated in terms of a normalized quadratic form of the variables to be tested for spatial dependence. Moran's original formulation assumed that the variables are observed and based the quadratic form on a weight matrix with zero or one elements, depending on whether or not two units were considered neighbors. Cliff and Ord (1973, 1981) considered testing for spatial dependence in the disturbance process of a classical linear regression model, and generalized the test statistic to a quadratic form of ordinary least squares residuals, allowing for general weight matrices.[2] They derived the finite sample moments of the test statistic under the assumption of normality. 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. He also discusses its close conceptual connection to the Durbin-Watson test statistic in the time series literature. King (1980, 1981) demonstrated that the Moran 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 optimality properties of the Moran I test is given in Hillier and Martellosio (2018), including a discussion of conditions under which the Moran I test is a Uniformly Most Powerful Invariant test. Kelejian and Prucha (2001) introduced a central limit theorem (CLT) for linear-quadratic forms, and used that result to establish the limiting distribution of the Moran $\mathcal{I}$ test statistic as $N(0, 1)$ under a fairly general set of assumptions. They allowed for heteroskedasticity, which facilitates, among other things, applications to models with limited dependent variables, and they introduced necessary modifications for the Moran $\mathcal{I}$ test statistic to accommodate endogenous regressors. Pinkse (1998, 2004) also considered Moran $\mathcal{I}$ flavored tests, including applications to discrete choice models.

Anselin (1988), Anselin and Rey (1991) and Anselin, Bera, Florax and Yoon (1996) considered LM and modified LM tests for spatial autoregressive model with spatially autoregressive disturbances, and provide extensive Monte Carlo results on their small sample properties. Baltagi and Li (2000), Baltagi, Song and Koh (2003), and Baltagi, Song Jung and Koh (2007) derived LM test for first order spatial panel data models, and also analyzed their small sample behavior based on an extensive Monte Carlo study.[3] For cross sectional data Born and Breitung (2011) considered LM tests for first order spatial dependence in the dependent variable and the disturbances, allowing for unknown heteroskedasticity. Baltagi and Yang (2013) considered small sample improved LM tests, and Yang (2015) provided a bootstrap refinement. Robinson and Rossi (2014) introduced improved LM tests based on an Edgeworth expansion.[4]

[2] Of course, in the absence of regressors this setup included the original Moran $\mathcal{I}$ test statistic as a special case.

[3] Pesaran (2004), Pesaran, Ulla and Yagamata (2008) and Baltagi, Feng and Kao (2012) also considered LM flavored tests for cross sectional dependence for panel data. Those tests are based on sample correlations and not specifically geared towards network generated dependence.

[4] Of course, there is also a large literature on ML and GMM estimation of Cliff-Ord type spatial models, which allows for likelihood ratio and Wald-type testing of the significance of spatial autoregressive parameters. Robinson and

One problem in using the Moran $\mathcal{I}$ and available LM tests for spatial models is that researchers are often unsure about how to specify the weight matrix employed by the test. Take a spatial network as an example, the weight w_{ij} could be a binary indicator variable depending on whether or not i and j are neighbors, or w_{ij} could be specified as the inverse of the geographical distance between i and j, etc. Researchers may consequently adopt a sequential testing procedure based on different specifications of the weight matrix. The sequential testing procedure, however, raises issues regarding the overall significance level of the test. Motivated by this problem we define in this paper a single test statistic, which in its simplest form is defined as a weighted inner product of a vector of quadratic forms, with each quadratic form corresponding to a different weight matrix. In this sense the test statistic combines a set of Moran $\mathcal{I}$ tests into a single test.

Our generalizations of the Moran $\mathcal{I}$ test differentiates between two uses. The first generalization is geared towards testing for cross sectional dependence in the disturbance process. We refer to the corresponding test statistic as the $\mathcal{I}_u^2$ test statistic. The second generalization is geared towards testing for cross sectional dependence in the dependent variable, which may be due to the dependence on the outcomes, observed exogenous characteristics, and/or unobserved characteristics of other units in the network. We refer to the corresponding test statistic as the $\mathcal{I}_y^2$ test statistic. We establish the limiting distributions 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. We also show that if the data generating process (DGP) under the alternative hypothesis is of the form of a higher order spatial autoregressive and/or spatial moving average process, then the generalized Moran $\mathcal{I}$ test statistics can be viewed as LM test statistics. Since our test statistics contain the Moran $\mathcal{I}$ test statistic as a special case, this result generalizes the findings of Burridge (1980).

Although the generalized Moran $\mathcal{I}$ test can be interpreted as a LM test if the model under the alternative hypothesis has a certain specification, the validity and implementation of the Moran $\mathcal{I}$ test does not rely on a specific model under the alternative hypothesis. By its construction the generalized Moran $\mathcal{I}$ test should be useful for detecting network generated dependence in wide range of situations. This includes, as we elaborate later in this paper, situations where weight matrix representing the network topology is misspecified and/or endogenous.

The paper also generalizes Lieberman's (1994) Laplace approximation of the expected values of powers of ratios of quadratic forms to those of ratios of products of powers of linear-quadratic forms over powers of quadratic forms. This generalization is then used to develop certain small sample standardized versions of the generalized Moran $\mathcal{I}$ test statistics.

The paper is organized as follows: In Section 2 we introduce and give an intuitive motivation for the generalized Moran $\mathcal{I}$ test statistics within a simplified setup. We also develop small sample

Rossi (2015) introduced an alternative test based on the biased OLS estimator for a first order spatial autoregressive model, using Edgeworth expansions and the bootstrap to appropriately adjust the size of the test.

standardized versions of those test statistics. In Section 3 we consider a generalized setup, which allows for both endogeneity and unknown heteroskedasticity, and we establish the limiting distribution of our generalized Moran $\mathcal{I}$ test statistics for this setup. In Section 4 we report on a Monte Carlo Study, which explores the small sample properties of our test statistics.[5] This includes results on the properties of the test statistics in situations where the network topology is endogenous. The Monte Carlo results are encouraging and suggest that the generalized Moran $\mathcal{I}$ tests perform well in a wide range of situations. Concluding remarks are given in the Section 5. All technical details are relegated to an appendix, and a supplementary appendix, which will be made available online. Throughout the paper we adopt the following notation. For a square matrix $\mathbf{A}$, let $\overline{\mathbf{A}} = (\mathbf{A}+\mathbf{A}')/2$.

2 Test Statistics and Motivation

In the following we introduce and provide motivations for two generalizations of the Moran $\mathcal{I}$ test statistic. For simplicity of exposition we assume in this section that under the null hypothesis the data are generated by a classical linear regression model. This setup will be generalized in the subsequent sections. The two generalizations of the Moran $\mathcal{I}$ test statistic will be referred to as the $\mathcal{I}_u^2(q)$ test statistic and the $\mathcal{I}_y^2(q)$ test statistic, respectively. The first test statistic is geared towards testing for cross sectional dependence in the unobserved disturbance process, while the second is geared towards testing for cross sectional dependence in the observed dependent variable. The Moran $\mathcal{I}$ test statistic is covered as a special case in that the statistics are such that $\mathcal{I}^2 = \mathcal{I}_u^2(1)$.

The test statistics are first motivated by considering a situation where the researcher is unsure about which weight matrix in a set of weight matrices would best represent the structure of the network under the alternative. We then provide a more formal motivation for the $\mathcal{I}_u^2(q)$ and $\mathcal{I}_y^2(q)$ test statistics by showing that both test statistics can be derived as LM test statistics within the context of Cliff-Ord type spatial processes.

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

Towards giving an intuitive introduction to the $\mathcal{I}_u^2(q)$ test statistic, suppose a set of cross sectional data is generated by the linear regression model[6]

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \tag{2}$$

where $\mathbf{y} = [y_1, \cdots, y_n]'$ is an $n \times 1$ vector of observations on the dependent variable, $\mathbf{X} = [x_{ik}]$ is an $n \times K_x$ matrix of observations on K_x nonstochastic exogenous variables, and $\mathbf{u} = [u_1, \cdots, u_n]'$ is an

[5] This study expands on some early small sample results given in Drukker and Prucha (2013).

[6] Although we do not indicate this explicitly in our notation, here and in the following all variables are allowed to depend on the sample size, i.e., are allowed to formulate triangular arrays.

$n \times 1$ vector of regression disturbances, which for the moment are assumed to be distributed with mean zero and variance σ^2.[7] The researcher wants to test the hypothesis that the disturbances are cross sectionally uncorrelated, i.e.,

$$H_0^u : \text{cov}(\mathbf{u}) = \sigma^2 \mathbf{I}_n,$$

against the alternative H_1^u that the disturbances are cross sectionally correlated.

As discussed, Cliff and Ord (1973, 1981) introduced an important class of models for spatial networks. Towards motivating our test statistic we follow that literature and keep track of links between cross sectional units in a network by an $n \times n$ weight matrix $\mathbf{W} = [w_{ij}]$, with its elements accounting for the relative directional strength of the links. If no direct link exists between units i and j we have $w_{ij} = 0$. Furthermore, in this literature $\mathbf{W}$ is typically normalized by setting $w_{ii} = 0$. Clearly under H_0^u we then have $\text{E}(\mathbf{u}'\mathbf{W}\mathbf{u}) = \sigma^2 \text{tr}(\mathbf{W}) = 0$, while under H_1^u we generally have $\text{E}(\mathbf{u}'\mathbf{W}\mathbf{u}) = \text{tr}[\mathbf{W}\text{E}(\mathbf{u}\mathbf{u}')] \neq 0$. This motivates the following standard Moran $\mathcal{I}$ test statistic for testing that the disturbance process in (2) satisfies H_0^u:

$$\mathcal{I} = \frac{\widetilde{\mathbf{u}}'\mathbf{W}\widetilde{\mathbf{u}}}{[2\widetilde{\sigma}^4 \text{tr}(\overline{\mathbf{W}}^2)]^{1/2}} \tag{3}$$

where $\widetilde{\mathbf{u}} = \mathbf{y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}$ with $\widetilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ denotes the ordinary least squares (OLS) residuals, and $\widetilde{\sigma}^2 = n^{-1}\widetilde{\mathbf{u}}'\widetilde{\mathbf{u}}$ is the corresponding estimator for σ^2. Kelejian and Prucha (2001) establish under a fairly general set of regularity conditions that under H_0^u we have $\mathcal{I} \xrightarrow{d} N(0,1)$, and thus $\mathcal{I}^2 \xrightarrow{d} \chi^2(1)$.

We emphasize that the weights given to respective links in $\mathbf{W}$ are generally considered to be reflective of some measure of proximity between units, but do not depend on an explicit indexing of units by location. By extending the notion of proximity from geographical proximity to economic 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.

The above introduction of the Moran $\mathcal{I}$ test statistic was intuitive and did not specify a particular form for the disturbance process. Now suppose that, under the alternative hypothesis H_1^u, the disturbance process is either a first-order spatial autoregressive, i.e., $\mathbf{u} = \rho\mathbf{W}\mathbf{u} + \boldsymbol{\varepsilon}$, or a first-order spatial moving-average process, i.e., $\mathbf{u} = \boldsymbol{\varepsilon} + \rho\mathbf{W}\boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon} = [\varepsilon_1, \cdots, \varepsilon_n]'$ is an $n \times 1$ vector of independent zero mean innovations. Burridge (1980) provided a formal motivation for the Moran $\mathcal{I}$ test by establishing that for both forms of the disturbance process $\mathcal{I}^2$ is identical to the LM test statistic based on a Gaussian likelihood for testing $\rho = 0$.

One practical problem with the Moran $\mathcal{I}$ test statistic is that empirical researchers are often unsure about the specification of $\mathbf{W}$. Thus it is of interest to consider a generalized Moran $\mathcal{I}$ test for situations where the researcher is not sure whether $\mathbf{W}_1, \mathbf{W}_2, \cdots$, or $\mathbf{W}_q$ or some linear combination

[7] The assumption of homoskedasticity will be relaxed later on.

of those matrices properly model the network topology. Towards introducing such a generalization, let

$$\widetilde{\mathbf{V}}^U = \begin{bmatrix} \widetilde{\mathbf{u}}'\mathbf{W}_1\widetilde{\mathbf{u}} \\ \vdots \\ \widetilde{\mathbf{u}}'\mathbf{W}_q\widetilde{\mathbf{u}} \end{bmatrix} \quad \text{and} \quad \widetilde{\mathbf{\Phi}}^{UU} = \begin{bmatrix} 2\widetilde{\sigma}^4\mathrm{tr}(\overline{\mathbf{W}}_1\overline{\mathbf{W}}_1) & \cdots & 2\widetilde{\sigma}^4\mathrm{tr}(\overline{\mathbf{W}}_1\overline{\mathbf{W}}_q) \\ \vdots & \ddots & \vdots \\ 2\widetilde{\sigma}^4\mathrm{tr}(\overline{\mathbf{W}}_q\overline{\mathbf{W}}_1) & \cdots & 2\widetilde{\sigma}^4\mathrm{tr}(\overline{\mathbf{W}}_q\overline{\mathbf{W}}_q) \end{bmatrix}.$$

The matrix $n^{-1}\widetilde{\mathbf{\Phi}}^{UU}$ will be seen to be a consistent estimator for the variance-covariance (VC) matrix of the limiting distribution of $n^{-1/2}\widetilde{\mathbf{V}}^U$. The researcher could now test H_0^u using the test statistic:

$$\mathcal{I}_u^2(q) = \widetilde{\mathbf{V}}^{U\prime}(\widetilde{\mathbf{\Phi}}^{UU})^{-1}\widetilde{\mathbf{V}}^U. \tag{4}$$

The above statistic generalizes the (squared) Moran $\mathcal{I}$ test statistic. It may be viewed as combining q Moran $\mathcal{I}$ tests 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. Of course, for $q = 1$ the $\mathcal{I}_u^2(q)$ test delivers the Moran $\mathcal{I}$ test as a special case.

In the following we will show that under the above assumptions (and some further regularity conditions), we have $\mathcal{I}_u^2(q) \xrightarrow{d} \chi^2(q)$ assuming that H_0^u is true. We note that this result remains true even if $\mathbf{X}$ includes the spatial lags of some exogenous variables.

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

In the following we introduce the $\mathcal{I}_y^2(q)$ test statistic. This statistic is a further generalization of the Moran $\mathcal{I}$ test statistic. It is geared towards situations where the researcher is interested in testing for network generated dependence in the dependent variable $\mathbf{y}$. Such dependence could stem from spillovers or interactions in the dependent variable, exogenous variables and/or the disturbances between cross-sectional units.

Towards motivating the $\mathcal{I}_y^2(q)$ test statistic, suppose again that a set of cross sectional data is generated by the linear regression model (2). However, we assume now that the researcher wants to test the more general hypotheses that (i) the mean of the dependent variable of the i-th unit only depends on exogenous variables specific to the i-th unit, and thus is not affected by changes in the exogenous variables of the other units, and (ii) the dependent variable is uncorrelated across units. That is, the researcher wants to test

$$H_0^y : \mathrm{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} \text{ and } \mathrm{cov}(\mathbf{y}) = \sigma^2\mathbf{I}_n, \tag{5}$$

where $\mathbf{X} = [x_{ik}]$ is a non-stochastic matrix with $\partial x_{ik}/\partial x_{jl} = 0$ for $i \neq j$ (i.e. $\mathbf{X}$ does not include spatial lags), against the alternative H_1^y that H_0^y is false.[8]

[8] The linear dependence of $\mathrm{E}(y_i)$ on $x_{i1}, \ldots, x_{iK}$ is only maintained for ease of exposition, and the assumption could be extended to allow for $\mathrm{E}(y_i)$ to depend nonlinearly on $x_{i1}, \ldots, x_{iK}$.

Towards motivating the $\mathcal{I}_y^2(q)$ test statistic, assume furthermore for a moment that under the alternative hypothesis the set of cross sectional data is generated by

$$\mathbf{y} = \lambda\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{X}\boldsymbol{\gamma} + \mathbf{u}. \tag{6}$$

Clearly, model (6) allows for potential network generated dependence in the endogenous and exogenous variables via the spatial lags $\mathbf{Wy}$ and $\mathbf{WX}$, in addition to potential network generated dependence via the disturbance process. The reduced form of model (6) is given by

$$\mathbf{y} = (\mathbf{I}_n - \lambda\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta} + (\mathbf{I}_n - \lambda\mathbf{W})^{-1}\mathbf{W}\mathbf{X}\boldsymbol{\gamma} + (\mathbf{I}_n - \lambda\mathbf{W})^{-1}\mathbf{u}. \tag{7}$$

The null hypothesis H_0^y specified in (5) will typically only hold if $H_0^u : \text{cov}(\mathbf{u}) = \sigma^2\mathbf{I}_n$ holds. Clearly, if H_0^u holds, then a test for H_0^y is equivalent to a test for $\lambda = 0$ and $\boldsymbol{\gamma} = \mathbf{0}$. Observe that under H_0^y we have $\mathbf{y} - \mathbf{X}\boldsymbol{\beta} = \mathbf{u}$ and $\text{E}[(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{W}\mathbf{y}] = \sigma^2\text{tr}(\mathbf{W}) = 0$ and $\text{E}[(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{W}\mathbf{X}] = \mathbf{0}$. On the other hand, under H_1^y we have $\mathbf{y} - \mathbf{X}\boldsymbol{\beta} = \lambda\mathbf{W}(\mathbf{I}_n - \lambda\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta} + (\mathbf{I}_n - \lambda\mathbf{W})^{-1}\mathbf{W}\mathbf{X}\boldsymbol{\gamma} + (\mathbf{I}_n - \lambda\mathbf{W})^{-1}\mathbf{u}$ and thus $\text{E}[(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{W}\mathbf{y}] \neq 0$ and $\text{E}[(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{W}\mathbf{X}] \neq \mathbf{0}$ if either $\lambda \neq 0$ or $\boldsymbol{\gamma} \neq \mathbf{0}$. This suggests that in constructing a test statistic for H_0^y we should also consider $\mathbf{u}'\mathbf{W}\mathbf{y}$ and $\mathbf{u}'\mathbf{W}\mathbf{X}$ in addition to $\mathbf{u}'\mathbf{W}\mathbf{u}$.

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 $\mathbf{W}_1, \mathbf{W}_2, \cdots$, or $\mathbf{W}_q$ or some linear combination of those matrices properly model the network topology. Let $\widehat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}$ with $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ denote the residuals of the restricted OLS estimator, and let $\widehat{\sigma}^2 = n^{-1}\widehat{\mathbf{u}}'\widehat{\mathbf{u}}$ denote the corresponding estimator for σ^2. Let $\widehat{\mathbf{V}}^Y = [\widehat{\mathbf{u}}'\mathbf{W}_1\mathbf{y}, \cdots, \widehat{\mathbf{u}}'\mathbf{W}_q\mathbf{y}]'$, $\widehat{\mathbf{V}}^X = [\widehat{\mathbf{u}}'\mathbf{W}_1\mathbf{X}, \cdots, \widehat{\mathbf{u}}'\mathbf{W}_q\mathbf{X}]'$, and $\widehat{\mathbf{V}}^U = [\widehat{\mathbf{u}}'\mathbf{W}_1\widehat{\mathbf{u}}, \cdots, \widehat{\mathbf{u}}'\mathbf{W}_q\widehat{\mathbf{u}}]'$, and define

$$\widehat{\mathbf{V}} = \begin{bmatrix} \widehat{\mathbf{V}}^Y \\ \widehat{\mathbf{V}}^X \\ \widehat{\mathbf{V}}^U \end{bmatrix} \qquad \text{and} \qquad \widehat{\boldsymbol{\Phi}} = \begin{bmatrix} \widehat{\boldsymbol{\Phi}}^{YY} & \widehat{\boldsymbol{\Phi}}^{YX} & \widehat{\boldsymbol{\Phi}}^{YU} \\ (\widehat{\boldsymbol{\Phi}}^{YX})' & \widehat{\boldsymbol{\Phi}}^{XX} & \mathbf{0} \\ (\widehat{\boldsymbol{\Phi}}^{YU})' & \mathbf{0} & \widehat{\boldsymbol{\Phi}}^{UU} \end{bmatrix}, \tag{8}$$

where

$$\begin{aligned}
\widehat{\boldsymbol{\Phi}}^{YY} &= [2\widehat{\sigma}^4\text{tr}(\overline{\mathbf{W}}_r\overline{\mathbf{W}}_s) + \widehat{\sigma}^2\widehat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{W}_r'\mathbf{M}_X\mathbf{W}_s\mathbf{X}\widehat{\boldsymbol{\beta}}]_{r,s=1,\cdots,q}, \\
\widehat{\boldsymbol{\Phi}}^{YX} &= [\widehat{\sigma}^2\widehat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{W}_r'\mathbf{M}_X\mathbf{W}_s\mathbf{X}]_{r,s=1,\cdots,q}, \\
\widehat{\boldsymbol{\Phi}}^{YU} &= [2\widehat{\sigma}^4\text{tr}(\overline{\mathbf{W}}_r\overline{\mathbf{W}}_s)]_{r,s=1,\cdots,q}, \\
\widehat{\boldsymbol{\Phi}}^{XX} &= [\widehat{\sigma}^2\mathbf{X}'\mathbf{W}_r'\mathbf{M}_X\mathbf{W}_s\mathbf{X}]_{r,s=1,\cdots,q}, \\
\widehat{\boldsymbol{\Phi}}^{UU} &= [2\widehat{\sigma}^4\text{tr}(\overline{\mathbf{W}}_r\overline{\mathbf{W}}_s)]_{r,s=1,\cdots,q},
\end{aligned}$$

with $\mathbf{M}_X = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Under certain regularity conditions, $n^{-1}\widehat{\boldsymbol{\Phi}}$ can be shown to be a consistent estimator for the limiting VC matrix of $n^{-1/2}\widehat{\mathbf{V}}$.

As $\mathbf{y} = \mathbf{X}\widehat{\boldsymbol{\beta}} + \widehat{\mathbf{u}}$, $\widehat{\mathbf{u}}'\mathbf{W}_r\mathbf{y} = \widehat{\mathbf{u}}'\mathbf{W}_r\mathbf{X}\widehat{\boldsymbol{\beta}} + \widehat{\mathbf{u}}'\mathbf{W}_r\widehat{\mathbf{u}}$ for $r = 1, \cdots, q$, and, consequently, $n^{-1}\widehat{\boldsymbol{\Phi}}$ is singular. To avoid the singularity problem of the VC matrix, we formulate the test statistic for H_0^y as

$$\mathcal{I}_y^2(q) = (\mathrm{L}\widehat{\mathbf{V}})'(\mathrm{L}\widehat{\boldsymbol{\Phi}}\mathrm{L}')^{-1}(\mathrm{L}\widehat{\mathbf{V}}). \tag{9}$$

where L is a selector matrix of rank $(K_x + 1)q$, consisting of rows of a conformable identity matrix, that ensures $\mathrm{L}\widehat{\boldsymbol{\Phi}}\mathrm{L}'$ is nonsingular. The selector matrix L is not unique. To fix ideas, we consider the selector matrix

$$\mathrm{L} = \begin{bmatrix} \mathbf{0}_{q\times q} & \mathbf{I}_{K_x q} & \mathbf{0}_{q\times q} \\ \mathbf{0}_{q\times q} & \mathbf{0}_{q\times K_x q} & \mathbf{I}_q \end{bmatrix}. \tag{10}$$

The following proposition shows that the test statistic defined in (9) with the selector matrix given by (10) is numerically equivalent to a test statistic based on $\widehat{\mathbf{V}}$ and a generalized inverse of $\widehat{\boldsymbol{\Phi}}$.

Proposition 1. *Let* L *be a selector matrix given by (10). Then,*

$$\mathcal{I}_y^2(q) = (\mathrm{L}\widehat{\mathbf{V}})'(\mathrm{L}\widehat{\boldsymbol{\Phi}}\mathrm{L}')^{-1}(\mathrm{L}\widehat{\mathbf{V}}) = \widehat{\mathbf{V}}'\widehat{\boldsymbol{\Phi}}^{+}\widehat{\mathbf{V}},$$

where $\widehat{\boldsymbol{\Phi}}^{+}$ denotes the Moore-Penrose generalized inverse of $\widehat{\boldsymbol{\Phi}}$.

The selector matrix (10) eliminates $\widehat{\mathbf{u}}'\mathbf{W}_1\mathbf{y}, \cdots, \widehat{\mathbf{u}}'\mathbf{W}_q\mathbf{y}$ from $\widehat{\mathbf{V}}$. Alternatively we could eliminate $\widehat{\mathbf{u}}'\mathbf{W}_1\mathbf{x}_k, \cdots, \widehat{\mathbf{u}}'\mathbf{W}_q\mathbf{x}_k$, where $\mathbf{x}_k$ denotes the k-th column of $\mathbf{X}$, or $\widehat{\mathbf{u}}'\mathbf{W}_1\widehat{\mathbf{u}}, \cdots, \widehat{\mathbf{u}}'\mathbf{W}_q\widehat{\mathbf{u}}$ from $\widehat{\mathbf{V}}$, which yields the same numerical value for $\mathcal{I}_y^2(q)$. In the following we will show that under the above assumptions (and some further regularity conditions), and assuming that H_0^y is true, we have $\mathcal{I}_y^2(q) \xrightarrow{d} \chi^2(\mathrm{rank}(\mathrm{L}))$.

Remark 1. Pötscher (1985) encountered a similar problem using LM tests for the orders of time series ARMA models. He defined his LM test statistic in terms of the linearly dependent score vector and a generalized inverse of its VC matrix. Proposition 1 shows that our test statistic for H_0^y is in line with the LM test statistic in Pötscher (1985).

Remark 2. The selector matrix L given by (10) has rank $(K_x + 1)q$. The test statistic $\mathcal{I}_y^2(q)$ defined in (9) can be generalized by considering some general selector matrix L of rank less than $(K_x + 1)q$. Thus, our test will be applicable to a wide range of hypotheses. For example, suppose a researcher wants to test for the absence of spillovers in the dependent variable, the disturbances, and a subset of the exogenous variables $\mathbf{X}_\star = \mathbf{X}\mathrm{L}_\star'$, where $\mathrm{L}_\star$ is a selector matrix consisting of rows of the identity matrix $\mathbf{I}_{K_x}$. That is, under the alternative hypothesis, the set of cross sectional data is assumed to be generated by $\mathbf{y} = \lambda\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{X}_\star\boldsymbol{\gamma}_\star + \mathbf{u}$. Then, the researcher would consider a test statistic defined in (9) with $\mathrm{L} = \mathrm{diag}(\mathbf{I}_q, \mathbf{I}_q \otimes \mathrm{L}_\star, \mathbf{I}_q)$.

Remark 3. In this subsection, $\widehat{\mathbf{V}}^U$ and $\widehat{\mathbf{\Phi}}^{UU}$ appearing in the $\mathcal{I}_y^2(q)$ test statistic are defined in the same way as $\widetilde{\mathbf{V}}^U$ and $\widetilde{\mathbf{\Phi}}^{UU}$ appearing in the $\mathcal{I}_u^2(q)$ test statistic. However, we emphasize that, as shown below, for more general models, where some of the regressors under the null hypothesis are endogenous, the VC matrices for $\widetilde{\mathbf{V}}^U$ and $\widehat{\mathbf{V}}^U$ employed respectively by the $\mathcal{I}_u^2(q)$ and $\mathcal{I}_y^2(q)$ test statistics could be different.

2.3 Equivalence of the Generalized Moran $\mathcal{I}$ Tests and LM Tests

The above motivations for the $\mathcal{I}_u^2(q)$ and $\mathcal{I}_y^2(q)$ test statistics are intuitive. In this subsection we also provide a more formal motivation by showing that both test statistics can be established as LM test statistics within the context of Cliff-Ord type spatial processes. We continue utilizing the notation defined above. In particular, we assume that under the null hypotheses H_0^u and H_0^y the data are generated by the linear regression model (2) with $\mathbf{u} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$.

First, suppose that under the alternative hypothesis the data are generated by a linear regression model, where the disturbances follow a spatial autoregressive process of order q, for short, a SAR(q) process, i.e.,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \qquad \mathbf{u} = \textstyle\sum_{r=1}^{q} \rho_r \mathbf{W}_r \mathbf{u} + \boldsymbol{\varepsilon}, \tag{11}$$

with $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$. The regressor matrix $\mathbf{X}$ and the weight matrices $\mathbf{W}_1, \cdots, \mathbf{W}_q$ are assumed to be nonstochastic, $\mathbf{X}$ is assumed to have full column rank, and $\mathbf{I}_n - \sum_{r=1}^{q} \rho_r \mathbf{W}_r$ is assumed to be nonsingular. Clearly under this setup the null hypothesis $H_0^u : \mathrm{cov}(\mathbf{u}) = \sigma^2 \mathbf{I}_n$ is equivalent to $H_0^y : \mathrm{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ and $\mathrm{cov}(\mathbf{y}) = \sigma^2 \mathbf{I}_n$, and can be formulated equivalently as $H_0 : \boldsymbol{\rho} = \mathbf{0}$, where $\boldsymbol{\rho} = [\rho_1, \cdots, \rho_q]'$. The following proposition establishes the equivalence of the $\mathcal{I}_u^2(q)$ and LM test for model (11).

Proposition 2. *Suppose the above stated assumptions hold for model (11) and $\widetilde{\mathbf{\Phi}}^{UU}$ is nonsingular. Let LM_u denote the LM test statistic for $H_0 : \boldsymbol{\rho} = \mathbf{0}$. Then $\mathrm{LM}_u = \mathcal{I}_u^2(q)$.*

Next, suppose that under the alternative hypothesis the data generating process is more general, and modeled as a SAR(q) process in $\mathbf{y}$ and $\mathbf{u}$, and with spatial lags in $\mathbf{X}$, i.e.,

$$\mathbf{y} = \textstyle\sum_{r=1}^{q} \lambda_r \mathbf{W}_r \mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \sum_{r=1}^{q} \mathbf{W}_r \mathbf{X} \boldsymbol{\gamma}_r + \mathbf{u}, \qquad \mathbf{u} = \sum_{r=1}^{q} \rho_r \mathbf{W}_r \mathbf{u} + \boldsymbol{\varepsilon}. \tag{12}$$

We continue to maintain the assumptions stated above, and furthermore assume that $\mathbf{I}_n - \sum_{r=1}^{q} \lambda_r \mathbf{W}_r$ is nonsingular. Under this setup the null hypothesis, $H_0^y : \mathrm{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ and $\mathrm{cov}(\mathbf{y}) = \sigma^2 \mathbf{I}_n$, can be formulated equivalently as $H_0 : \boldsymbol{\lambda} = \mathbf{0}, \boldsymbol{\rho} = \mathbf{0}, \boldsymbol{\gamma}_1 = \cdots = \boldsymbol{\gamma}_q = \mathbf{0}$, where $\boldsymbol{\lambda} = [\lambda_1, \cdots, \lambda_q]'$ and $\boldsymbol{\rho} = [\rho_1, \cdots, \rho_q]'$. The following proposition establishes the equivalence of the $\mathcal{I}_y^2(q)$ and LM test for model (12).

Proposition 3. *Suppose the above stated assumptions hold for model (12) and* $\mathrm{L}\widehat{\mathbf{\Phi}}\mathrm{L}'$*, with* L *defined in (10), is nonsingular. Let* LM_y *denote the LM test statistic for* $H_0 : \boldsymbol{\lambda} = \mathbf{0}, \boldsymbol{\rho} = \mathbf{0}, \boldsymbol{\gamma}_1 = \cdots = \boldsymbol{\gamma}_q = \mathbf{0}$*. Then* $\mathrm{LM}_y = \mathcal{I}_y^2(q)$*.*

Remark 4. As shown in their proofs, Propositions 2 and 3 continue to hold if the disturbance process of models (11) and (12) is specified as a spatial moving average process of order q, or more generally, as a spatial autoregressive moving-average process of order $(\overline{q}, \underline{q})$, for short, a spatial ARMA$(\overline{q}, \underline{q})$ process, where $\mathbf{u} = \sum_{r=1}^{\overline{q}} \rho_r \mathbf{W}_r \mathbf{u} + \boldsymbol{\varepsilon} + \sum_{r=\overline{q}+1}^{q} \rho_r \mathbf{W}_r \boldsymbol{\varepsilon}$ and $q = \overline{q} + \underline{q}$. We note further that a generalized version of Proposition 3 also holds if we allow for different orders of spatial lags with different weight matrices for $\mathbf{y}$, $\mathbf{X}$ and $\mathbf{u}$ in model (12).

Remark 5. Although the above propositions establish that the $\mathcal{I}_u^2(q)$ and $\mathcal{I}_y^2(q)$ test statistics coincide with the LM test statistics when under the alternative the data generating process is defined by models (11) and (12), respectively, we emphasize that the validity and implementation of the generalized Moran $\mathcal{I}$ tests do not require the assumption of a specific data generating process under the alternative hypothesis.

2.4 Consistency of the Generalized Moran $\mathcal{I}$ Tests

We maintain the following assumption regarding the weight matrices $\mathbf{W}_1, \cdots, \mathbf{W}_q$ considered by the researcher to capture the proximity of cross sectional units in a network.

Assumption 1. $\mathbf{W}_1, \cdots, \mathbf{W}_q$ *are* $n \times n$ *nonstochastic weight matrices with zero diagonals and with their row and column sums bounded in absolute value, uniformly in* n*, by some finite constant.*

This boundedness assumption is quite standard in the literature, and satisfied for normalized weight matrices, which are typically used in empirical work; see, e.g., Kelejian and Prucha (2010) for further discussions. Next, suppose that under the alternative hypothesis the data are generated by (11), where the elements of $\boldsymbol{\varepsilon}$ are i.i.d. $(0, \sigma^2)$ innovations with finite $(4+\delta)$-th moments for some $\delta > 0$ and the weight matrices $\mathbf{W}_1, \cdots, \mathbf{W}_q$ satisfy Assumption 1. Observe that, in model (11), $\mathrm{E}(\mathbf{u}) = \mathbf{0}$ and $\mathbf{\Omega}_u = \mathrm{cov}(\mathbf{u}) = \sigma^2 \mathbf{R}^{-1} \mathbf{R}'^{-1}$, where $\mathbf{R} = \mathbf{I}_n - \sum_{r=1}^{q} \rho_r \mathbf{W}_r$. Assume that the row and column sums of $\mathbf{R}^{-1}$ are uniformly bounded in absolute value, and let $\overline{\sigma}^2 = \mathrm{E}\widetilde{\sigma}^2 + o(1) = n^{-1}\mathrm{tr}(\mathbf{\Omega}_u)$ with $0 < c_\sigma \leq \overline{\sigma}^2 \leq C_\sigma < \infty$ for some constants c_σ and C_σ. Assume further that the elements of $\mathbf{X}$ are uniformly bounded, and $\lim_{n\to\infty} n^{-1}\mathbf{X}'\mathbf{X}$ is finite and nonsingular. Let $\boldsymbol{\mu}^U = [\mathrm{tr}(\mathbf{W}_1 \mathbf{\Omega}_u), \cdots, \mathrm{tr}(\mathbf{W}_q \mathbf{\Omega}_u)]'$ denote the mean of $\widetilde{\mathbf{V}}^U$, and let $\mathbf{\Phi}^{UU} = [2\overline{\sigma}^4 \mathrm{tr}(\overline{\mathbf{W}}_r \overline{\mathbf{W}}_s)]_{r,s=1,\cdots,q}$ denote the non-stochastic counter part of $\widetilde{\mathbf{\Phi}}^{UU}$. Observe that under the alternative hypothesis in general $\boldsymbol{\mu}^U \neq \mathbf{0}$. More specifically, suppose that under the alternative hypothesis $0 < c_\mu \leq n^{-1} \left|\boldsymbol{\mu}^U\right| \leq C_\mu < \infty$, and $0 < c_\phi \leq n^{-1}\lambda_{\min}(\mathbf{\Phi}^{UU}) \leq n^{-1}\lambda_{\max}(\mathbf{\Phi}^{UU}) \leq C_\phi < \infty$ for some constants c_μ, C_μ, c_ϕ and C_ϕ, where $\lambda_{\min}(\mathbf{\Phi}^{UU})$ and $\lambda_{\max}(\mathbf{\Phi}^{UU})$ denote the smallest and largest eigenvalues of $\mathbf{\Phi}^{UU}$. The following proposition establishes the consistency of the $\mathcal{I}_u^2(q)$ test statistic defined in (4).

Proposition 4. *Suppose that the data are generated by model (11) and the above stated assumptions hold. Then, under the alternative hypothesis, we have* $\lim_{n\to\infty} \Pr(\mathcal{I}_u^2(q) \leq \gamma) = 0$ *for any* $\gamma > 0$.

Now, suppose that under the alternative hypothesis the data are generated by (12). We continue to maintain the assumptions stated above, and furthermore assume that $\mathbf{S} = \mathbf{I}_n - \sum_{r=1}^{q} \lambda_r \mathbf{W}_r$ is nonsingular and the row and column sums of $\mathbf{S}^{-1}$ are uniformly bounded in absolute value. Let $\mathbf{d} = \mathrm{E}\widehat{\mathbf{u}} = \mathbf{M}_X \mathbf{S}^{-1}(\mathbf{X}\boldsymbol{\beta} + \sum_{r=1}^{q} \mathbf{W}_r \mathbf{X}\boldsymbol{\gamma}_r)$ and let $\underline{\sigma}^2 = \mathrm{E}\widehat{\sigma}^2 + o(1) = n^{-1}\mathbf{d}'\mathbf{d} + n^{-1}\mathrm{tr}(\mathbf{S}'^{-1}\mathbf{S}^{-1}\mathbf{\Omega}_u)$, with $0 < \underline{c}_\sigma \leq \underline{\sigma}^2 \leq \underline{C}_\sigma < \infty$ for some constants $\underline{c}_\sigma$ and $\underline{C}_\sigma$. Let $\underline{\boldsymbol{\mu}} = [\underline{\boldsymbol{\mu}}^{X\prime}, \underline{\boldsymbol{\mu}}^{U\prime}]'$ and $\underline{\mathbf{\Phi}} = \mathrm{diag}(\underline{\mathbf{\Phi}}^{XX}, \underline{\mathbf{\Phi}}^{UU})$ where $\underline{\boldsymbol{\mu}}^X = [\mathbf{d}'\mathbf{W}_1\mathbf{X}, \cdots, \mathbf{d}'\mathbf{W}_q\mathbf{X}]'$ and $\underline{\boldsymbol{\mu}}^U = [\mathbf{d}'\mathbf{W}_1\mathbf{d} + \mathrm{tr}(\mathbf{S}'^{-1}\mathbf{W}_1\mathbf{S}^{-1}\mathbf{\Omega}_u), \cdots, \mathbf{d}'\mathbf{W}_q\mathbf{d} + \mathrm{tr}(\mathbf{S}'^{-1}\mathbf{W}_q\mathbf{S}^{-1}\mathbf{\Omega}_u)]'$ denote in essence the means of $\widehat{\mathbf{V}}^X$ and $\widehat{\mathbf{V}}^U$, and $\underline{\mathbf{\Phi}}^{XX} = [\underline{\sigma}^2 \mathbf{X}'\mathbf{W}_r'\mathbf{M}_X\mathbf{W}_s\mathbf{X}]_{r,s=1,\cdots,q}$ and $\underline{\mathbf{\Phi}}^{UU} = [2\underline{\sigma}^4 \mathrm{tr}(\overline{\mathbf{W}}_r \overline{\mathbf{W}}_s)]_{r,s=1,\cdots,q}$ denote the non-stochastic counter parts of $\widehat{\mathbf{\Phi}}^{XX}$ and $\widehat{\mathbf{\Phi}}^{UU}$, respectively. Observe that under the alternative hypothesis in general $\underline{\boldsymbol{\mu}} \neq \mathbf{0}$. More specifically, suppose that under the alternative hypothesis $0 < \underline{c}_\mu \leq n^{-1}\left|\underline{\boldsymbol{\mu}}\right| \leq \underline{C}_\mu < \infty$, and $0 < \underline{c}_\phi \leq n^{-1}\lambda_{\min}(\underline{\mathbf{\Phi}}) \leq n^{-1}\lambda_{\max}(\underline{\mathbf{\Phi}}) \leq \underline{C}_\phi < \infty$ for some constants $\underline{c}_\mu, \underline{C}_\mu, \underline{c}_\phi$ and $\underline{C}_\phi$, where $\lambda_{\min}(\underline{\mathbf{\Phi}})$ and $\lambda_{\max}(\underline{\mathbf{\Phi}})$ denote the smallest and largest eigenvalues of $\underline{\mathbf{\Phi}}$. The following proposition establishes the consistency of the $\mathcal{I}_y^2(q)$ test statistic defined in (9).

Proposition 5. *Suppose that the data are generated by model (12) and the above stated assumptions hold. Then, under the alternative hypothesis, we have* $\lim_{n\to\infty} \Pr(\mathcal{I}_y^2(q) \leq \gamma) = 0$ *for any* $\gamma > 0$.

2.5 Standardization Using Laplace Approximation

Cliff and Ord (1973, 1981) derived the exact mean and variance of the Moran $\mathcal{I}$ test statistic under normality, and introduced a corresponding small sample standardized version of that statistic. In the following we introduce an approximately standardized version of the $\mathcal{I}_y^2(q)$ test statistic, which is developed without assuming that the disturbances are normally distributed. An approximately standardized version of the $\mathcal{I}_u^2(q)$ test statistic is implicitly defined as a special case. Our approximately standardized version of the $\mathcal{I}_y^2(q)$ test statistic is based on a Laplace approximation.

As in the previous section, let $\widehat{\mathbf{u}}$ denote the OLS residuals and $\widehat{\sigma}_{\mathrm{u}}^2 = (n - K_x)^{-1}\widehat{\mathbf{u}}'\widehat{\mathbf{u}}$ denote the unbiased estimator for σ^2. Now consider $\widehat{\sigma}_{\mathrm{u}}^{-2}\widehat{\mathbf{V}} = [\widehat{\sigma}_{\mathrm{u}}^{-2}\widehat{\mathbf{V}}^{Y\prime}, \widehat{\sigma}_{\mathrm{u}}^{-2}\widehat{\mathbf{V}}^{X\prime}, \widehat{\sigma}_{\mathrm{u}}^{-2}\widehat{\mathbf{V}}^{U\prime}]'$ and let

$$\boldsymbol{\mu}_L = \mathrm{E}_L[\widehat{\sigma}_{\mathrm{u}}^{-2}\widehat{\mathbf{V}}] \qquad \text{and} \qquad \mathbf{\Phi}_L = \mathrm{E}_L[\widehat{\sigma}_{\mathrm{u}}^{-4}\widehat{\mathbf{V}}\widehat{\mathbf{V}}'] \tag{13}$$

denote the Laplace approximations of $\mathrm{E}[\widehat{\sigma}_{\mathrm{u}}^{-2}\widehat{\mathbf{V}}]$ and $\mathrm{E}[\widehat{\sigma}_{\mathrm{u}}^{-4}\widehat{\mathbf{V}}\widehat{\mathbf{V}}']$. Furthermore, let $\widehat{\boldsymbol{\mu}}_L$ and $\widehat{\mathbf{\Phi}}_L$ denote estimators for $\boldsymbol{\mu}_L$ and $\mathbf{\Phi}_L$. Our approximately standardized version of the $\mathcal{I}_y^2(q)$ test statistic, which was given in (9), is then defined as

$$\mathcal{I}^2_{y,S}(q) = \left[\mathrm{L}(\widehat{\sigma}_{\mathrm{u}}^{-2}\widehat{\mathbf{V}} - \widehat{\boldsymbol{\mu}}_L)\right]' \left[\mathrm{L}(\widehat{\boldsymbol{\Phi}}_L - \widehat{\boldsymbol{\mu}}_L\widehat{\boldsymbol{\mu}}_L')\mathrm{L}'\right]^{-1} \left[\mathrm{L}(\widehat{\sigma}_{\mathrm{u}}^{-2}\widehat{\mathbf{V}} - \widehat{\boldsymbol{\mu}}_L)\right]. \tag{14}$$

The explicit expressions for $\boldsymbol{\mu}_L$ and $\boldsymbol{\Phi}_L$ and their corresponding estimators $\widehat{\boldsymbol{\mu}}_L$ and $\widehat{\boldsymbol{\Phi}}_L$ are given in Appendix B.2.

In the following we give a brief outline of the derivation of $\boldsymbol{\mu}_L$ and $\boldsymbol{\Phi}_L$. Let $\mathbf{W}$ be generic for the weight matrices $\mathbf{W}_1, \cdots, \mathbf{W}_q$, then it is readily seen that generic elements of $\widehat{\sigma}_{\mathrm{u}}^{-2}\widehat{\mathbf{V}}^Y$, $\widehat{\sigma}_{\mathrm{u}}^{-2}\widehat{\mathbf{V}}^X$, and $\widehat{\sigma}_{\mathrm{u}}^{-2}\widehat{\mathbf{V}}^U$ are, respectively, given by

$$\widehat{\sigma}_{\mathrm{u}}^{-2}\mathbf{y}'\mathbf{W}'\widehat{\mathbf{u}} = \frac{\mathbf{u}'\mathbf{A}_y\mathbf{u} + \mathbf{a}_y'\mathbf{u}}{\mathbf{u}'\mathbf{S}\mathbf{u}}, \quad \widehat{\sigma}_{\mathrm{u}}^{-2}\mathbf{x}_k'\mathbf{W}'\widehat{\mathbf{u}} = \frac{\mathbf{a}_k'\mathbf{u}}{\mathbf{u}'\mathbf{S}\mathbf{u}}, \quad \widehat{\sigma}_{\mathrm{u}}^{-2}\widehat{\mathbf{u}}'\mathbf{W}'\widehat{\mathbf{u}} = \frac{\mathbf{u}'\mathbf{A}_u\mathbf{u}}{\mathbf{u}'\mathbf{S}\mathbf{u}} \tag{15}$$

with $\mathbf{A}_y = (\mathbf{M}_X\mathbf{W} + \mathbf{W}'\mathbf{M}_X)/2$, $\mathbf{a}_y = \mathbf{M}_X\mathbf{W}\mathbf{X}\boldsymbol{\beta}$, $\mathbf{a}_k = \mathbf{M}_X\mathbf{W}\mathbf{x}_k$, $\mathbf{A}_u = \mathbf{M}_X\overline{\mathbf{W}}\mathbf{M}_X$, and $\mathbf{S} = (n-K_x)^{-1}\mathbf{M}_X$. Form (15) we see that each elements of $\widehat{\sigma}_{\mathrm{u}}^{-2}\widehat{\mathbf{V}}$ can be expressed as a ratio of a linear, quadratic or linear-quadratic form over a quadratic form in $\mathbf{u}$. To obtain the Laplace approximation of the moments of $\widehat{\sigma}_{\mathrm{u}}^{-2}\widehat{\mathbf{V}}$, we extend Theorem 1 in Lieberman (1994) and introduce a proposition on the Laplace approximation of the moments of ratios of linear-quadratic forms over quadratic forms.[9]

Proposition 6. *Let* $\mathbf{u}$ *be an* $n \times 1$ *random vector, let* $\mathbf{A}$, $\mathbf{B}$ *and* $\mathbf{S}$ *be symmetric nonstochastic* $n \times n$ *matrices with* $\mathbf{S}$ *positive definite, and let* $\mathbf{a}$, $\mathbf{b}$ *be nonstochastic* $n \times 1$ *vectors. Assuming the existence of the joint moment generating function for* $\mathbf{u}'\mathbf{A}\mathbf{u} + \mathbf{a}'\mathbf{u}$, $\mathbf{u}'\mathbf{B}\mathbf{u} + \mathbf{b}'\mathbf{u}$ *and* $\mathbf{u}'\mathbf{S}\mathbf{u}$, $M(t_a, t_b, t) = \mathrm{E}\exp\left[t_a(\mathbf{u}'\mathbf{A}\mathbf{u} + \mathbf{a}'\mathbf{u}) + t_b(\mathbf{u}'\mathbf{B}\mathbf{u} + \mathbf{b}'\mathbf{u}) + t\mathbf{u}'\mathbf{S}\mathbf{u}\right]$ *and the subsequent expectations, we have the following Laplace approximation:*

$$\mathrm{E}\left[\left(\frac{\mathbf{u}'\mathbf{A}\mathbf{u} + \mathbf{a}'\mathbf{u}}{\mathbf{u}'\mathbf{S}\mathbf{u}}\right)^p \left(\frac{\mathbf{u}'\mathbf{B}\mathbf{u} + \mathbf{b}'\mathbf{u}}{\mathbf{u}'\mathbf{S}\mathbf{u}}\right)^q\right] \simeq \frac{\mathrm{E}\left[(\mathbf{u}'\mathbf{A}\mathbf{u} + \mathbf{a}'\mathbf{u})^p(\mathbf{u}'\mathbf{B}\mathbf{u} + \mathbf{b}'\mathbf{u})^q\right]}{\left[\mathrm{E}(\mathbf{u}'\mathbf{S}\mathbf{u})\right]^{p+q}}. \tag{16}$$

For nonnegative integers p and q, with $p + q = 1$. Proposition 6 delivers an approximation of the mean, and with $p+q = 2$ an approximation of the second moments of ratios of linear-quadratic forms over quadratic forms. Thus, using Proposition 6, together with Lemma A.1 in Kelejian and Prucha (2010), we can obtain the explicit expressions for $\boldsymbol{\mu}_L$ and $\boldsymbol{\Phi}_L$. The proof of Proposition 6 and a detailed derivation of the Laplace approximated moments are given in a Supplementary Appendix, which will be made available online.

[9] Lieberman (1994, Theorem 1) derived expressions for the Laplace approximation of the moments of ratios of quadratic forms. More specifically, Lieberman's theorem is obtained as a special case of Proposition 6 corresponding to $\mathbf{a} = \mathbf{0}$ and $q = 0$.

3 Test Statistics in General Settings

In the following we generalize the underlying setup for the generalized Moran $\mathcal{I}$ test statistics. In the previous sections we assumed that the regressor matrix $\mathbf{X}$ is nonstochastic. The previous section also assumes that the disturbances are homoskedastic. Those assumptions allowed for an intuitive motivation of our generalization of the Moran $\mathcal{I}$ test statistic. However, those assumptions may not be appropriate for certain applications. Thus in the following we expand the $\mathcal{I}_u^2(q)$ and $\mathcal{I}_y^2(q)$ tests to situations where the regressors are allowed to be endogenous under the null hypothesis, and where the disturbances are allowed to be heteroskedastic.

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

We first consider an extension of the $\mathcal{I}_u^2(q)$ test statistic, and suppose that the data are generated by the following linear regression model:

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\theta} + \mathbf{u} \tag{17}$$

where $\mathbf{y} = [y_1, \cdots, y_n]'$ is an $n \times 1$ vector of observations on the dependent variable, $\mathbf{Z} = [z_{ik}]$ is a $n \times K$ matrix of observations on K regressors, $\mathbf{u} = [u_1, \cdots, u_n]'$ is a $n \times 1$ vector of regression disturbances, and $\boldsymbol{\theta}$ is a $K \times 1$ vector of parameters.

In contrast to model (2), in the above model we allow for the regressors to be correlated with the disturbances, i.e., we allow for $\mathrm{E}(\mathbf{u}|\mathbf{Z}) \neq \mathbf{0}$. For example, the regressor matrix could be of the form $\mathbf{Z} = [\mathbf{Wy}, \mathbf{X}, \mathbf{WX}, \mathbf{Y}_o, \mathbf{WY}_o]$ where $\mathbf{Y}_o$ is a matrix of "outside" endogenous variables. Of course, the above setup contains model (2) as a special case.

We also assume the availability of an instrumental variable (IV) matrix $\mathbf{H}$ for which $\mathrm{E}(\mathbf{u}|\mathbf{H}) = \mathbf{0}$. For example, if $\underline{\mathbf{X}} = [\mathbf{X}, \mathbf{X}_o]$ denotes the set of all exogenous variables in the underlying system that generates the endogenous variables $[\mathbf{y}, \mathbf{Y}_o]$ in (17), then, $\mathbf{H}$ could be a subset the linearly independent columns of $\underline{\mathbf{X}}, \mathbf{W}\underline{\mathbf{X}}, \mathbf{W}^2\underline{\mathbf{X}}, \cdots$.

Now suppose the researcher wants to test the hypothesis that the disturbances are cross sectionally uncorrelated, i.e., H_0^u : cov($\mathbf{u}$) is diagonal, against the alternative H_1^u that the disturbances are cross sectionally correlated. To formally derive the properties of our test statistic we assume furthermore that under H_0^u the following conditions hold.

Assumption 2. *Suppose there are G endogenous variables in model (17).*
(i) *Let $\boldsymbol{\varepsilon}$ be a $(Gn) \times 1$ vector of i.i.d. $(0,1)$ innovations with finite $(2+\delta)$-th moments for some $\delta > 0$. Let $\boldsymbol{\Xi} = [\boldsymbol{\Xi}_1, \cdots, \boldsymbol{\Xi}_G]$ where $\boldsymbol{\Xi}_1, \cdots, \boldsymbol{\Xi}_G$ are $n \times n$ nonstochastic diagonal matrices with uniformly bounded diagonal elements. Then, $\mathbf{u} = \boldsymbol{\Xi}\boldsymbol{\varepsilon}$.*
(ii) *The elements of the instrument matrix $\mathbf{H}$ are nonstochastic and uniformly bounded. Furthermore, $\mathrm{plim}_{n\to\infty} n^{-1}\mathbf{H}'\mathbf{Z}$ is finite with full column rank, and $\lim_{n\to\infty} n^{-1}\mathbf{H}'\mathbf{H}$ is finite and nonsin-*

gular.

(iii) $n^{-1}\mathbf{Z}'\mathbf{W}_r\mathbf{Z} = O_p(1)$ *and* $n^{-1}\mathbf{Z}'\overline{\mathbf{W}}_r\mathbf{u} = n^{-1}\mathrm{E}(\mathbf{Z}'\overline{\mathbf{W}}_r\mathbf{u}) + o_p(1)$ *with* $n^{-1}\mathrm{E}(\mathbf{Z}'\overline{\mathbf{W}}_r\mathbf{u}) = O(1)$ *for* $r = 1, \cdots, q$.

Assumption 2 (i) is formulated fairly generally, to cover a wide range of data generating processes. For example, for $G > 1$ model (17) may represent one structural or a partially reduced form equation of a simultaneous system of G equations, see, e.g., Kelejian and Prucha (2004) and Cohen-Cole, Liu and Zenou (2017) for simultaneous systems with network spillovers. The g-th $n \times 1$ subvector of $\boldsymbol{\varepsilon}$ may then be viewed as the vector of innovations entering the g-th equation. Assumption 2 (ii) is in line with much of the literature on GMM estimation of Cliff-Ord type spatial models. Assumption 2 (iii) is, e.g., satisfied if the data are generated by a simultaneous system as considered in the above references.

The 2SLS estimator for $\boldsymbol{\theta}$ is given by $\widetilde{\boldsymbol{\theta}} = (\widetilde{\mathbf{Z}}'\widetilde{\mathbf{Z}})^{-1}\widetilde{\mathbf{Z}}'\mathbf{y}$, where $\widetilde{\mathbf{Z}} = \mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\mathbf{Z}$. Under the above assumption, it is readily seen that $\widetilde{\boldsymbol{\theta}}$ is a $n^{1/2}$-consistent estimator for $\boldsymbol{\theta}$. Let $\widetilde{\mathbf{V}}^U = [\widetilde{\mathbf{u}}'\mathbf{W}_1\widetilde{\mathbf{u}}, \cdots, \widetilde{\mathbf{u}}'\mathbf{W}_q\widetilde{\mathbf{u}}]'$ and

$$\widetilde{\boldsymbol{\Phi}}^{UU} = [2\mathrm{tr}(\overline{\mathbf{W}}_r\widetilde{\boldsymbol{\Sigma}}\overline{\mathbf{W}}_s\widetilde{\boldsymbol{\Sigma}}) + 4\widetilde{\mathbf{u}}'\overline{\mathbf{W}}_r(\mathbf{Z} - \widetilde{\mathbf{Z}})(\widetilde{\mathbf{Z}}'\widetilde{\mathbf{Z}})^{-1}\widetilde{\mathbf{Z}}'\widetilde{\boldsymbol{\Sigma}}\widetilde{\mathbf{Z}}(\widetilde{\mathbf{Z}}'\widetilde{\mathbf{Z}})^{-1}(\mathbf{Z} - \widetilde{\mathbf{Z}})'\overline{\mathbf{W}}_s\widetilde{\mathbf{u}}]_{r,s=1,\cdots,q} \tag{18}$$

where $\widetilde{\mathbf{u}} = [\widetilde{u}_1, \cdots, \widetilde{u}_n]' = \mathbf{y} - \mathbf{Z}\widetilde{\boldsymbol{\theta}}$ and $\widetilde{\boldsymbol{\Sigma}} = \mathrm{diag}(\widetilde{u}_i^2)$. Then, the generalized $\mathcal{I}_u^2(q)$ statistic is given by

$$\mathcal{I}_u^2(q) = \widetilde{\mathbf{V}}^{U\prime}(\widetilde{\boldsymbol{\Phi}}^{UU})^{-1}\widetilde{\mathbf{V}}^U. \tag{19}$$

Theorem 1. *Suppose the null hypothesis* H_0^u *and Assumptions 1 and 2 hold. Then* $n^{-1}\widetilde{\boldsymbol{\Phi}}^{UU} - n^{-1}\boldsymbol{\Phi}^{UU} = o_p(1)$*, where* $\boldsymbol{\Phi}^{UU}$ *is defined in (A.1) of the appendix. Furthermore, provided the smallest eigenvalues of* $n^{-1}\boldsymbol{\Phi}^{UU}$ *are bounded away form zero,*

$$\mathcal{I}_u^2(q) = \widetilde{\mathbf{V}}^{U\prime}(\widetilde{\boldsymbol{\Phi}}^{UU})^{-1}\widetilde{\mathbf{V}}^U \xrightarrow{d} \chi^2(q).$$

For $q = 1$ the above theorem contains results given in Kelejian and Prucha (2001) regarding the asymptotic distribution of the Moran $\mathcal{I}$ test statistic as a special case. For $q \geq 1$ and exogenous regressors, i.e., $\mathbf{Z} = \mathbf{X}$, the theorem delivers results given in Robinson (2008) as a special case. For $\mathbf{Z} = \mathbf{X}$ it is readily seen that $\mathrm{plim} n^{-1}\mathbf{Z}'\overline{\mathbf{W}}_s\widetilde{\mathbf{u}} = 0$. Consequently in this case the expression for the estimator for $\boldsymbol{\Phi}^{UU}$ can be simplified to $\widetilde{\boldsymbol{\Phi}}^{UU} = [2\mathrm{tr}(\overline{\mathbf{W}}_r\widetilde{\boldsymbol{\Sigma}}\overline{\mathbf{W}}_s\widetilde{\boldsymbol{\Sigma}})]$. If additionally the disturbances are homoskedastic, we have $\widetilde{\boldsymbol{\Phi}}^{UU} = [2\widetilde{\sigma}^4\mathrm{tr}(\overline{\mathbf{W}}_r\overline{\mathbf{W}}_s)]$. The square of the classical Moran $\mathcal{I}$ test statistic is now readily seen to be given by $\mathcal{I}_u^2(1)$. As an important by-product of the above theorem we thus see that the standard Moran $\mathcal{I}$ test statistic implicitly assumes that all regressors are exogenous. The expressions in (18) and (19) provide the adjustments needed for the test statistic to converge to a standard limiting distribution in case the model contains endogenous regressors.

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

Now suppose we want to test for the absence of network generated cross sectional dependence in the dependent variable. As discussed above, such dependence could stem from spillovers or interactions between cross-sectional units in the dependent variable, the regressors and/or the disturbances. More specifically, suppose the researcher wants to test (i) that the mean of the dependent variable of the i-th unit only depends on exogenous variables specific to the i-th unit, and thus is not affected by changes in the exogenous variables of the other units, and (ii) that the dependent variable is uncorrelated across units. More compactly, suppose the researcher wants to test the following hypothesis for the linear regression model (17):

$$H_0^y : \mathrm{E}(\mathbf{y}) = \underline{\mathbf{X}}\boldsymbol{\beta} \text{ and } \mathrm{cov}(\mathbf{y}) \text{ is diagonal,} \tag{20}$$

where $\underline{\mathbf{X}} = [\underline{x}_{ik}]$ is a non-stochastic matrix with $\partial \underline{x}_{ik}/\partial \underline{x}_{jl} = 0$ for $i \neq j$ (i.e. $\underline{\mathbf{X}}$ does not include spatial lags), against the alternative H_1^y that H_0^y is false. Recall that in this section the regressors $\mathbf{Z}$ are allowed to be endogenous. To complete our specification we assume that under H_0^y the reduced form for $\mathbf{Z}$ is given by

$$\mathbf{Z} = \underline{\mathbf{X}}\boldsymbol{\Pi} + \mathbf{E} \tag{21}$$

where $\mathbf{E} = [e_{ik}]$ is a matrix of reduced form disturbances with zero means. Of course, this implies that $\boldsymbol{\beta} = \boldsymbol{\Pi}\boldsymbol{\theta}$ in (20).

Our generalization of the $\mathcal{I}_y^2(q)$ given below necessitates that we are able to estimate the correlation structure between the disturbances $\mathbf{u}$ and $\mathbf{E}$. Thus the setup assumes that we are able to observe $\underline{\mathbf{X}}$. Consequently we assume in the following that the instrumental variable matrix $\mathbf{H}$ equals $\underline{\mathbf{X}}$. We maintain that the following additional conditions hold under the null hypothesis.

Assumption 3. *Suppose there are G endogenous variables in model (17).*
(i) *Let $\boldsymbol{\varepsilon}$ be a $(Gn)\times 1$ vector of i.i.d. $(0,1)$ innovations with finite $(2+\delta)$-th moments for some $\delta > 0$, and let $\mathbf{e}_k$ denote the k-th column of $\mathbf{E} = \mathbf{Z} - \underline{\mathbf{X}}\boldsymbol{\Pi}$. Let $\boldsymbol{\Xi} = [\boldsymbol{\Xi}_1, \cdots, \boldsymbol{\Xi}_G]$ and $\boldsymbol{\Psi}_k = [\boldsymbol{\Psi}_{1k}, \cdots, \boldsymbol{\Psi}_{Gk}]$, where $\boldsymbol{\Xi}_1, \cdots, \boldsymbol{\Xi}_G$ and $\boldsymbol{\Psi}_{1k}, \cdots, \boldsymbol{\Psi}_{Gk}$ are $n \times n$ nonstochastic diagonal matrices with uniformly bounded diagonal elements. Then, $\mathbf{u} = \boldsymbol{\Xi}\boldsymbol{\varepsilon}$ and $\mathbf{e}_k = \boldsymbol{\Psi}_k\boldsymbol{\varepsilon}$.*
(ii) *The matrix of exogenous regressors $\underline{\mathbf{X}} = [\underline{x}_{ik}]$ is nonstochastic with uniformly bounded elements and $\partial \underline{x}_{ik}/\partial \underline{x}_{jl} = 0$ for $i \neq j$. Furthermore, $\mathrm{plim}_{n\to\infty} n^{-1}\underline{\mathbf{X}}'\mathbf{Z}$ is finite with full column rank, and $\lim_{n\to\infty} n^{-1}\underline{\mathbf{X}}'\underline{\mathbf{X}}$ is finite and nonsingular.*

The above assumption is in essence an expansion of Assumption 2, and a discussion similar to that given in the context of Assumption 2 also applies to the above assumption.

The 2SLS estimator for the parameters $\boldsymbol{\theta}$ of model (17) is given by $\widehat{\boldsymbol{\theta}} = (\widehat{\mathbf{Z}}'\mathbf{Z})^{-1}\widehat{\mathbf{Z}}'\mathbf{y}$, where $\widehat{\mathbf{Z}} = \underline{\mathbf{X}}(\underline{\mathbf{X}}'\underline{\mathbf{X}})^{-1}\underline{\mathbf{X}}'\mathbf{Z}$. Let $\widehat{\boldsymbol{\Sigma}} = \mathrm{diag}(\widehat{u}_i^2)$, $\widehat{\boldsymbol{\Sigma}}_k = \mathrm{diag}(\widehat{u}_i\widehat{\epsilon}_{ik})$, and $\widehat{\boldsymbol{\Sigma}}_{kl} = \mathrm{diag}(\widehat{\epsilon}_{ik}\widehat{\epsilon}_{il})$, where $\widehat{u}_i$ is the

i-th element of $\widehat{\mathbf{u}} = \mathbf{y} - \mathbf{Z}\widehat{\boldsymbol{\theta}}$ and $\widehat{\epsilon}_{ik}$ is the (i,k)-th element of $\widehat{\mathbf{E}} = \mathbf{Z} - \widehat{\mathbf{Z}}$. Let

$$\widehat{\mathbf{V}} = \begin{bmatrix} \widehat{\mathbf{V}}^Y \\ \widehat{\mathbf{V}}^Z \\ \widehat{\mathbf{V}}^U \end{bmatrix} \qquad \text{and} \qquad \widehat{\boldsymbol{\Phi}} = \begin{bmatrix} \widehat{\boldsymbol{\Phi}}^{YY} & \widehat{\boldsymbol{\Phi}}^{YZ} & \widehat{\boldsymbol{\Phi}}^{YU} \\ (\widehat{\boldsymbol{\Phi}}^{YZ})' & \widehat{\boldsymbol{\Phi}}^{ZZ} & \widehat{\boldsymbol{\Phi}}^{ZU} \\ (\widehat{\boldsymbol{\Phi}}^{YU})' & (\widehat{\boldsymbol{\Phi}}^{ZU})' & \widehat{\boldsymbol{\Phi}}^{UU} \end{bmatrix}, \tag{22}$$

where

$$\begin{aligned} \widehat{\mathbf{V}}^Y &= [\widehat{\mathbf{u}}'\mathbf{W}_1\mathbf{y}, \cdots, \widehat{\mathbf{u}}'\mathbf{W}_q\mathbf{y}]', \quad \widehat{\mathbf{V}}^Z = [\widehat{\mathbf{u}}'\mathbf{W}_1\mathbf{Z}, \cdots, \widehat{\mathbf{u}}'\mathbf{W}_q\mathbf{Z}]', \quad \widehat{\mathbf{V}}^U = [\widehat{\mathbf{u}}'\mathbf{W}_1\widehat{\mathbf{u}}, \cdots, \widehat{\mathbf{u}}'\mathbf{W}_q\widehat{\mathbf{u}}]', \\ \widehat{\boldsymbol{\Phi}}^{YY} &= [\widehat{\phi}_{rs}^{YY}], \quad \widehat{\boldsymbol{\Phi}}^{YZ} = [\widehat{\boldsymbol{\phi}}_{rs}^{YZ}], \quad \widehat{\boldsymbol{\Phi}}^{YU} = [\widehat{\phi}_{rs}^{YU}], \quad \widehat{\boldsymbol{\Phi}}^{ZZ} = [\widehat{\boldsymbol{\phi}}_{rs}^{ZZ}], \quad \widehat{\boldsymbol{\Phi}}^{ZU} = [\widehat{\boldsymbol{\phi}}_{rs}^{ZU}], \quad \widehat{\boldsymbol{\Phi}}^{UU} = [\widehat{\phi}_{rs}^{UU}], \end{aligned}$$

with

$$\begin{aligned} \widehat{\phi}_{rs}^{YY} &= 2\mathrm{tr}(\overline{\mathbf{W}}_r\widehat{\boldsymbol{\Sigma}}\overline{\mathbf{W}}_s\widehat{\boldsymbol{\Sigma}}) + 2\textstyle\sum_{k=1}^K \widehat{\theta}_k \mathrm{tr}(\overline{\mathbf{W}}_s\widehat{\boldsymbol{\Sigma}}\mathbf{W}_r\widehat{\boldsymbol{\Sigma}}_k + \overline{\mathbf{W}}_r\widehat{\boldsymbol{\Sigma}}\mathbf{W}_s\widehat{\boldsymbol{\Sigma}}_k) \\ &\quad + \textstyle\sum_{k=1}^K\sum_{l=1}^K \widehat{\theta}_k\widehat{\theta}_l \mathrm{tr}(\mathbf{W}_r\widehat{\boldsymbol{\Sigma}}_k\mathbf{W}_s\widehat{\boldsymbol{\Sigma}}_l + \mathbf{W}_r\widehat{\boldsymbol{\Sigma}}_{kl}\mathbf{W}_s'\widehat{\boldsymbol{\Sigma}}) \\ &\quad + \widehat{\boldsymbol{\theta}}'\widehat{\mathbf{Z}}'\mathbf{W}_r'\mathbf{M}_{\widehat{Z}}\widehat{\boldsymbol{\Sigma}}\mathbf{M}_{\widehat{Z}}\mathbf{W}_s\widehat{\mathbf{Z}}\widehat{\boldsymbol{\theta}}, \\ \widehat{\boldsymbol{\phi}}_{rs}^{YZ} &= \Big[2\mathrm{tr}(\overline{\mathbf{W}}_r\widehat{\boldsymbol{\Sigma}}\mathbf{W}_s\widehat{\boldsymbol{\Sigma}}_l) + \textstyle\sum_{k=1}^K \widehat{\theta}_k \mathrm{tr}(\mathbf{W}_r\widehat{\boldsymbol{\Sigma}}_k\mathbf{W}_s\widehat{\boldsymbol{\Sigma}}_l + \mathbf{W}_r\widehat{\boldsymbol{\Sigma}}_{kl}\mathbf{W}_s'\widehat{\boldsymbol{\Sigma}})\Big]_{l=1,\cdots,K} \\ &\quad + \widehat{\boldsymbol{\theta}}'\widehat{\mathbf{Z}}'\mathbf{W}_r'\mathbf{M}_{\widehat{Z}}\widehat{\boldsymbol{\Sigma}}\mathbf{M}_{\widehat{Z}}\mathbf{W}_s\widehat{\mathbf{Z}}, \\ \widehat{\phi}_{rs}^{YU} &= 2\mathrm{tr}(\overline{\mathbf{W}}_r\widehat{\boldsymbol{\Sigma}}\overline{\mathbf{W}}_s\widehat{\boldsymbol{\Sigma}}) + 2\textstyle\sum_{k=1}^K \widehat{\theta}_k \mathrm{tr}(\mathbf{W}_r\widehat{\boldsymbol{\Sigma}}_k\overline{\mathbf{W}}_s\widehat{\boldsymbol{\Sigma}}), \\ \widehat{\boldsymbol{\phi}}_{rs}^{ZZ} &= \Big[\mathrm{tr}(\mathbf{W}_r\widehat{\boldsymbol{\Sigma}}_k\mathbf{W}_s\widehat{\boldsymbol{\Sigma}}_l + \mathbf{W}_r\widehat{\boldsymbol{\Sigma}}_{kl}\mathbf{W}_s'\widehat{\boldsymbol{\Sigma}})\Big]_{k,l=1,\cdots,K} + \widehat{\mathbf{Z}}'\mathbf{W}_r'\mathbf{M}_{\widehat{Z}}\widehat{\boldsymbol{\Sigma}}\mathbf{M}_{\widehat{Z}}\mathbf{W}_s\widehat{\mathbf{Z}}, \\ \widehat{\boldsymbol{\phi}}_{rs}^{ZU} &= \Big[2\mathrm{tr}(\mathbf{W}_r\widehat{\boldsymbol{\Sigma}}_k\overline{\mathbf{W}}_s\widehat{\boldsymbol{\Sigma}})\Big]_{k=1,\cdots,K}, \\ \widehat{\phi}_{rs}^{UU} &= 2\mathrm{tr}(\overline{\mathbf{W}}_r\widehat{\boldsymbol{\Sigma}}\overline{\mathbf{W}}_s\widehat{\boldsymbol{\Sigma}}), \end{aligned}$$

and $\mathbf{M}_{\widehat{Z}} = \mathbf{I}_n - \widehat{\mathbf{Z}}(\widehat{\mathbf{Z}}'\widehat{\mathbf{Z}})^{-1}\widehat{\mathbf{Z}}'$. As shown in the proof of the next theorem, $n^{-1}\widehat{\boldsymbol{\Phi}}$ is a consistent estimator for the VC matrix of $n^{-1/2}\widehat{\mathbf{V}}$. The generalized $\mathcal{I}_y^2(q)$ statistic is now given by

$$\mathcal{I}_y^2(q) = (\mathrm{L}\widehat{\mathbf{V}})'(\mathrm{L}\widehat{\boldsymbol{\Phi}}\mathrm{L}')^{-1}(\mathrm{L}\widehat{\mathbf{V}}), \tag{23}$$

where L is a selector matrix such that $\mathrm{L}\widehat{\boldsymbol{\Phi}}\mathrm{L}'$ is nonsingular. We have the following result regarding its limiting distribution:

Theorem 2. *Suppose the null hypothesis H_0^y hold and Assumptions 1 and 3 hold. Then $n^{-1}\widehat{\boldsymbol{\Phi}} - n^{-1}\boldsymbol{\Phi} = o_p(1)$ where $\boldsymbol{\Phi}$ is defined in (A.2) in the appendix. Furthermore, provided the smallest eigenvalue of $n^{-1}\mathrm{L}\widehat{\boldsymbol{\Phi}}\mathrm{L}'$ is bounded away from zero,*

$$\mathcal{I}_y^2(q) = (\mathrm{L}\widehat{\mathbf{V}})'(\mathrm{L}\widehat{\boldsymbol{\Phi}}\mathrm{L}')^{-1}(\mathrm{L}\widehat{\mathbf{V}}) \xrightarrow{d} \chi^2(\mathrm{rank}(\mathrm{L})).$$

The above theorem contains the results stated in Section 2 regarding the limiting distribution of the $\mathcal{I}_y^2(q)$ under the assumptions that all regressors are exogenous and the disturbances are homoskedastic as a special case. Observing that in this case $\mathbf{Z} = \mathbf{X}$, $\boldsymbol{\theta} = \boldsymbol{\beta}$, $\mathbf{E} = \mathbf{0}$, $\widehat{\boldsymbol{\Sigma}} = \widehat{\sigma}^2 \mathbf{I}_n$, and $\widehat{\boldsymbol{\Sigma}}_k = \widehat{\boldsymbol{\Sigma}}_{kl} = \mathbf{0}$ the above expression for $\widehat{\boldsymbol{\Phi}}$ is readily seen to reduce to the expression for $\widehat{\boldsymbol{\Phi}}$ given in (8).

The reason for the inclusion of the selector matrix L in the definition of the $\mathcal{I}_y^2(q)$ test statistic in (23) is similar to that discussed in Section 2.2. Note that the elements of $\widehat{\mathbf{V}}$ are linearly dependent, since $\mathbf{y} = \mathbf{Z}\widehat{\boldsymbol{\theta}} + \widehat{\mathbf{u}}$. Consequently $n^{-1}\widehat{\boldsymbol{\Phi}}$ is singular, at least in the limit. The selector matrix L ensures that $\mathrm{L}\widehat{\boldsymbol{\Phi}}\mathrm{L}'$ is nonsingular.

3.3 Endogenous and Misspecified Weight Matrices

Now we revisit the simple social interaction model (1) introduced in Section 1. Given root-n consistent estimators for the model parameters we can test for the absence of network generated cross sectional dependence in this model (i.e. $\lambda = \gamma = \rho = 0$) using, e.g., a Wald test. However, consistent estimation of (1) can be difficult in certain situations. For example, suppose a generic element of the weight matrix $\mathbf{W} = [w_{ij}]$ is given by

$$w_{ij} = f(\boldsymbol{\xi}_{ij}, \eta_{ij})$$

where $\boldsymbol{\xi}_{ij}$ is a vector of observable pair-specific characteristics and η_{ij} captures unobservable pair-specific heterogeneity. If η_{ij} is correlated with ε_i and ε_j, the innovations in the social interaction model (1), the weight matrix $\mathbf{W}$ is endogenous. In this case, without parametric assumptions on $f(\cdot)$ and the correlation structure between η_{ij} and $\boldsymbol{\varepsilon}$, consistent estimation of model (1) can be quite challenging.[10] The generalized Moran $\mathcal{I}$ tests, on the other hand, is relatively easier to implement. To be more specific, we could construct some auxiliary weight matrices, based on exogenous elements of $\boldsymbol{\xi}_{ij}$. For example, suppose $\xi_{ij,1}$ indicates whether individuals i and j are of the same gender, then we could construct a weight matrix with its (i, j)-th element being $\xi_{ij,1}$. Then, we can carry out the generalized Moran $\mathcal{I}$ test with these auxiliary weight matrices. As the auxiliary weight matrices are exogenous, the generalized Moran $\mathcal{I}$ test should have the proper size. However, the power of the test depends on how well linear combinations of the auxiliary weight matrices approximate the true weight matrix $\mathbf{W}$. We conduct Monte Carlo simulations to investigate the power of the test in the next section.

Second, data on social networks are often based on self-reported surveys and are subject to substantial measurement errors. For example, in the National Longitudinal Study of Adolescent

[10] For some recent progress on identification and estimation of models with endogenous networks, see, e.g., Qu and Lee (2015) and Johnsson and Moon (2016).

Health (Add Health),[11] students in sampled schools were asked to nominate their best friends, up to five males and five females, from the school roster. Although friendship is typically considered reciprocal, more than half of the friend nominations are not reciprocal in the Add Health data, indicating a possible measurement error in the self-reported friendship information. With the misspecified weight matrix, existing estimators (e.g. Bramoullé et al., 2009 and Lee, Liu and Lin, 2010) of (1) is likely to be inconsistent. In contrast, the generalized Moran $\mathcal{I}$ test is robust with misspecified weight matrices and can be useful in such situations.

4 Monte Carlo Study

In the following we report on the results of a Monte Carlo study of the finite sample properties of the $\mathcal{I}_u^2(q)$ and $\mathcal{I}_y^2(q)$ test statistics. Our specifications of the data generating process are geared towards social networks. The main findings of the study are reported in Tables 1-6 below. The full results of our Monte Carlo study are collected in an auxiliary appendix, which will be made available online on our web sites. In that appendix we also report on results were the specification of the data generating process is motivated by spatial networks.

In our Monte Carlo study we consider first data generating processes where the weight matrices representing the network structure are exogenous, and then data generating processes with endogenous network structures. Each Monte Carlo experiment is based on 10,000 repetitions.

4.1 Exogenous Network Structure

For this part of the study the $n \times 1$ vector of observations on the dependent variable $\mathbf{y}$ is generated from the following model:

$$\mathbf{y} = \lambda_1 \mathbf{W}_1 \mathbf{y} + \lambda_2 \mathbf{W}_2 \mathbf{y} + \beta \mathbf{z} + \gamma_1 \mathbf{W}_1 \mathbf{z} + \gamma_2 \mathbf{W}_2 \mathbf{z} + \mathbf{u}, \qquad \mathbf{u} = \rho_1 \mathbf{W}_1 \mathbf{u} + \rho_2 \mathbf{W}_2 \mathbf{u} + \mathbf{v}. \tag{24}$$

The observations on the regressor $\mathbf{z}$ are assumed to be generated by the reduced form

$$\mathbf{z} = \mathbf{x} + \tau \mathbf{e}, \tag{25}$$

where $\mathbf{x}$ is generated independently from Uniform$[0, 5]$. The distributions of $\mathbf{e}$ and $\mathbf{v}$ will be defined later, but we note that $\mathbf{e}$ and $\mathbf{v}$ will be correlated. The coefficient τ can take on two possible values: 0 and 1. The regressor $\mathbf{z}$ is exogenous if $\tau = 0$, and endogenous if $\tau = 1$. We set $\beta = 1$ in the DGP.

To generate the weight matrices used in the Monte Carlo experiments we partition n individuals into equal-sized groups with m individuals in each group. Let $\boldsymbol{\xi}_1 = [\xi_{11}, \cdots, \xi_{n1}]'$ be an $n \times 1$ vector of i.i.d. binary random variables taking values -1 and 1 with equal probability, and let

[11] For more information on Add Health data, see http://www.cpc.unc.edu/projects/addhealth.

$\boldsymbol{\xi}_2 = [\xi_{12}, \cdots, \xi_{n2}]'$ be an $n \times 1$ vector of i.i.d. discrete random variables taking values $1, 2, \cdots, 10$ with equal probability. Let $\mathbf{1}(\cdot)$ denote an indicator function that equals one if its argument is true and zero otherwise. Now define $\mathbf{W}_1^* = [w_{ij,1}^*]$ where $w_{ij,1}^* = \mathbf{1}(\xi_{i1} = \xi_{j1})$ if i and j are in the same group and $w_{ij,1}^* = 0$ otherwise, and $\mathbf{W}_2^* = [w_{ij,2}^*]$ where $w_{ij,2}^* = \left(1 + |\xi_{i2} - \xi_{j2}|\right)^{-1}$ if i and j are in the same group and $w_{ij,2}^* = 0$ otherwise. The weight matrices $\mathbf{W}_1 = [w_{ij,1}]$ and $\mathbf{W}_2 = [w_{ij,2}]$ are then obtained by normalizing $\mathbf{W}_1^*$ and $\mathbf{W}_2^*$ such that $w_{ij,r} = w_{ij,r}^* / \max_i \sum_{j=1}^n w_{ij,r}^*$ for $r = 1, 2$.[12] For an exemplary interpretation, suppose ξ_{i1} is an indicator for the gender of an individual, and ξ_{i2} represents the income decile of an individual. Then $\mathbf{W}_1^*$ and $\mathbf{W}_2^*$ reflect, respectively, network links based on the gender of the individuals and on the similarity in their incomes.

We keep the weight matrices $\mathbf{W}_1$ and $\mathbf{W}_2$ and the exogenous variable $\mathbf{x}$ constant for all simulation repetitions.

4.1.1 Simulations with Exogenous z and Homoskedastic Innovations

For our first set of Monte Carlo simulations the data are generated from (24)-(25) with $\tau = 0$. In this case, the regressor $\mathbf{z}$ ($= \mathbf{x}$) is exogenous. Also, for this set of simulations the innovations $\mathbf{v}$ are taken to be homoskedastic. More specifically, the innovations are generated as $\mathbf{v} = \sqrt{2}\boldsymbol{\varepsilon}$, where the elements of the $n \times 1$ vector $\boldsymbol{\varepsilon}$ are i.i.d. $(0, 1)$. We explore two alternative distributions of $\boldsymbol{\varepsilon}$: (a) the standard normal distribution $N(0, 1)$ and (b) the standardized log-normal distribution $(e^2 - e)^{-1/2}(\ln N(0, 1) - e^{1/2})$. Under $H_0^y : \lambda_1 = \lambda_2 = \gamma_1 = \gamma_2 = \rho_1 = \rho_2 = 0$ we have $R^2 \approx 0.5$ for both distributions of $\boldsymbol{\varepsilon}$.

For the Monte Carlo simulations pertaining to the $\mathcal{I}_u^2(q)$ test we set $\lambda_1 = \lambda_2 = \gamma_1 = \gamma_2 = 0$, and experiment with $\rho_1, \rho_2 \in \{0, 0.2, 0.4\}$ in the DGP. We consider the test statistics defined by (4) based on $\widetilde{\mathbf{V}}^U = \widetilde{\mathbf{u}}'\mathbf{W}_1\widetilde{\mathbf{u}}$, $\widetilde{\mathbf{V}}^U = \widetilde{\mathbf{u}}'\mathbf{W}_2\widetilde{\mathbf{u}}$, and $\widetilde{\mathbf{V}}^U = [\widetilde{\mathbf{u}}'\mathbf{W}_1\widetilde{\mathbf{u}}, \widetilde{\mathbf{u}}'\mathbf{W}_2\widetilde{\mathbf{u}}]'$ respectively, where $\widetilde{\mathbf{u}}$ denotes the OLS residuals from regressing $\mathbf{y}$ on $\mathbf{z}$. The simulation results for the $\mathcal{I}_u^2(q)$ tests, small sample standardized $\mathcal{I}_u^2(q)$ tests, and the corresponding Bonferroni tests[13] are reported in Table 1. For the Monte Carlo simulations pertaining to the $\mathcal{I}_y^2(q)$ test we experiment with $\lambda_1, \lambda_2 \in \{0, 0.1, 0.2\}$, $\rho_1, \rho_2 \in \{0, 0.2, 0.4\}$ and $\gamma_1, \gamma_2 \in \{0, 0.2, 0.5\}$ in the DGP. We consider the test statistics defined by (9) based on $\mathrm{L}\widehat{\mathbf{V}} = [\widehat{\mathbf{u}}'\mathbf{W}_1\mathbf{z}, \widehat{\mathbf{u}}'\mathbf{W}_1\widehat{\mathbf{u}}]'$ and $\mathrm{L}\widehat{\mathbf{V}} = [\widehat{\mathbf{u}}'\mathbf{W}_2\mathbf{z}, \widehat{\mathbf{u}}'\mathbf{W}_2\widehat{\mathbf{u}}]'$ with $q = 1$, and $\mathrm{L}\widehat{\mathbf{V}} = [\widehat{\mathbf{u}}'\mathbf{W}_1\mathbf{z}, \widehat{\mathbf{u}}'\mathbf{W}_2\mathbf{z}, \widehat{\mathbf{u}}'\mathbf{W}_1\widehat{\mathbf{u}}, \widehat{\mathbf{u}}'\mathbf{W}_2\widehat{\mathbf{u}}]'$ with $q = 2$, respectively, where $\widehat{\mathbf{u}}$ also denotes the OLS residuals from regressing $\mathbf{y}$ on $\mathbf{z}$. The simulation results for the $\mathcal{I}_y^2(q)$ tests, small sample standardized $\mathcal{I}_y^2(q)$ tests, and the corresponding Bonferroni tests[14] are reported in Table 2.

[12] Note $w_{ij,k} = w_{ij,k}^* / \min\{\max_i \sum_{j=1}^n w_{ij,k}^*, \max_j \sum_{i=1}^n w_{ij,k}^*\} = w_{ij,k}^* / \max_i \sum_{j=1}^n w_{ij,k}^*$ for $k = 1, 2$ as $\mathbf{W}_1^*$ and $\mathbf{W}_2^*$ are symmetric,

[13] Let p_r denote the p-value of the $\mathcal{I}_u^2(1)$ test statistic based on $\widetilde{\mathbf{V}}^U = \widetilde{\mathbf{u}}'\mathbf{W}_r\widetilde{\mathbf{u}}$, for $r = 1, 2$. The Bonferroni test with the nominal size α rejects H_0^u if $\min\{p_r\} \leq \alpha/2$.

[14] Let p_r denote the p-value of the $\mathcal{I}_y^2(1)$ test statistic based on $\mathrm{L}\widehat{\mathbf{V}} = [\widehat{\mathbf{u}}'\mathbf{W}_r\mathbf{z}, \widehat{\mathbf{u}}'\mathbf{W}_r\widehat{\mathbf{u}}]'$, for $r = 1, 2$. The Bonferroni test with the nominal size α rejects H_0^y if $\min\{p_r\} \leq \alpha/2$.

[Tables 1 and 2 approximately here]

We start with a discussion of the results on the size of the tests under the null hypotheses H_0^u and H_0^y. More specifically, the lines in Table 1 corresponding to $\rho_1 = \rho_2 = 0$ and those in Table 2 corresponding to $\lambda_1 = \lambda_2 = \rho_1 = \rho_2 = \gamma_1 = \gamma_2 = 0$ report the actual size of the tests. When the innovations $\boldsymbol{\varepsilon}$ follow the standard normal distribution, the actual sizes of the $\mathcal{I}_u^2(q)$ and $\mathcal{I}_y^2(q)$ tests and their small sample standardized versions are close to the asymptotic nominal size of 0.05. This is in contrast to the Bonferroni test, which under-rejects the null hypothesis. When the innovations $\boldsymbol{\varepsilon}$ follow the log-normal distribution, the actual size of the $\mathcal{I}_u^2(2)$ and $\mathcal{I}_y^2(4)$ tests based on both $\mathbf{W}_1$ and $\mathbf{W}_2$ is closer to the asymptotic nominal size than that of the $\mathcal{I}_u^2(1)$ and $\mathcal{I}_y^2(2)$ tests based on a single weight matrix ($\mathbf{W}_1$ or $\mathbf{W}_2$). Furthermore, the actual sizes of the small sample standardized $\mathcal{I}_u^2(q)$ and $\mathcal{I}_y^2(q)$ tests are closer to the asymptotic nominal size than those without small sample standardization.

We next discuss the results pertaining to the power of the tests under the alternative. Overall, we find that the power increases to one as the amount of cross sectional dependence increases and the small sample standardized tests improve the power of the tests. For the $\mathcal{I}_u^2(q)$ tests, we now consider the case with $\rho_1 \neq 0$ and $\rho_2 = 0$ in more detail. In this case, we expect the $\mathcal{I}_u^2(1)$ test based on $\mathbf{W}_1$ to outperform the $\mathcal{I}_u^2(2)$ test based on $\mathbf{W}_1$ and $\mathbf{W}_2$ as the former incorporates the information $\rho_2 = 0$ while the latter does not. Indeed, we find some relative loss in power from the $\mathcal{I}_u^2(2)$ test but the loss is mostly modest. We next consider the case with $\rho_2 \neq 0$. In this case, we expect the $\mathcal{I}_u^2(2)$ test to outperform the $\mathcal{I}_u^2(1)$ test based on $\mathbf{W}_1$ as the former incorporates the information about $\mathbf{W}_2$. An inspection of Table 1 shows that the results are consistent with this conjecture. When $\rho_1 = 0$ and $\rho_2 \neq 0$ (resp., $\rho_1 \neq 0$ and $\rho_2 = 0$), the results of the $\mathcal{I}_u^2(1)$ test based on $\mathbf{W}_1$ (resp., $\mathbf{W}_2$) provide some insight into the performance of $\mathcal{I}_u^2(q)$ tests when the weight matrices are misspecified. The results suggest that with misspecified weight matrices there is loss in power relative to tests with correctly specified weight matrices. Nevertheless, the power of the test with misspecified weight matrices increases with the amount of cross sectional dependence. Similar remarks apply for the results of the $\mathcal{I}_y^2(q)$ tests.

[Tables 3 and 4 approximately here]

To investigate the performance of the test when the number of candidate weight matrices q is large, we generate weight matrices $\mathbf{W}_1, \cdots, \mathbf{W}_q$ in the same way as $\mathbf{W}_1$. More specifically, let ξ_{ir} be an i.i.d. binary random variable taking values -1 and 1 with equal probability, for $i = 1, \cdots, n$ and $r = 1, \cdots, q$. Let $\mathbf{W}_r^* = [w_{ij,r}^*]$, where $w_{ij,r}^* = \mathbf{1}(\xi_{ir} = \xi_{jr})$ if i and j are in the same group and $w_{ij,r}^* = 0$ otherwise, for $r = 1, \cdots, q$. The matrices $\mathbf{W}_r = [w_{ij,r}]$ are then obtained by normalizing $\mathbf{W}_r^*$ such that $w_{ij,r} = w_{ij,r}^* / \max_i \sum_{j=1}^n w_{ij,r}^*$. For the Monte Carlo simulations pertaining to the

$\mathcal{I}_u^2(q)$ test we experiment with $\rho_1 \in \{0, 0.2, 0.4, 0.6, 0.8\}$ in the DGP:

$$\mathbf{y} = \beta \mathbf{x} + \mathbf{u}, \qquad \mathbf{u} = \rho_1 \mathbf{W}_1 \mathbf{u} + \mathbf{v}.$$

For the Monte Carlo simulations pertaining to the $\mathcal{I}_y^2(q)$ test we experiment with $\lambda_1 \in \{0, 0.1, 0.2\}$, $\rho_1 \in \{0, 0.2, 0.4\}$ and $\gamma_1 \in \{0, 0.2, 0.5\}$ in the DGP:

$$\mathbf{y} = \lambda_1 \mathbf{W}_1 \mathbf{y} + \beta \mathbf{x} + \gamma_1 \mathbf{W}_1 \mathbf{x} + \mathbf{u}, \qquad \mathbf{u} = \rho_1 \mathbf{W}_1 \mathbf{u} + \mathbf{v}.$$

In both DGPs, $\beta = 1$, $\mathbf{x}$ is generated independently from Uniform$[0, 5]$, and $\mathbf{v} = \sqrt{2}\boldsymbol{\varepsilon}$ with $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{I}_n)$.

The simulation results for the $\mathcal{I}_u^2(q)$ and $\mathcal{I}_y^2(q)$ tests, their small sample standardized tests, and the corresponding Bonferroni tests are reported in Tables 3 and 4. The actual sizes of the $\mathcal{I}_u^2(q)$ and $\mathcal{I}_y^2(q)$ tests and their small sample standardized versions are close to the asymptotic nominal size of 0.05. The Bonferroni test, on the other hand, under-rejects the null hypothesis. The downward size distortion of the Bonferroni test is more severe as q increases. Also, as expected, the power of the $\mathcal{I}_u^2(q)$ and $\mathcal{I}_y^2(q)$ tests reduces as q increases. The small sample standardized tests improve the power of the tests.

4.1.2 Simulations with Endogenous z and Heteroskedastic Innovations

For our next set of Monte Carlo simulations the data are generated again from (24)-(25). In contrast to the previous simulations the regressor $\mathbf{z}$ is endogenous by setting $\tau = 1$ in (25). Also, for this set of simulations the innovations are taken to be heteroskedastic. More specifically, let $r_i = (1+\xi_{i1}/2)^{1/2}$, then for each simulation repetition, we generate $\mathbf{v} = (v_1, \cdots, v_n)'$ and $\mathbf{e} = (e_1, \cdots, e_n)'$ such that $v_i = r_i \epsilon_{i1}$ and $e_i = r_i \epsilon_{i2}$, where ϵ_{i1} and ϵ_{i2} are respectively the i-th elements of $\boldsymbol{\epsilon}_1$ and $\boldsymbol{\epsilon}_2$ with

$$\begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \end{bmatrix} \sim \mathrm{N}\left(\mathbf{0}, \begin{bmatrix} \mathbf{I}_n & 0.5\mathbf{I}_n \\ 0.5\mathbf{I}_n & \mathbf{I}_n \end{bmatrix}\right).$$

For the Monte Carlo simulations pertaining to the $\mathcal{I}_u^2(q)$ test we set $\gamma_1 = \gamma_2 = 0$, and experiment with $\lambda_1, \lambda_2 \in \{0, 0.2\}$ and $\rho_1, \rho_2 \in \{0, 0.2, 0.4\}$ in the DGP. We consider the test statistics defined by (19) based on $\widetilde{\mathbf{V}}^U = \widetilde{\mathbf{u}}'\mathbf{W}_1\widetilde{\mathbf{u}}$, $\widetilde{\mathbf{V}}^U = \widetilde{\mathbf{u}}'\mathbf{W}_2\widetilde{\mathbf{u}}$, and $\widetilde{\mathbf{V}}^U = [\widetilde{\mathbf{u}}'\mathbf{W}_1\widetilde{\mathbf{u}}, \widetilde{\mathbf{u}}'\mathbf{W}_2\widetilde{\mathbf{u}}]'$ respectively, where $\widetilde{\mathbf{u}}$ denotes the 2SLS residuals from regressing $\mathbf{y}$ on $\mathbf{z}$ with the IV $\mathbf{x}$ when $\lambda_1 = \lambda_2 = 0$ and denotes the 2SLS residuals from regressing $\mathbf{y}$ on $[\mathbf{W}_1\mathbf{y}, \mathbf{W}_2\mathbf{y}, \mathbf{z}]$ with IVs $[\mathbf{W}_1\mathbf{x}, \mathbf{W}_2\mathbf{x}, \mathbf{x}]$ when $\lambda_1 = \lambda_2 = 0.2$. The simulation results for the $\mathcal{I}_u^2(q)$ tests and the Bonferroni test are reported in Table 5. For the Monte Carlo simulations pertaining to the $\mathcal{I}_y^2(q)$ test we experiment with $\lambda_1, \lambda_2 \in \{0, 0.1, 0.2\}$, $\rho_1, \rho_2 \in \{0, 0.2, 0.4\}$ and $\gamma_1, \gamma_2 \in \{0, 0.2, 0.5\}$ in the DGP. We consider the test statistics defined by (23) based on $\mathrm{L}\widehat{\mathbf{V}} = [\widehat{\mathbf{u}}'\mathbf{W}_1\mathbf{z}, \widehat{\mathbf{u}}'\mathbf{W}_1\widehat{\mathbf{u}}]'$ and $\mathrm{L}\widehat{\mathbf{V}} = [\widehat{\mathbf{u}}'\mathbf{W}_2\mathbf{z}, \widehat{\mathbf{u}}'\mathbf{W}_2\widehat{\mathbf{u}}]'$ with

$q = 1$, and $\mathrm{L}\widehat{\mathbf{V}} = [\widehat{\mathbf{u}}'\mathbf{W}_1\mathbf{z}, \widehat{\mathbf{u}}'\mathbf{W}_2\mathbf{z}, \widehat{\mathbf{u}}'\mathbf{W}_1\widehat{\mathbf{u}}, \widehat{\mathbf{u}}'\mathbf{W}_2\widehat{\mathbf{u}}]'$ with $q = 2$, respectively, where $\widehat{\mathbf{u}}$ also denotes the 2SLS residuals from regressing $\mathbf{y}$ on $\mathbf{z}$ with the IV $\mathbf{x}$. The simulation results for the $\mathcal{I}_y^2(q)$ tests are reported in Table 6.

[Tables 5 and 6 approximately here]

Similar to the results reported in the previous section, we find 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, while the Bonferroni test under-rejects the null hypothesis. The power of the tests increases to one as the amount of cross sectional dependence increases. Furthermore, when the cross sectional dependence is based on $\mathbf{W}_1$ but not on $\mathbf{W}_2$, there is some modest loss in power from the $\mathcal{I}_u^2(2)$ and $\mathcal{I}_y^2(2)$ tests based on both $\mathbf{W}_1$ and $\mathbf{W}_2$ relative to the $\mathcal{I}_u^2(1)$ and $\mathcal{I}_y^2(1)$ tests based on $\mathbf{W}_1$ only. However, when the cross sectional dependence is based on both $\mathbf{W}_1$ and $\mathbf{W}_2$, the $\mathcal{I}_u^2(2)$ and $\mathcal{I}_y^2(2)$ tests with both $\mathbf{W}_1$ and $\mathbf{W}_2$ outperform the $\mathcal{I}_u^2(1)$ and $\mathcal{I}_y^2(1)$ test based on $\mathbf{W}_1$ only.

4.2 Endogenous Network Structure

For this part of the simulation study, we assume the vector of dependent variables $\mathbf{y}$ of the n cross-sectional units is generated by

$$\mathbf{y} = \lambda\mathbf{W}_0\mathbf{y} + \beta\mathbf{x} + \gamma\mathbf{W}_0\mathbf{x} + \mathbf{u}, \qquad \mathbf{u} = \rho\mathbf{W}_0\mathbf{u} + \mathbf{v}, \tag{26}$$

where $\beta = 1$, $\mathbf{x}$ is generated independently from Uniform$[0, 5]$, and $\mathbf{v} = \sqrt{2}\boldsymbol{\varepsilon}$ with $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{I}_n)$, and where the elements of $\mathbf{W}_0$ are functions of the weight matrices $\mathbf{W}_1$ and $\mathbf{W}_2$ defined after (24)-(25).

More specifically, to generate the weight matrix, let

$$d_{ij} = \overline{w}_{ij,1} + \overline{w}_{ij,2} + \zeta_{ij}, \tag{27}$$

where $\overline{w}_{ij,1}$ and $\overline{w}_{ij,2}$ are standardized $w_{ij,1}$ and $w_{ij,2}$ respectively[15] and ζ_{ij} is a random innovation. Let $\mathbf{W}_0^* = [w_{ij,0}^*]$ where $w_{ij,0}^* = \mathbf{1}(d_{ij} > 0)$ if i and j are in the same group and $w_{ij,0}^* = 0$ otherwise. The weight matrix $\mathbf{W}_0 = [w_{ij,0}]$ is then obtained by normalizing $\mathbf{W}_0^*$ such that $w_{ij,0} = w_{ij,0}^* / \max_i \sum_{j=1}^n w_{ij,0}^*$. Abstractly, the design of the weight matrix $\mathbf{W}_0$ is motivated by a friendship network based on the simple homophily link formation model (27), where two individuals i and j in the same group are more likely to form a link, i.e. $w_{ij,0}^* = 1$, if they are of the same gender (captured by $w_{ij,1}$) and have similar incomes (captured by $w_{ij,2}$).

We consider both exogenous and endogenous $\mathbf{W}_0$ in the Monte Carlo experiment. For the exogenous $\mathbf{W}_0$, the random innovations ζ_{ij} in (27) are generated independently from $N(0, 1)$. For

[15] We standardize $w_{ij,1}$ and $w_{ij,2}$ so that $\overline{w}_{ij,1}$ and $\overline{w}_{ij,2}$ have empirical distributions with zero mean and unit variance.

the endogenous $\mathbf{W}_0$, the random innovations in (27) are given by $\zeta_{ij} = 2^{-1/2}(\varepsilon_i + \varepsilon_j)$, where ε_i and ε_j are random innovations of the outcome equation (26).

For the Monte Carlo simulations pertaining to the $\mathcal{I}_u^2(q)$ test we set $\lambda = \gamma = 0$, and experiment with $\rho \in \{0, 0.2, 0.4, 0.6, 0.8\}$ in the DGP. We consider the test statistics defined by (19) based on $\widetilde{\mathbf{V}}^U = \widetilde{\mathbf{u}}'\mathbf{W}_0\widetilde{\mathbf{u}}$, $\widetilde{\mathbf{V}}^U = \widetilde{\mathbf{u}}'\mathbf{W}_1\widetilde{\mathbf{u}}$, $\widetilde{\mathbf{V}}^U = \widetilde{\mathbf{u}}'\mathbf{W}_2\widetilde{\mathbf{u}}$, and $\widetilde{\mathbf{V}}^U = [\widetilde{\mathbf{u}}'\mathbf{W}_1\widetilde{\mathbf{u}}, \widetilde{\mathbf{u}}'\mathbf{W}_2\widetilde{\mathbf{u}}]'$ respectively, where $\widetilde{\mathbf{u}}$ denotes the OLS residuals from regressing $\mathbf{y}$ on $\mathbf{x}$. The simulation results for the $\mathcal{I}_u^2(q)$ tests are reported in Table 7. For the Monte Carlo simulations pertaining to the $\mathcal{I}_y^2(q)$ test we experiment with $\lambda \in \{0, 0.1, 0.2\}$, $\rho \in \{0, 0.2, 0.4\}$ and $\gamma \in \{0, 0.2, 0.5\}$ in the DGP. We consider the test statistics defined by (23) based on $\mathrm{L}\widehat{\mathbf{V}} = [\widehat{\mathbf{u}}'\mathbf{W}_0\mathbf{x}, \widehat{\mathbf{u}}'\mathbf{W}_0\widehat{\mathbf{u}}]'$, $\mathrm{L}\widehat{\mathbf{V}} = [\widehat{\mathbf{u}}'\mathbf{W}_1\mathbf{x}, \widehat{\mathbf{u}}'\mathbf{W}_1\widehat{\mathbf{u}}]'$ and $\mathrm{L}\widehat{\mathbf{V}} = [\widehat{\mathbf{u}}'\mathbf{W}_2\mathbf{x}, \widehat{\mathbf{u}}'\mathbf{W}_2\widehat{\mathbf{u}}]'$ with $q = 1$, and $\mathrm{L}\widehat{\mathbf{V}} = [\widehat{\mathbf{u}}'\mathbf{W}_1\mathbf{x}, \widehat{\mathbf{u}}'\mathbf{W}_2\mathbf{x}, \widehat{\mathbf{u}}'\mathbf{W}_1\widehat{\mathbf{u}}, \widehat{\mathbf{u}}'\mathbf{W}_2\widehat{\mathbf{u}}]'$ with $q = 2$, respectively, where $\widehat{\mathbf{u}}$ also denotes the OLS residuals from regressing $\mathbf{y}$ on $\mathbf{x}$. The simulation results for the $\mathcal{I}_y^2(q)$ tests are reported in Table 8.

[Tables 7 and 8 approximately here]

When $\mathbf{W}_0$ is exogenous, we expect the $\mathcal{I}_u^2(1)$ and $\mathcal{I}_y^2(1)$ tests based on $\mathbf{W}_0$ to outperform the $\mathcal{I}_u^2(q)$ and $\mathcal{I}_y^2(q)$ tests based on $\mathbf{W}_1$ and/or $\mathbf{W}_2$ as the former is based on the true weight matrix in the DGP. Indeed, we find some relative loss in power from the $\mathcal{I}_u^2(q)$ and $\mathcal{I}_y^2(q)$ tests based on $\mathbf{W}_1$ and/or $\mathbf{W}_2$ but the loss is mostly modest.

When $\mathbf{W}_0$ is endogenous we expect the $\mathcal{I}_u^2(1)$ and $\mathcal{I}_y^2(1)$ tests based on $\mathbf{W}_0$ to be distorted in size, but expect the tests based on $\mathbf{W}_1$ and/or $\mathbf{W}_2$ to be properly sized. Indeed, we find that the actual sizes of the $\mathcal{I}_u^2(q)$ and $\mathcal{I}_y^2(q)$ tests based on $\mathbf{W}_1$ and/or $\mathbf{W}_2$ are close to the asymptotic nominal size of 0.05, while the $\mathcal{I}_u^2(1)$ and $\mathcal{I}_y^2(1)$ tests based on $\mathbf{W}_0$ over-reject the null hypotheses. In particular, the size distortion is more severe (the actual size is 1) for the $\mathcal{I}_y^2(1)$ test with $\mathbf{W}_0$. For the $\mathcal{I}_u^2(q)$ and $\mathcal{I}_y^2(q)$ tests based on $\mathbf{W}_1$ and/or $\mathbf{W}_2$, the power increases to one as the amount of cross sectional dependence increases. The $\mathcal{I}_u^2(2)$ and $\mathcal{I}_y^2(2)$ tests based on both $\mathbf{W}_1$ and $\mathbf{W}_2$ improve the power of the $\mathcal{I}_u^2(1)$ and $\mathcal{I}_y^2(1)$ tests based on $\mathbf{W}_1$ or $\mathbf{W}_2$ only.

5 Conclusion

In this paper we introduced generalizations of the Moran $\mathcal{I}$ tests for network generated cross sectional dependence in the disturbance process, and in the dependent variable and the exogenous covariates. Those test incorporate information from multiple weight matrices. Weight matrices are frequently used to model the proximity between cross sectional units in a network. Our tests are intuitively motivated by the fact that empirical researchers are often faced with multiple potential choices for the weight matrices, but are unsure about the proper selection. While our tests are intuitive, they are also shown to have a formal interpretation as Lagrange Multiplier tests. 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. We also derive small sample standardized variants of the test statistics based on a Laplace approximation.

We also conduct Monte Carlo experiments to investigate the finite sample performance of the generalized Moran $\mathcal{I}$ 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. Furthermore, in situations where the network links are endogenous, the proposed tests could be used to test for cross sectional dependence generated by the endogenous network, using weight matrices based on exogenous proximity measures underlying the link formation process.

Our generalized Moran $\mathcal{I}$ tests are specification tests. As with any specification test, a closely related issue is the performance of pre-test estimators. The latter is a serious and difficult issue, without a general consensus in the profession on how to best deal with this issue – see Leeb and Pöetscher (2005, 2008, 2009) for a general discussion of pre-test estimators. It would be of interest to explore the behavior of pre-test estimators in connection with the generalized Moran $\mathcal{I}$ tests in future research. This should include an exploration of the effect on the estimation of impact measures. For large sample sizes, where relative efficiency may be of a lesser concern, it may be prudent to estimate an encompassing model with several weight matrices to avoid the pre-testing problem. Of course, opinions may differ, and future research may attempt to shed more light on this issue. Another area of interesting future research would be an extension of the $\mathcal{I}_u^2(q)$ and $\mathcal{I}_y^2(q)$ tests to panel data.

References

Anselin, L. (1988). Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity, *Geographical Analysis* **20**: 1–17.

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

Anselin, L. and Rey, S. (1991). Properties of tests for spatial dependence in linear regression models, *Geographical Analysis* **23**: 112–131.

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. and Li, D. (2000). LM tests for functional form and spatial correlation. Econometric Society World Congress 2000 Contributed Papers.

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

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.

Cohen-Cole, E., Liu, X. and Zenou, Y. (2017). Multivariate choices and identification of social interactions. Forthcoming in *Journal of Applied Econometrics*.

Conley, T. (1999). GMM estimation with cross sectional dependence, *Journal of Econometrics* **92**: 1–45.

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

Drukker, D. M. and Prucha, I. R. (2013). Small sample properties of the $I^2(q)$ test statistic for spatial dependence, *Spatial Economic Analysis* **8**: 271–292.

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

Jenish, N. and Prucha, I. R. (2009). Central limit theorems and uniform laws of large numbers for arrays of random fields, *Journal of Econometrics* **150**: 86–98.

Jenish, N. and Prucha, I. R. (2012). On spatial processes and asymptotic inference under near-epoch dependence, *Journal of Econometrics* **170**: 178–190.

Johnsson, I. and Moon, H. R. (2016). Estimation of peer effects in endogenous social networks: Control function approach. University of Southern California working paper.

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.

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. (2015). Dynamic spatial panel models: Networks, common shocks, and sequential exogeneity. Working paper, University of Maryland.

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.

Leeb, H. and Pötscher, B. M. (2005). Model selection and inference: Facts and fiction, *Econometric Theory* **21**: 21–59.

Leeb, H. and Pötscher, B. M. (2008). Sparse estimators and the oracle property, or the return of Hodges' estimator, *Journal of Econometrics* **142**: 201–211.

Leeb, H. and Pötscher, B. M. (2009). On the distribution of penalized maximum likelihood estimators: The LASSO, SCAD, and thresholding, *Journal of Multivariate Analysis* **100**: 2065–2082.

Lieberman, O. (1994). A Laplace approximation to the moments of a ratio of quadratic forms, *Biometrika* **81**: 681–690.

Liu, X. and Lee, L. F. (2010). GMM estimation of social interaction models with centrality, *Journal of Econometrics* **159**: 99–115.

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.

Olver, F. W. J. (1997). *Asymptotics and Special Functions*, CRC Press, New York.

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.

Pinkse, J. (1998). A consistent nonparametric test for serial independence, *Journal of Econometrics* **84**: 205–232.

Pinkse, J. (2004). Moran-flavored tests with nuisance parameters: examples, *in* L. Anselin and R. Florax (eds), *Advances in Spatial Econometrics*, Springer, pp. 67–78.

Pötscher, B. M. (1985). The behavior of the Lagrange multiplier test in testing the orders of ARMA models, *Metrika* **32**: 129–150.

Qu, X. and Lee, L. F. (2015). Estimating a spatial autoregressive model with an endogenous spatial weight matrix, *Journal of Econometrics* **184**: 209–232.

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

Robinson, P. M. and Rossi, F. (2014). Improved lagrange multiplier tests in spatial autoregressions, *Econometrics Journal* **17**: 139–164.

Robinson, P. M. and Rossi, F. (2015). Refined tests for spatial correlation, *Econometric Theory* **31**: 1–32.

Trenkler, G. and Schipp, B. (1993). Generalized inverses of partitioned matrices, *Econometric Theory* **9**: 530–533.

Yang, Z. L. (2015). LM tests of spatial dependence based on bootstrap critical values, *Journal of Econometrics* **185**: 33–59.

A Proofs of Theorems

Proof of Theorem 1. Observe that

$$n^{-1/2}\widetilde{\mathbf{u}}'\mathbf{W}_r\widetilde{\mathbf{u}} = n^{-1/2}\mathbf{u}'\overline{\mathbf{W}}_r\mathbf{u} - 2(n^{-1}\mathbf{Z}'\overline{\mathbf{W}}_r\mathbf{u})'n^{1/2}(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) + (\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta})'(n^{-1}\mathbf{Z}'\overline{\mathbf{W}}_r\mathbf{Z})n^{1/2}(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta})$$

and

$$n^{1/2}(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \left(n^{-1}\widetilde{\mathbf{Z}}'\mathbf{Z}\right)^{-1} n^{-1/2}\widetilde{\mathbf{Z}}'\mathbf{u} = \mathbf{P}_n' n^{-1/2}\mathbf{H}'\mathbf{u}$$

with $\mathbf{P}_n = n(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\mathbf{Z}\left[\mathbf{Z}'\mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\mathbf{Z}\right]^{-1}$. In light of Assumptions 2, $n^{-1}\mathbf{Z}'\overline{\mathbf{W}}_r\mathbf{Z} = O_p(1)$ and

$$n^{1/2}(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \mathbf{P}'n^{-1/2}\mathbf{H}'\boldsymbol{\Xi}\boldsymbol{\varepsilon} + o_p(1) = O_p(1),$$

where $\mathbf{P} = \mathbf{Q}_{HH}^{-1}\mathbf{Q}_{HZ}(\mathbf{Q}_{HZ}'\mathbf{Q}_{HH}^{-1}\mathbf{Q}_{HZ})^{-1}$ with $\mathbf{Q}_{HZ} = \text{plim}_{n\to\infty} n^{-1}\mathbf{H}'\mathbf{Z}$ and $\mathbf{Q}_{HH} = \lim_{n\to\infty} n^{-1}\mathbf{H}'\mathbf{H}$. Hence,

$$(n^{-1}\mathbf{Z}'\overline{\mathbf{W}}_r\mathbf{u})'n^{1/2}(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) = n^{-1}\text{E}(\mathbf{Z}'\overline{\mathbf{W}}_r\mathbf{u})'\mathbf{P}'n^{-1/2}\mathbf{H}'\boldsymbol{\Xi}\boldsymbol{\varepsilon} + o_p(1)$$

and

$$n^{-1/2}\widetilde{\mathbf{u}}'\mathbf{W}_r\widetilde{\mathbf{u}} = n^{-1/2}(\boldsymbol{\varepsilon}'\mathbf{C}_r\boldsymbol{\varepsilon} + \mathbf{c}_r'\boldsymbol{\varepsilon}) + o_p(1),$$

where $\mathbf{C}_r = \boldsymbol{\Xi}'\overline{\mathbf{W}}_r\boldsymbol{\Xi}$ and $\mathbf{c}_r = -2\boldsymbol{\Xi}'\mathbf{H}\mathbf{P}[n^{-1}\text{E}(\mathbf{Z}'\overline{\mathbf{W}}_r\mathbf{u})]$. Observe that $\mathbf{C}_r$ is a symmetric matrix with a zero diagonal. It follows from Lemma A.1 in Kelejian and Prucha (2010) that under H_0^u we have $\text{E}(\boldsymbol{\varepsilon}'\mathbf{C}_r\boldsymbol{\varepsilon} + \mathbf{c}_r'\boldsymbol{\varepsilon}) = 0$ and $\text{cov}[\boldsymbol{\varepsilon}'\mathbf{C}_r\boldsymbol{\varepsilon} + \mathbf{c}_r'\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon}'\mathbf{C}_s\boldsymbol{\varepsilon} + \mathbf{c}_s'\boldsymbol{\varepsilon}] = 2\text{tr}(\mathbf{C}_r\mathbf{C}_s) + \mathbf{c}_r'\mathbf{c}_s$ for $r, s = 1, \cdots, q$. Let

$$\boldsymbol{\Phi}^{UU} = [2\text{tr}(\mathbf{C}_r\mathbf{C}_s) + \mathbf{c}_r'\mathbf{c}_s]_{r,s=1,\cdots,q}. \tag{A.1}$$

Under the maintained assumptions, the row and column sums of $\mathbf{C}_r$ are uniformly bounded in absolute value by some finite constant, and the elements of $\mathbf{c}_r$ are uniformly bounded in absolute value by some finite constant. It follows that $n^{-1}\boldsymbol{\Phi}^{UU} = O(1)$. Since the smallest eigenvalues of $n^{-1}\boldsymbol{\Phi}^{UU}$ are bounded away from zero it follows immediately from Theorem A.1 in Kelejian and Prucha (2010) that

$$(\boldsymbol{\Phi}^{UU})^{-1/2}\widetilde{\mathbf{V}}^U = (n^{-1}\boldsymbol{\Phi}^{UU})^{-1/2}\begin{bmatrix} n^{-1/2}(\boldsymbol{\varepsilon}'\mathbf{C}_1\boldsymbol{\varepsilon} + \mathbf{c}_1'\boldsymbol{\varepsilon}) \\ \vdots \\ n^{-1/2}(\boldsymbol{\varepsilon}'\mathbf{C}_q\boldsymbol{\varepsilon} + \mathbf{c}_q'\boldsymbol{\varepsilon}) \end{bmatrix} + o_p(1) \xrightarrow{d} N(0, \mathbf{I}_q).$$

The above discussion also establishes that $n^{-1/2}(\boldsymbol{\varepsilon}'\mathbf{C}_1\boldsymbol{\varepsilon} + \mathbf{c}_1'\boldsymbol{\varepsilon}) = O_p(1)$. To show the desired result, it now suffices to show that $n^{-1}\widetilde{\boldsymbol{\Phi}}^{UU} = n^{-1}\boldsymbol{\Phi}^{UU} + o_p(1)$. From (A.1) it follows that the (r, s)-th element of $n^{-1}\mathbf{\Phi}^{UU}$ can be written as

$$2n^{-1}\mathrm{tr}(\overline{\mathbf{W}}_r\mathbf{\Sigma}\overline{\mathbf{W}}_{\mathbf{s}}\mathbf{\Sigma}) + 4[n^{-1}\mathrm{E}(\mathbf{u}'\overline{\mathbf{W}}_s\mathbf{Z})]\mathbf{P}'\left[n^{-1}\mathbf{H}'\mathbf{\Sigma}\mathbf{H}\right]\mathbf{P}[n^{-1}\mathrm{E}(\mathbf{Z}'\overline{\mathbf{W}}_s\mathbf{u})]$$

with $\mathbf{\Sigma} = \mathrm{E}\mathbf{u}\mathbf{u}' = \mathbf{\Xi}\mathbf{\Xi}'$. From (18) we see that the (r,s)-th element of $n^{-1}\widetilde{\mathbf{\Phi}}^{UU}$ can be written as

$$\begin{aligned} & 2n^{-1}\mathrm{tr}(\overline{\mathbf{W}}_r\widetilde{\mathbf{\Sigma}}\overline{\mathbf{W}}_s\widetilde{\mathbf{\Sigma}}) + 4\left[n^{-1}\widetilde{\mathbf{u}}'\overline{\mathbf{W}}_r(\mathbf{Z}-\widetilde{\mathbf{Z}})\right]\left[n^{-1}\widetilde{\mathbf{Z}}'\widetilde{\mathbf{Z}}\right]^{-1} n^{-1}\widetilde{\mathbf{Z}}'\widetilde{\mathbf{\Sigma}}\widetilde{\mathbf{Z}}\left[n^{-1}\widetilde{\mathbf{Z}}'\widetilde{\mathbf{Z}}\right]^{-1}\left[n^{-1}(\mathbf{Z}-\widetilde{\mathbf{Z}})'\overline{\mathbf{W}}_s\widetilde{\mathbf{u}}\right] \\ = \; & 2n^{-1}\mathrm{tr}(\overline{\mathbf{W}}_r\widetilde{\mathbf{\Sigma}}\overline{\mathbf{W}}_s\widetilde{\mathbf{\Sigma}}) + 4\left[n^{-1}\widetilde{\mathbf{u}}'\overline{\mathbf{W}}_r(\mathbf{Z}-\widetilde{\mathbf{Z}})\right]\mathbf{P}_n'\left[n^{-1}\mathbf{H}'\widetilde{\mathbf{\Sigma}}\mathbf{H}\right]\mathbf{P}_n\left[n^{-1}(\mathbf{Z}-\widetilde{\mathbf{Z}})'\overline{\mathbf{W}}_s\widetilde{\mathbf{u}}\right], \end{aligned}$$

In light of Assumptions 2, we have $\mathbf{P}_n = \mathbf{P} + o_p(1) = O_p(1)$, $n^{-1}\mathbf{Z}'\overline{\mathbf{W}}_r\widetilde{\mathbf{u}} = n^{-1}\mathbf{Z}'\overline{\mathbf{W}}_r\mathbf{u} - n^{-1}\mathbf{Z}'\overline{\mathbf{W}}_s\mathbf{Z}(\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}) = n^{-1}\mathrm{E}(\mathbf{Z}'\overline{\mathbf{W}}_r\mathbf{u})+o_p(1) = O_p(1)$, and $n^{-1}\widetilde{\mathbf{Z}}'\overline{\mathbf{W}}_r\widetilde{\mathbf{u}} = n^{-1}\widetilde{\mathbf{Z}}'\overline{\mathbf{W}}_r\mathbf{u}-n^{-1}\widetilde{\mathbf{Z}}'\overline{\mathbf{W}}_s\mathbf{Z}(\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}) = o_p(1)$. By similar argument as in Lemmata C.3-C.5 in Kelejian and Prucha (2010) it follows that $n^{-1}\mathrm{tr}(\overline{\mathbf{W}}_r\widetilde{\mathbf{\Sigma}}\overline{\mathbf{W}}_s\widetilde{\mathbf{\Sigma}}) = n^{-1}\mathrm{tr}(\overline{\mathbf{W}}_r\mathbf{\Sigma}\overline{\mathbf{W}}_s\mathbf{\Sigma}) + o_p(1) = O_p(1)$ and $n^{-1}\mathbf{H}'\widetilde{\mathbf{\Sigma}}\mathbf{H} = n^{-1}\mathbf{H}'\mathbf{\Sigma}\mathbf{H} + o_p(1) = O_p(1)$. The consistency of $n^{-1}\widetilde{\mathbf{\Phi}}^{UU}$ now follows from standard argumentation. □

The following two Lemmata will be used in the proof of Theorem 2.

Lemma A.1. *Suppose* $\mathbf{Z}$ *is defined by (21) and suppose Assumption 3 holds. Let* $\mathbf{W}$ *be an* $n \times n$ *nonstochastic zero-diagonal matrix with its row and column sums uniformly bounded in absolute value by some finite constant. Then*

$$\begin{aligned} n^{-1}\mathbf{Z}'\mathbf{W}\mathbf{Z} &= n^{-1}\mathbf{\Pi}'\underline{\mathbf{X}}'\mathbf{W}\underline{\mathbf{X}}\mathbf{\Pi} + o_p(1) = O_p(1) \\ n^{-1}\mathbf{Z}'\mathbf{W}\mathbf{u} &= o_p(1). \end{aligned}$$

Proof. Let $\mathbf{z}_k$ denote the k-th column of $\mathbf{Z}$ and $\boldsymbol{\pi}_k$ denote the k-th column of $\mathbf{\Pi}$. In light of Assumption 3 we have

$$\begin{aligned} \mathbf{z}_k'\mathbf{W}\mathbf{z}_l &= \boldsymbol{\pi}_k'\underline{\mathbf{X}}'\mathbf{W}\underline{\mathbf{X}}\boldsymbol{\pi}_l + \boldsymbol{\varepsilon}'\mathbf{\Psi}_k'\mathbf{W}\mathbf{\Psi}_l\boldsymbol{\varepsilon} + [\mathbf{\Psi}_k'\mathbf{W}\underline{\mathbf{X}}\boldsymbol{\pi}_l + \mathbf{\Psi}_l'\mathbf{W}'\underline{\mathbf{X}}\boldsymbol{\pi}_k]'\boldsymbol{\varepsilon} \\ &= \boldsymbol{\pi}_k'\underline{\mathbf{X}}'\mathbf{W}\underline{\mathbf{X}}\boldsymbol{\pi}_l + \boldsymbol{\varepsilon}'\mathbf{D}_1\boldsymbol{\varepsilon} + \mathbf{d}_1'\boldsymbol{\varepsilon} \end{aligned}$$

with $\mathbf{D}_1 = (\mathbf{\Psi}_k'\mathbf{W}\mathbf{\Psi}_l + \mathbf{\Psi}_l'\mathbf{W}'\mathbf{\Psi}_k)/2$ and $\mathbf{d}_1 = \mathbf{\Psi}_k'\mathbf{W}\underline{\mathbf{X}}\boldsymbol{\pi}_l + \mathbf{\Psi}_l'\mathbf{W}'\underline{\mathbf{X}}\boldsymbol{\pi}_k$. Observe that under the maintained assumptions the diagonal elements of $\mathbf{D}_1$ are zero. Consequently, employing the formulae for the mean and variance of linear quadratic forms given by Lemma A.1 in Kelejian and Prucha (2010) we have

$$\mathrm{E}[\mathbf{z}_k'\mathbf{W}\mathbf{z}_l] = \boldsymbol{\pi}_k'\underline{\mathbf{X}}'\mathbf{W}\underline{\mathbf{X}}\boldsymbol{\pi}_l, \qquad \text{and} \qquad \mathrm{cov}[\mathbf{z}_k'\mathbf{W}\mathbf{z}_l] = 2\mathrm{tr}(\mathbf{D}_1^2) + \mathbf{d}_1'\mathbf{d}_1.$$

Given the maintained assumptions $n^{-1}\boldsymbol{\pi}_k'\underline{\mathbf{X}}'\mathbf{W}\underline{\mathbf{X}}\boldsymbol{\pi}_l = O(1)$, $n^{-1}\mathrm{tr}(\mathbf{D}_1^2) = O(1)$, and $n^{-1}\mathbf{d}_1'\mathbf{d}_1 =$

$O(1)$. Thus

$$\mathrm{E}[n^{-1}\mathbf{z}_k'\mathbf{W}\mathbf{z}_l] = O(1) \qquad \text{and} \qquad \mathrm{cov}[n^{-1}\mathbf{z}_k'\mathbf{W}\mathbf{z}_l] = o(1),$$

which implies $n^{-1}\mathbf{z}_k'\mathbf{W}\mathbf{z}_l = \mathrm{E}[n^{-1}\mathbf{z}_k'\mathbf{W}\mathbf{z}_l] + o_p(1) = \boldsymbol{\pi}_k'\underline{\mathbf{X}}'\mathbf{W}\underline{\mathbf{X}}\boldsymbol{\pi}_l + o_p(1) = O_p(1)$.

Next observe that in light of Assumption 3 we have

$$\mathbf{z}_k'\mathbf{W}\mathbf{u} = \boldsymbol{\varepsilon}'\boldsymbol{\Psi}_k'\mathbf{W}\boldsymbol{\Xi}\boldsymbol{\varepsilon} + \boldsymbol{\pi}_k'\underline{\mathbf{X}}'\mathbf{W}\boldsymbol{\Xi}\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}'\mathbf{D}_2\boldsymbol{\varepsilon} + \mathbf{d}_2'\boldsymbol{\varepsilon}$$

with $\mathbf{D}_2 = (\boldsymbol{\Psi}_k'\mathbf{W}\boldsymbol{\Xi} + \boldsymbol{\Xi}'\mathbf{W}\boldsymbol{\Psi}_k)/2$ and $\mathbf{d}_2 = \boldsymbol{\Xi}'\mathbf{W}'\underline{\mathbf{X}}\boldsymbol{\pi}_k$. Observe that, under maintained assumptions, the diagonal elements of $\mathbf{D}_2$ are zero. Consequently, employing again the formulae for the mean and variance of linear quadratic forms given by Lemma A.1 in Kelejian and Prucha (2010) we have

$$\mathrm{E}[\mathbf{z}_k'\mathbf{W}\mathbf{u}] = 0, \qquad \text{and} \qquad \mathrm{cov}[\mathbf{z}_k'\mathbf{W}\mathbf{u}] = 2\mathrm{tr}(\mathbf{D}_2^2) + \mathbf{d}_2'\mathbf{d}_2.$$

Given the maintained assumptions $n^{-1}\mathrm{tr}(\mathbf{D}_2^2) = O(1)$ and $n^{-1}\mathbf{d}_2'\mathbf{d}_2 = O(1)$. Thus $cov[n^{-1}\mathbf{z}_k'\mathbf{W}\mathbf{u}] = o(1)$, and $n^{-1}\mathbf{z}_k'\mathbf{W}\mathbf{u} = \mathrm{E}[n^{-1}\mathbf{z}_k'\mathbf{W}\mathbf{u}] + o_p(1) = o_p(1)$. □

Lemma A.2. *Suppose the data generating process is defined by (17) and (21), and suppose Assumptions 1 and 3 hold. Suppose furthermore that*

$$n^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = n^{-1/2}\mathbf{T}'\mathbf{u} + o_p(1) = O_p(1),$$

where $\mathbf{T}$ *is a nonstochastic* $n \times K$ *matrix with its elements uniformly bounded in absolute value by some finite constant. Let* $\widehat{\mathbf{u}} = \mathbf{y} - \mathbf{Z}\widehat{\boldsymbol{\theta}}$ *and let* $\mathbf{W}$ *be an* $n \times n$ *nonstochastic zero-diagonal matrix with its row and column sums uniformly bounded in absolute value by some finite constant. Then*

$$\begin{aligned} n^{-1/2}\mathbf{y}'\mathbf{W}_r'\widehat{\mathbf{u}} &= n^{-1/2}\boldsymbol{\varepsilon}'\mathbf{A}_r\boldsymbol{\varepsilon} + n^{-1/2}\mathbf{a}_r'\boldsymbol{\varepsilon} + o_p(1) \\ n^{-1/2}\mathbf{z}_k'\mathbf{W}_r'\widehat{\mathbf{u}} &= n^{-1/2}\boldsymbol{\varepsilon}'\mathbf{B}_{k,r}\boldsymbol{\varepsilon} + n^{-1/2}\mathbf{b}_{k,r}'\boldsymbol{\varepsilon} + o_p(1) \\ n^{-1/2}\widehat{\mathbf{u}}'\mathbf{W}_r'\widehat{\mathbf{u}} &= n^{-1/2}\boldsymbol{\varepsilon}'\mathbf{C}_r\boldsymbol{\varepsilon} + o_p(1), \end{aligned}$$

where

$$\begin{aligned} \mathbf{A}_r &= \boldsymbol{\Xi}'\overline{\mathbf{W}}_r\boldsymbol{\Xi} + \frac{1}{2}\sum_{k=1}^{K}\theta_k[\boldsymbol{\Xi}'\mathbf{W}_r\boldsymbol{\Psi}_k + \boldsymbol{\Psi}_k'\mathbf{W}_r'\boldsymbol{\Xi}], \\ \mathbf{a}_r &= \boldsymbol{\Xi}'\left[\mathbf{I}_n - n^{-1}\mathbf{T}\boldsymbol{\Pi}'\underline{\mathbf{X}}'\right]\mathbf{W}_r\underline{\mathbf{X}}\boldsymbol{\Pi}\boldsymbol{\theta}, \\ \mathbf{B}_{k,r} &= \frac{1}{2}[\boldsymbol{\Xi}'\mathbf{W}_r\boldsymbol{\Psi}_k + \boldsymbol{\Psi}_k'\mathbf{W}_r'\boldsymbol{\Xi}], \\ \mathbf{b}_{k,r} &= \boldsymbol{\Xi}'\left[\mathbf{I}_n - n^{-1}\mathbf{T}\boldsymbol{\Pi}'\underline{\mathbf{X}}'\right]\mathbf{W}_r\underline{\mathbf{X}}\boldsymbol{\pi}_k, \end{aligned}$$

and $\mathbf{C}_r = \boldsymbol{\Xi}'\overline{\mathbf{W}}_r\boldsymbol{\Xi}$*.*

Proof. Observing that $\widehat{\mathbf{u}} = \mathbf{u} - \mathbf{Z}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})$, employing Assumptions 1, 3 and Lemma A.1, and recalling from the proof of that lemma that $\mathbf{z}_k'\mathbf{W}'\mathbf{u} = \boldsymbol{\varepsilon}'\boldsymbol{\Psi}_k'\mathbf{W}'\boldsymbol{\Xi}\boldsymbol{\varepsilon} + \boldsymbol{\pi}_k'\underline{\mathbf{X}}'\mathbf{W}'\boldsymbol{\Xi}\boldsymbol{\varepsilon}$ it follows that

$$\begin{aligned} & n^{-1/2}\mathbf{y}'\mathbf{W}_r'\widehat{\mathbf{u}} \\ = \; & n^{-1/2}\mathbf{u}'\mathbf{W}_r'\mathbf{u} + n^{-1/2}\boldsymbol{\theta}'\mathbf{Z}'\mathbf{W}_r'\mathbf{u} - \boldsymbol{\theta}'[n^{-1}\mathbf{Z}'\mathbf{W}_r'\mathbf{Z}]n^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) - [n^{-1}\mathbf{u}'\mathbf{W}_r'\mathbf{Z}]n^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \\ = \; & n^{-1/2}[\boldsymbol{\varepsilon}'\boldsymbol{\Xi}'\mathbf{W}_r'\boldsymbol{\Xi}\boldsymbol{\varepsilon} + \textstyle\sum_{k=1}^{K}\theta_k\boldsymbol{\varepsilon}'\boldsymbol{\Psi}_k'\mathbf{W}_r'\boldsymbol{\Xi}\boldsymbol{\varepsilon} + \boldsymbol{\theta}'\boldsymbol{\Pi}'\underline{\mathbf{X}}'\mathbf{W}_r'\boldsymbol{\Xi}\boldsymbol{\varepsilon} - n^{-1}\boldsymbol{\theta}'\boldsymbol{\Pi}'\underline{\mathbf{X}}'\mathbf{W}_r'\underline{\mathbf{X}}\boldsymbol{\Pi}\mathbf{T}'\boldsymbol{\Xi}\boldsymbol{\varepsilon}] + o_p(1), \end{aligned}$$

which verifies that $n^{-1/2}\mathbf{y}'\mathbf{W}_r'\widehat{\mathbf{u}} = n^{-1/2}\boldsymbol{\varepsilon}'\mathbf{A}_r\boldsymbol{\varepsilon} + n^{-1/2}\mathbf{a}_r'\boldsymbol{\varepsilon} + o_p(1)$ as claimed.

Next observe that

$$\begin{aligned} & n^{-1/2}\mathbf{z}_k'\mathbf{W}_r'\widehat{\mathbf{u}} \\ = \; & n^{-1/2}\boldsymbol{\pi}_k'\underline{\mathbf{X}}'\mathbf{W}_r'\mathbf{u} + n^{-1/2}\boldsymbol{\varepsilon}'\boldsymbol{\Psi}_k'\mathbf{W}_r'\mathbf{u} - n^{-1}\mathbf{z}_k'\mathbf{W}_r'\mathbf{Z}n^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \\ = \; & n^{-1/2}[\boldsymbol{\varepsilon}'\boldsymbol{\Psi}_k'\mathbf{W}_r'\boldsymbol{\Xi}\boldsymbol{\varepsilon} + \boldsymbol{\pi}_k'\underline{\mathbf{X}}'\mathbf{W}_r'\boldsymbol{\Xi}\boldsymbol{\varepsilon} - n^{-1}\boldsymbol{\pi}_k'\underline{\mathbf{X}}'\mathbf{W}_r'\underline{\mathbf{X}}\boldsymbol{\Pi}\mathbf{T}'\boldsymbol{\Xi}\boldsymbol{\varepsilon}] + o_p(1), \end{aligned}$$

which verifies that $n^{-1/2}\mathbf{z}_k'\mathbf{W}_r'\widehat{\mathbf{u}} = n^{-1/2}\boldsymbol{\varepsilon}'\mathbf{B}_{k,r}\boldsymbol{\varepsilon} + n^{-1/2}\mathbf{b}_{k,r}'\boldsymbol{\varepsilon} + o_p(1)$ as claimed.

Finally, in light of Assumptions 1 and 3, it follows from Lemma A.1 that

$$\begin{aligned} & n^{-1/2}\widehat{\mathbf{u}}'\mathbf{W}_r'\widehat{\mathbf{u}} \\ = \; & n^{-1/2}\mathbf{u}'\mathbf{W}_r'\mathbf{u} - 2[n^{-1}\mathbf{u}'\overline{\mathbf{W}}_r\mathbf{Z}]n^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) + (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})'[n^{-1}\mathbf{Z}'\mathbf{W}_r\mathbf{Z}]n^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \\ = \; & n^{-1/2}\boldsymbol{\varepsilon}'\boldsymbol{\Xi}'\mathbf{W}_r'\boldsymbol{\Xi}\boldsymbol{\varepsilon} + o_p(1), \end{aligned}$$

which verifies that $n^{-1/2}\widehat{\mathbf{u}}'\mathbf{W}_r'\widehat{\mathbf{u}} = n^{-1/2}\boldsymbol{\varepsilon}'\mathbf{C}_r\boldsymbol{\varepsilon} + o_p(1)$ as claimed.

□

Proof of Theorem 2. Observe that

$$n^{-1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = (n^{-1}\widehat{\mathbf{Z}}'\widehat{\mathbf{Z}})^{-1}n^{-1/2}\widehat{\mathbf{Z}}'\mathbf{u} = n^{-1/2}\widehat{\mathbf{T}}'\mathbf{u}$$

where $\widehat{\mathbf{T}} = \widehat{\mathbf{Z}}(n^{-1}\widehat{\mathbf{Z}}'\widehat{\mathbf{Z}})^{-1} = \underline{\mathbf{X}}\widehat{\mathbf{P}}$ with $\widehat{\mathbf{P}} = (n^{-1}\underline{\mathbf{X}}'\underline{\mathbf{X}})^{-1}(n^{-1}\underline{\mathbf{X}}'\mathbf{Z})[(n^{-1}\mathbf{Z}'\underline{\mathbf{X}})(n^{-1}\underline{\mathbf{X}}'\underline{\mathbf{X}})^{-1}(n^{-1}\underline{\mathbf{X}}'\mathbf{Z})]^{-1}$. Let $\mathbf{T} = \underline{\mathbf{X}}\mathbf{P}$ with $\mathbf{P} = \mathbf{Q}_{XX}^{-1}\mathbf{Q}_{XZ}[\mathbf{Q}_{XZ}\mathbf{Q}_{XX}^{-1}\mathbf{Q}_{XZ}]^{-1}$, and observe that under the maintained assumptions $\mathbf{T}$ is a nonstochastic $n \times K$ matrix with its elements uniformly bounded in absolute value by some finite constant. Furthermore, $\widehat{\mathbf{P}} \xrightarrow{p} \mathbf{P}$ and $n^{-1/2}\underline{\mathbf{X}}'\mathbf{u} = O_p(1)$, and thus

$n^{-1/2}\widehat{\mathbf{T}}'\mathbf{u} = n^{-1/2}\mathbf{T}'\mathbf{u} + o_p(1)$. Now let

$$\mathbf{V}_\circ = \begin{bmatrix} \mathbf{V}_\circ^Y \\ \mathbf{V}_\circ^Z \\ \mathbf{V}_\circ^U \end{bmatrix}, \quad \mathbf{V}_\circ^Y = \begin{bmatrix} \boldsymbol{\varepsilon}'\mathbf{A}_1\boldsymbol{\varepsilon} + \mathbf{a}_1'\boldsymbol{\varepsilon} \\ \vdots \\ \boldsymbol{\varepsilon}'\mathbf{A}_q\boldsymbol{\varepsilon} + \mathbf{a}_q'\boldsymbol{\varepsilon} \end{bmatrix}, \quad \mathbf{V}_\circ^U = \begin{bmatrix} \boldsymbol{\varepsilon}'\mathbf{C}_1\boldsymbol{\varepsilon} \\ \vdots \\ \boldsymbol{\varepsilon}'\mathbf{C}_q\boldsymbol{\varepsilon} \end{bmatrix}$$

$$\mathbf{V}_\circ^Z = \begin{bmatrix} \mathbf{V}_{\circ,1}^Z \\ \vdots \\ \mathbf{V}_{\circ,q}^Z \end{bmatrix}, \quad \mathbf{V}_{\circ,r}^Z = \begin{bmatrix} \boldsymbol{\varepsilon}'\mathbf{B}_{1,r}\boldsymbol{\varepsilon} + \mathbf{b}_{1,r}'\boldsymbol{\varepsilon} \\ \vdots \\ \boldsymbol{\varepsilon}'\mathbf{B}_{K,r}\boldsymbol{\varepsilon} + \mathbf{b}_{K,r}'\boldsymbol{\varepsilon} \end{bmatrix},$$

with $\mathbf{A}_r, \mathbf{B}_{k,r}, \mathbf{C}_r, \mathbf{a}_r$ and $\mathbf{b}_{k,r}$ as defined in Lemma A.2, then by that lemma $n^{-1/2}\widehat{\mathbf{V}} - n^{-1/2}\mathbf{V}_\circ = o_p(1)$. Furthermore, observing that $\mathbf{A}_r, \mathbf{B}_{k,r}$ and $\mathbf{C}_r$ are symmetric matrices with zero diagonals, it follows from Lemma A.1 in Kelejian and Prucha (2010) that

$$\boldsymbol{\Phi} = \mathrm{E}\mathbf{V}_\circ\mathbf{V}_\circ' = \begin{bmatrix} \boldsymbol{\Phi}^{YY} & \boldsymbol{\Phi}^{YZ} & \boldsymbol{\Phi}^{YU} \\ (\boldsymbol{\Phi}^{YZ})' & \boldsymbol{\Phi}^{ZZ} & \boldsymbol{\Phi}^{ZU} \\ (\boldsymbol{\Phi}^{YU})' & (\boldsymbol{\Phi}^{ZU})' & \boldsymbol{\Phi}^{UU} \end{bmatrix} \tag{A.2}$$

where

$$\begin{aligned}
\underset{q\times q}{\boldsymbol{\Phi}^{YY}} &= [\phi_{rs}^{YY}]_{r,s=1,\cdots,q}, & \underset{1\times 1}{\phi_{rs}^{YY}} &= 2\mathrm{tr}(\mathbf{A}_r\mathbf{A}_s) + \mathbf{a}_r'\mathbf{a}_s \\
\underset{q\times qK}{\boldsymbol{\Phi}^{YZ}} &= [\boldsymbol{\phi}_{rs}^{YZ}]_{r,s=1,\cdots,q}, & \underset{1\times K}{\boldsymbol{\phi}_{rs}^{YZ}} &= [2\mathrm{tr}(\mathbf{A}_r\mathbf{B}_{l,s}) + \mathbf{a}_r'\mathbf{b}_{l,s}]_{l=1,\cdots,K} \\
\underset{q\times q}{\boldsymbol{\Phi}^{YU}} &= [\phi_{rs}^{YU}]_{r,s=1,\cdots,q}, & \underset{1\times 1}{\phi_{rs}^{YU}} &= 2\mathrm{tr}(\mathbf{A}_r\mathbf{C}_s) \\
\underset{qK\times qK}{\boldsymbol{\Phi}^{ZZ}} &= [\boldsymbol{\phi}_{rs}^{ZZ}]_{r,s=1,\cdots,q}, & \underset{K\times K}{\boldsymbol{\phi}_{rs}^{ZZ}} &= [2\mathrm{tr}(\mathbf{B}_{k,r}\mathbf{B}_{l,s}) + \mathbf{b}_{k,r}'\mathbf{b}_{l,s}]_{k,l=1,\cdots,K} \\
\underset{qK\times q}{\boldsymbol{\Phi}^{ZU}} &= [\boldsymbol{\phi}_{rs}^{ZU}]_{r,s=1,\cdots,q}, & \underset{K\times 1}{\boldsymbol{\phi}_{rs}^{ZU}} &= [2\mathrm{tr}(\mathbf{B}_{k,r}\mathbf{C}_s)]_{k=1,\cdots,K} \\
\underset{q\times q}{\boldsymbol{\Phi}^{UU}} &= [\phi_{rs}^{UU}]_{r,s=1,\cdots,q}, & \underset{1\times 1}{\phi_{rs}^{UU}} &= 2\mathrm{tr}(\mathbf{C}_r\mathbf{C}_s),
\end{aligned}$$

Under the maintained assumptions the row and column sums of $\mathbf{A}_r, \mathbf{B}_{k,r}$ and $\mathbf{C}_r$ are uniformly bounded in absolute value by some finite constant, and the elements of $\mathbf{a}_r$ and $\mathbf{b}_{k,r}$ are uniformly bounded in absolute value by some finite constant. Thus, $n^{-1}\boldsymbol{\Phi} = O(1)$. By assumption the smallest eigenvalue of $\mathrm{L}\boldsymbol{\Phi}\mathrm{L}'$ is bounded away from zero. Consequently it follows immediately from Lemma A.1 and Theorem A.1 in Kelejian and Prucha (2010) and Lemma A.2 that

$$(\mathrm{L}\boldsymbol{\Phi}\mathrm{L}')^{-1/2}(\mathrm{L}\widehat{\mathbf{V}}) = (n^{-1}\mathrm{L}\boldsymbol{\Phi}\mathrm{L}')^{-1/2}(n^{-1/2}\mathrm{L}\mathbf{V}_\circ) + o_p(1) \xrightarrow{d} N(0, \mathbf{I}_{(K_\star+2)q}).$$

Observe that $n^{-1/2}\mathrm{L}\mathbf{V}_\circ = O_p(1)$. To show the desired result it thus suffices to show that $n^{-1}\widehat{\boldsymbol{\Phi}} = n^{-1}\boldsymbol{\Phi} + o_p(1)$. Let $\boldsymbol{\Sigma} = \mathrm{E}\mathbf{u}\mathbf{u}' = \boldsymbol{\Xi}\boldsymbol{\Xi}'$, $\boldsymbol{\Sigma}_k = \mathrm{E}\boldsymbol{\varepsilon}_k\mathbf{u}' = \boldsymbol{\Psi}_k\boldsymbol{\Xi}'$, and $\boldsymbol{\Sigma}_{kl} = \mathrm{E}\boldsymbol{\varepsilon}_k\boldsymbol{\varepsilon}_l' = \boldsymbol{\Psi}_k\boldsymbol{\Psi}_l'$ and observe that

$$\begin{aligned}
\phi_{rs}^{YY} &= 2\text{tr}(\overline{\mathbf{W}}_r\mathbf{\Sigma}\overline{\mathbf{W}}_s\mathbf{\Sigma}) + 2\textstyle\sum_{k=1}^K \theta_k \text{tr}(\overline{\mathbf{W}}_s\mathbf{\Sigma}\mathbf{W}_r\mathbf{\Sigma}_k + \overline{\mathbf{W}}_r\mathbf{\Sigma}\mathbf{W}_s\mathbf{\Sigma}_k) \\
&\quad + \textstyle\sum_{k=1}^K \sum_{l=1}^K \theta_k\theta_l \text{tr}(\mathbf{W}_r\mathbf{\Sigma}_k\mathbf{W}_s\mathbf{\Sigma}_l + \mathbf{W}_r\mathbf{\Sigma}_{kl}\mathbf{W}_s'\mathbf{\Sigma}) \\
&\quad + \boldsymbol{\theta}'\mathbf{\Pi}'\underline{\mathbf{X}}'\mathbf{W}_r'(\mathbf{I}_n - n^{-1}\underline{\mathbf{X}}\mathbf{\Pi}\mathbf{T}')\mathbf{\Sigma}(\mathbf{I}_n - n^{-1}\mathbf{T}\mathbf{\Pi}'\underline{\mathbf{X}}')\mathbf{W}_s\underline{\mathbf{X}}\mathbf{\Pi}\boldsymbol{\theta}, \\
\boldsymbol{\phi}_{rs}^{YZ} &= \left[2\text{tr}(\overline{\mathbf{W}}_r\mathbf{\Sigma}\mathbf{W}_s\mathbf{\Sigma}_l) + \textstyle\sum_{k=1}^K \theta_k \text{tr}(\mathbf{W}_r\mathbf{\Sigma}_k\mathbf{W}_s\mathbf{\Sigma}_l + \mathbf{W}_r\mathbf{\Sigma}_{kl}\mathbf{W}_s'\mathbf{\Sigma})\right]_{l=1,\cdots,K} \\
&\quad + \boldsymbol{\theta}'\mathbf{\Pi}'\underline{\mathbf{X}}'\mathbf{W}_r'(\mathbf{I}_n - n^{-1}\underline{\mathbf{X}}\mathbf{\Pi}\mathbf{T}')\mathbf{\Sigma}(\mathbf{I}_n - n^{-1}\mathbf{T}\mathbf{\Pi}'\underline{\mathbf{X}}')\mathbf{W}_s\underline{\mathbf{X}}\mathbf{\Pi}, \\
\phi_{rs}^{YU} &= 2\text{tr}(\overline{\mathbf{W}}_r\mathbf{\Sigma}\overline{\mathbf{W}}_s\mathbf{\Sigma}) + 2\textstyle\sum_{k=1}^K \theta_k \text{tr}(\mathbf{W}_r\mathbf{\Sigma}_k\overline{\mathbf{W}}_s\mathbf{\Sigma}), \\
\boldsymbol{\phi}_{rs}^{ZZ} &= [\text{tr}(\mathbf{W}_r\mathbf{\Sigma}_k\mathbf{W}_s\mathbf{\Sigma}_l + \mathbf{W}_r\mathbf{\Sigma}_{kl}\mathbf{W}_s'\mathbf{\Sigma})]_{k,l=1,\cdots,K} \\
&\quad + \mathbf{\Pi}'\underline{\mathbf{X}}'\mathbf{W}_r'(\mathbf{I}_n - n^{-1}\underline{\mathbf{X}}\mathbf{\Pi}\mathbf{T}')\mathbf{\Sigma}(\mathbf{I}_n - n^{-1}\mathbf{T}\mathbf{\Pi}'\underline{\mathbf{X}}')\mathbf{W}_s\underline{\mathbf{X}}\mathbf{\Pi}, \\
\boldsymbol{\phi}_{rs}^{ZU} &= \left[2\text{tr}(\mathbf{W}_r\mathbf{\Sigma}_k\overline{\mathbf{W}}_s\mathbf{\Sigma})\right]_{k=1,\cdots,K}, \\
\phi_{rs}^{UU} &= 2\text{tr}(\overline{\mathbf{W}}_r\mathbf{\Sigma}\overline{\mathbf{W}}_s\mathbf{\Sigma}).
\end{aligned}$$

The estimators of $\widehat{\phi}_{rs}^{YY}$, $\widehat{\boldsymbol{\phi}}_{rs}^{YZ}$, $\widehat{\phi}_{rs}^{YU}$, $\widehat{\boldsymbol{\phi}}_{rs}^{ZZ}$, $\widehat{\boldsymbol{\phi}}_{rs}^{ZU}$, and $\widehat{\phi}_{rs}^{UU}$ defined in the theorem are obtained by making the following replacements in the above expressions:$\boldsymbol{\theta} \to \widehat{\boldsymbol{\theta}}$, $\mathbf{\Sigma} \to \widehat{\mathbf{\Sigma}}$, $\mathbf{\Sigma}_k \to \widehat{\mathbf{\Sigma}}_k$, $\mathbf{\Sigma}_{kl} \to \widehat{\mathbf{\Sigma}}_{kl}$, $\mathbf{\Pi} \to \widehat{\mathbf{\Pi}} = (\underline{\mathbf{X}}'\underline{\mathbf{X}})^{-\mathbf{1}}\underline{\mathbf{X}}'\mathbf{Z}$, $\underline{\mathbf{X}}\mathbf{\Pi} \to \underline{\mathbf{X}}\widehat{\mathbf{\Pi}} = \widehat{\mathbf{Z}}$, $\mathbf{T} \to \widehat{\mathbf{T}} = \widehat{\mathbf{Z}}(n^{-1}\widehat{\mathbf{Z}}'\widehat{\mathbf{Z}})^{-1}$. Given the maintained assumptions it follows from a similar argument as in Lemmata C.3-C.5 in Kelejian and Prucha (2010) that

$$\begin{aligned}
n^{-1}\text{tr}(\overline{\mathbf{W}}_r\widehat{\mathbf{\Sigma}}\overline{\mathbf{W}}_s\widehat{\mathbf{\Sigma}}) &= n^{-1}\text{tr}(\overline{\mathbf{W}}_r\mathbf{\Sigma}\overline{\mathbf{W}}_s\mathbf{\Sigma}) + o_p(1) = O_p(1), \\
n^{-1}\text{tr}(\overline{\mathbf{W}}_r\widehat{\mathbf{\Sigma}}\mathbf{W}_s\widehat{\mathbf{\Sigma}}_k) &= n^{-1}\text{tr}(\overline{\mathbf{W}}_r\mathbf{\Sigma}\mathbf{W}_s\mathbf{\Sigma}_k) + o_p(1) = O_p(1), \\
n^{-1}\text{tr}(\mathbf{W}_r\widehat{\mathbf{\Sigma}}_k\mathbf{W}_s\widehat{\mathbf{\Sigma}}_l) &= n^{-1}\text{tr}(\mathbf{W}_r\mathbf{\Sigma}_k\mathbf{W}_s\mathbf{\Sigma}_l) + o_p(1) = O_p(1), \\
n^{-1}\text{tr}(\mathbf{W}_r\widehat{\mathbf{\Sigma}}_{kl}\mathbf{W}_s'\widehat{\mathbf{\Sigma}}) &= n^{-1}\text{tr}(\mathbf{W}_r\mathbf{\Sigma}_{kl}\mathbf{W}_s'\mathbf{\Sigma}) + o_p(1) = O_p(1), \\
n^{-1}\underline{\mathbf{X}}'\widehat{\mathbf{\Sigma}}\underline{\mathbf{X}} &= n^{-1}\underline{\mathbf{X}}'\mathbf{\Sigma}\underline{\mathbf{X}} + o_p(1) = O_p(1), \\
n^{-1}\underline{\mathbf{X}}'\mathbf{W}_r'\widehat{\mathbf{\Sigma}}\underline{\mathbf{X}} &= n^{-1}\underline{\mathbf{X}}'\mathbf{W}_r'\mathbf{\Sigma}\underline{\mathbf{X}} + o_p(1) = O_p(1), \\
n^{-1}\underline{\mathbf{X}}'\mathbf{W}_r'\widehat{\mathbf{\Sigma}}\mathbf{W}_s\underline{\mathbf{X}} &= n^{-1}\underline{\mathbf{X}}'\mathbf{W}_r\mathbf{\Sigma}\mathbf{W}_s\underline{\mathbf{X}} + o_p(1) = O_p(1).
\end{aligned}$$

The consistency of $n^{-1}\widehat{\mathbf{\Phi}}$ now follows from standard argumentation. □

B Supplementary Appendix

B.1 Proofs of Propositions

Proof of Proposition 1. Let $\mathbf{\Gamma} = (\mathbf{I}_q \otimes \widehat{\boldsymbol{\beta}}', \mathbf{I}_q)'$, then in light of (8) and (10) we have

$$\widehat{\mathbf{V}}_* = \mathrm{L}\widehat{\mathbf{V}} = \begin{bmatrix} \widehat{\mathbf{V}}^X \\ \widehat{\mathbf{V}}^U \end{bmatrix} \qquad \text{and} \qquad \widehat{\mathbf{\Phi}}_* = \mathrm{L}\widehat{\mathbf{\Phi}}\mathrm{L}' = \begin{bmatrix} \widehat{\mathbf{\Phi}}^{XX} & 0 \\ 0 & \widehat{\mathbf{\Phi}}^{UU} \end{bmatrix},$$

and

$$\widehat{\mathbf{V}} = \begin{bmatrix} \mathbf{\Gamma}'\widehat{\mathbf{V}}_* \\ \widehat{\mathbf{V}}_* \end{bmatrix} \qquad \text{and} \qquad \widehat{\mathbf{\Phi}} = \begin{bmatrix} \mathbf{\Gamma}'\widehat{\mathbf{\Phi}}_*\mathbf{\Gamma} & \mathbf{\Gamma}'\widehat{\mathbf{\Phi}}_* \\ \widehat{\mathbf{\Phi}}_*\mathbf{\Gamma} & \widehat{\mathbf{\Phi}}_* \end{bmatrix}.$$

Next observe that the Moore-Penrose generalized inverse of $\widehat{\mathbf{\Phi}}$ is, as is readily checked, given by

$$\widehat{\mathbf{\Phi}}^+ = \begin{bmatrix} \mathbf{\Gamma}'\mathbf{A}\mathbf{\Gamma} & \mathbf{\Gamma}'\mathbf{A} \\ \mathbf{A}\mathbf{\Gamma} & \mathbf{A} \end{bmatrix},$$

where $\mathbf{A} = (\mathbf{I} + \mathbf{\Gamma}\mathbf{\Gamma}')^{-1}\widehat{\mathbf{\Phi}}_*^{-1}(\mathbf{I} + \mathbf{\Gamma}\mathbf{\Gamma}')^{-1}$. Consequently,

$$\begin{aligned} \widehat{\mathbf{V}}'\widehat{\mathbf{\Phi}}^+\widehat{\mathbf{V}} &= \begin{bmatrix} \mathbf{\Gamma}'\widehat{\mathbf{V}}_* \\ \widehat{\mathbf{V}}_* \end{bmatrix}' \begin{bmatrix} \mathbf{\Gamma}'\mathbf{A}\mathbf{\Gamma} & \mathbf{\Gamma}'\mathbf{A} \\ \mathbf{A}\mathbf{\Gamma} & \mathbf{A} \end{bmatrix} \begin{bmatrix} \mathbf{\Gamma}'\widehat{\mathbf{V}}_* \\ \widehat{\mathbf{V}}_* \end{bmatrix} \\ &= \widehat{\mathbf{V}}_*'(\mathbf{I} + \mathbf{\Gamma}\mathbf{\Gamma}')\mathbf{A}(\mathbf{I} + \mathbf{\Gamma}\mathbf{\Gamma}')\widehat{\mathbf{V}}_* \\ &= \widehat{\mathbf{V}}_*'\widehat{\mathbf{\Phi}}_*^{-1}\widehat{\mathbf{V}}_*, \end{aligned}$$

which proves the claim. □

In the following proofs, we assume the disturbance process of models (11) and (12) is specified as a more general spatial ARMA$(\overline{q}, \underline{q})$ process, where $\mathbf{u} = \sum_{r=1}^{\overline{q}} \rho_r \mathbf{W}_r \mathbf{u} + \boldsymbol{\varepsilon} + \sum_{r=\overline{q}+1}^{q} \rho_r \mathbf{W}_r \boldsymbol{\varepsilon}$ and $q = \overline{q} + \underline{q}$.

Proof of Proposition 2. Let $\mathbf{R} = (\mathbf{I}_n + \rho_{\overline{q}+1}\mathbf{W}_{\overline{q}+1} + \cdots + \rho_q \mathbf{W}_q)^{-1}(\mathbf{I}_n - \rho_1 \mathbf{W}_1 - \cdots - \rho_{\overline{q}}\mathbf{W}_{\overline{q}})$. If $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_n)$, under the stated assumptions, $\mathbf{u} \sim N(\mathbf{0}, \mathbf{\Omega}_u)$ and $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{\Omega}_u)$, where $\mathbf{\Omega}_u = \sigma^2\mathbf{R}^{-1}\mathbf{R}'^{-1}$. The log-likelihood function is given by

$$\ln L(\delta) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\det(\mathbf{\Omega}_u)| - \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{\Omega}_u^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \tag{B.1}$$

where $\boldsymbol{\delta} = [\boldsymbol{\rho}', \boldsymbol{\beta}', \sigma^2]'$. The first-order derivatives of the log-likelihood function (B.1) are

$$
\begin{aligned}
\frac{\partial \ln L}{\partial \rho_r} &= \mathrm{tr}[\frac{\partial \mathbf{R}}{\partial \rho_r}\mathbf{R}^{-1}] - \frac{1}{\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\mathbf{R}'\frac{\partial \mathbf{R}}{\partial \rho_r}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}),\\
\frac{\partial \ln L}{\partial \boldsymbol{\beta}} &= \frac{1}{\sigma^2}\mathbf{X}'\mathbf{R}'\mathbf{R}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}),\\
\frac{\partial \ln L}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\mathbf{R}'\mathbf{R}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}).
\end{aligned}
$$

Evaluated at the true parameter value, the mean of the second-order derivatives of the log-likelihood function (B.1) are

$$
\begin{aligned}
\mathrm{E}\frac{\partial^2 \ln L}{\partial \rho_r \partial \rho_s} &= -\mathrm{tr}[\frac{\partial \mathbf{R}}{\partial \rho_r}\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \rho_s}\mathbf{R}^{-1}] - \mathrm{tr}[(\frac{\partial \mathbf{R}}{\partial \rho_r}\mathbf{R}^{-1})'\frac{\partial \mathbf{R}}{\partial \rho_s}\mathbf{R}^{-1}],\\
\mathrm{E}\frac{\partial^2 \ln L}{\partial \rho_s \partial \boldsymbol{\beta}} &= \mathbf{0},\\
\mathrm{E}\frac{\partial^2 \ln L}{\partial \rho_s \partial \sigma^2} &= \frac{1}{\sigma^2}\mathrm{tr}[\frac{\partial \mathbf{R}}{\partial \rho_s}\mathbf{R}^{-1}].
\end{aligned}
$$

Observe that when $\boldsymbol{\rho} = \mathbf{0}$, $\mathbf{R} = \mathbf{I}_n$ and $\frac{\partial}{\partial \rho_r}\mathbf{R} = -\mathbf{W}_r$. Evaluated at the restricted ML estimator $\widetilde{\boldsymbol{\delta}}$,

$$
\frac{\partial \ln L}{\partial \rho_r}|_{\widetilde{\boldsymbol{\delta}}} = \frac{1}{\widehat{\sigma}^2}\widetilde{\mathbf{u}}'\mathbf{W}_r\widetilde{\mathbf{u}}, \qquad \frac{\partial \ln L}{\partial \boldsymbol{\beta}}|_{\widetilde{\boldsymbol{\delta}}} = \mathbf{0}, \qquad \frac{\partial \ln L}{\partial \sigma^2}|_{\widetilde{\boldsymbol{\delta}}} = 0,
$$

and

$$
\mathrm{E}\frac{\partial^2 \ln L}{\partial \rho_r \partial \rho_s}|_{\widetilde{\boldsymbol{\delta}}} = -2\mathrm{tr}(\overline{\mathbf{W}}_r\overline{\mathbf{W}}_s), \qquad \mathrm{E}\frac{\partial^2 \ln L}{\partial \rho_s \partial \boldsymbol{\beta}}|_{\widetilde{\boldsymbol{\delta}}} = \mathbf{0}, \qquad \mathrm{E}\frac{\partial^2 \ln L}{\partial \rho_s \partial \sigma^2}|_{\widetilde{\boldsymbol{\delta}}} = 0.
$$

Then, the LM test statistic is given by

$$
\begin{aligned}
\mathrm{LM}_u &= [\frac{\partial \ln L}{\partial \boldsymbol{\delta}'}]_{\widetilde{\boldsymbol{\delta}}}[-\mathrm{E}\frac{\partial^2 \ln L}{\partial \boldsymbol{\delta}\partial \boldsymbol{\delta}'}]_{\widetilde{\boldsymbol{\delta}}}^{-1}[\frac{\partial \ln L}{\partial \boldsymbol{\delta}}]_{\widetilde{\boldsymbol{\delta}}}\\
&= \begin{bmatrix} \frac{\partial \ln L}{\partial \boldsymbol{\rho}}|_{\widetilde{\boldsymbol{\delta}}} \\ \mathbf{0}_{K_x \times 1} \\ 0 \end{bmatrix}' \begin{bmatrix} -\mathrm{E}\frac{\partial^2 \ln L}{\partial \boldsymbol{\rho}\partial \boldsymbol{\rho}'}|_{\widetilde{\boldsymbol{\delta}}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathrm{E}\frac{\partial^2 \ln L}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'}|_{\widetilde{\boldsymbol{\delta}}} & -\mathrm{E}\frac{\partial^2 \ln L}{\partial \boldsymbol{\beta}\partial \sigma^2}|_{\widetilde{\boldsymbol{\delta}}} \\ \mathbf{0} & -\mathrm{E}\frac{\partial^2 \ln L}{\partial \boldsymbol{\beta}'\partial \sigma^2}|_{\widetilde{\boldsymbol{\delta}}} & -\mathrm{E}\frac{\partial^2 \ln L}{\partial (\sigma^2)^2}|_{\widetilde{\boldsymbol{\delta}}} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial \ln L}{\partial \boldsymbol{\rho}}|_{\widetilde{\boldsymbol{\delta}}} \\ \mathbf{0}_{K_x \times 1} \\ 0 \end{bmatrix}\\
&= [\frac{\partial \ln L}{\partial \boldsymbol{\rho}}]'_{\widetilde{\boldsymbol{\delta}}}[-\mathrm{E}\frac{\partial^2 \ln L}{\partial \boldsymbol{\rho}\partial \boldsymbol{\rho}'}]_{\widetilde{\boldsymbol{\delta}}}^{-1}[\frac{\partial \ln L}{\partial \boldsymbol{\rho}}]_{\widetilde{\boldsymbol{\delta}}}\\
&= \widetilde{\mathbf{V}}^{U\prime}(\widetilde{\boldsymbol{\Phi}}^{UU})^{-1}\widetilde{\mathbf{V}}^{U},
\end{aligned}
$$

which proves the claim. □

Proof of Proposition 3. Let $\mathbf{R} = (\mathbf{I}_n + \rho_{\overline{q}+1}\mathbf{W}_{\overline{q}+1} + \cdots + \rho_q \mathbf{W}_q)^{-1}(\mathbf{I}_n - \rho_1 \mathbf{W}_1 - \cdots - \rho_{\overline{q}}\mathbf{W}_{\overline{q}})$ and $\mathbf{S} = \mathbf{I}_n - \lambda_1 \mathbf{W}_1 - \cdots - \lambda_q \mathbf{W}_q$. If $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$, under the stated assumptions, $\mathbf{y} \sim N(\mathbf{S}^{-1}\boldsymbol{\mu}_y, \boldsymbol{\Omega}_y)$, where $\boldsymbol{\mu}_y = \mathbf{X}\boldsymbol{\beta} + \sum_{r=1}^q \mathbf{W}_r \mathbf{X}\boldsymbol{\gamma}_r$ and $\boldsymbol{\Omega}_y = \mathbf{S}^{-1}\boldsymbol{\Omega}_u \mathbf{S}'^{-1}$. The log-likelihood function is given by

$$\ln L(\theta) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\det(\boldsymbol{\Omega}_y)| - \frac{1}{2}(\mathbf{y} - \mathbf{S}^{-1}\boldsymbol{\mu}_y)'\boldsymbol{\Omega}_y^{-1}(\mathbf{y} - \mathbf{S}^{-1}\boldsymbol{\mu}_y), \tag{B.2}$$

where $\boldsymbol{\theta} = [\boldsymbol{\rho}', \boldsymbol{\lambda}', \boldsymbol{\gamma}_1', \cdots, \boldsymbol{\gamma}_q', \boldsymbol{\beta}', \sigma^2]'$. The first-order derivatives of the log-likelihood function (B.2) are

$$\begin{aligned}
\frac{\partial \ln L}{\partial \rho_r} &= \mathrm{tr}[\frac{\partial \mathbf{R}}{\partial \rho_r}\mathbf{R}^{-1}] - \frac{1}{\sigma^2}(\mathbf{Sy} - \mathbf{X}\boldsymbol{\beta} - \textstyle\sum_{s=1}^q \mathbf{W}_s \mathbf{X}\boldsymbol{\gamma}_s)'\mathbf{R}'\frac{\partial \mathbf{R}}{\partial \rho_r}(\mathbf{Sy} - \mathbf{X}\boldsymbol{\beta} - \textstyle\sum_{s=1}^q \mathbf{W}_s \mathbf{X}\boldsymbol{\gamma}_s) \\
\frac{\partial \ln L}{\partial \lambda_r} &= -\mathrm{tr}[\mathbf{W}_r \mathbf{S}^{-1}] + \frac{1}{\sigma^2}(\mathbf{Sy} - \mathbf{X}\boldsymbol{\beta} - \textstyle\sum_{s=1}^q \mathbf{W}_s \mathbf{X}\boldsymbol{\gamma}_s)'\mathbf{R}'\mathbf{R}\mathbf{W}_r \mathbf{y} \\
\frac{\partial \ln L}{\partial \boldsymbol{\gamma}_r} &= \frac{1}{\sigma^2}\mathbf{X}'\mathbf{W}_r'\mathbf{R}'\mathbf{R}(\mathbf{Sy} - \mathbf{X}\boldsymbol{\beta} - \textstyle\sum_{s=1}^q \mathbf{W}_s \mathbf{X}\boldsymbol{\gamma}_s) \\
\frac{\partial \ln L}{\partial \boldsymbol{\beta}} &= \frac{1}{\sigma^2}\mathbf{X}'\mathbf{R}'\mathbf{R}(\mathbf{Sy} - \mathbf{X}\boldsymbol{\beta} - \textstyle\sum_{s=1}^q \mathbf{W}_s \mathbf{X}\boldsymbol{\gamma}_s) \\
\frac{\partial \ln L}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(\mathbf{Sy} - \mathbf{X}\boldsymbol{\beta} - \textstyle\sum_{s=1}^q \mathbf{W}_s \mathbf{X}\boldsymbol{\gamma}_s)'\mathbf{R}'\mathbf{R}(\mathbf{Sy} - \mathbf{X}\boldsymbol{\beta} - \textstyle\sum_{s=1}^q \mathbf{W}_s \mathbf{X}\boldsymbol{\gamma}_s).
\end{aligned}$$

Evaluated at the true parameter value, the mean of the second-order derivatives of the log-likelihood function (B.2) are

$$\begin{aligned}
\mathrm{E}\frac{\partial^2 \ln L}{\partial \rho_r \partial \rho_s} &= -\mathrm{tr}[\frac{\partial \mathbf{R}}{\partial \rho_r}\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \rho_s}\mathbf{R}^{-1}] - \mathrm{tr}[(\frac{\partial \mathbf{R}}{\partial \rho_r}\mathbf{R}^{-1})'\frac{\partial \mathbf{R}}{\partial \rho_s}\mathbf{R}^{-1}], \\
\mathrm{E}\frac{\partial^2 \ln L}{\partial \rho_r \partial \lambda_s} &= \mathrm{tr}[\frac{\partial \mathbf{R}}{\partial \rho_r}\mathbf{W}_s \mathbf{S}^{-1}\mathbf{R}^{-1}] + \mathrm{tr}[\frac{\partial \mathbf{R}'}{\partial \rho_r}\mathbf{R}\mathbf{W}_s \mathbf{S}^{-1}(\mathbf{R}'\mathbf{R})^{-1}], \\
\mathrm{E}\frac{\partial^2 \ln L}{\partial \rho_r \partial \boldsymbol{\gamma}_s} &= \mathrm{E}\frac{\partial^2 \ln L}{\partial \rho_r \partial \boldsymbol{\beta}} = \mathbf{0}, \\
\mathrm{E}\frac{\partial^2 \ln L}{\partial \rho_r \partial \sigma^2} &= \frac{1}{\sigma^2}\mathrm{tr}[\frac{\partial \mathbf{R}}{\partial \rho_r}\mathbf{R}^{-1}],
\end{aligned}$$

$$
\begin{aligned}
\mathrm{E}\frac{\partial^2 \ln L}{\partial\lambda_r\partial\lambda_s} &= -\mathrm{tr}[\mathbf{W}_r\mathbf{S}^{-1}\mathbf{W}_s\mathbf{S}^{-1}] - \mathrm{tr}[(\mathbf{R}\mathbf{W}_r\mathbf{S}^{-1}\mathbf{R}^{-1})'\mathbf{R}\mathbf{W}_s\mathbf{S}^{-1}\mathbf{R}^{-1}] \\
&\quad -\frac{1}{\sigma^2}[\mathbf{R}\mathbf{W}_r\mathbf{S}^{-1}(\mathbf{X}\boldsymbol{\beta} + \textstyle\sum_{r=1}^q \mathbf{W}_r\mathbf{X}\boldsymbol{\gamma}_r)]'[\mathbf{R}\mathbf{W}_s\mathbf{S}^{-1}(\mathbf{X}\boldsymbol{\beta} + \textstyle\sum_{r=1}^q \mathbf{W}_r\mathbf{X}\boldsymbol{\gamma}_r)], \\
\mathrm{E}\frac{\partial^2 \ln L}{\partial\lambda_r\partial\boldsymbol{\gamma}_s} &= -\frac{1}{\sigma^2}\mathbf{X}'\mathbf{W}_s'\mathbf{R}'\mathbf{R}\mathbf{W}_r\mathbf{S}^{-1}(\mathbf{X}\boldsymbol{\beta} + \textstyle\sum_{r=1}^q \mathbf{W}_r\mathbf{X}\boldsymbol{\gamma}_r), \\
\mathrm{E}\frac{\partial^2 \ln L}{\partial\lambda_r\partial\boldsymbol{\beta}} &= -\frac{1}{\sigma^2}\mathbf{X}'\mathbf{R}'\mathbf{R}\mathbf{W}_r\mathbf{S}^{-1}(\mathbf{X}\boldsymbol{\beta} + \textstyle\sum_{r=1}^q \mathbf{W}_r\mathbf{X}\boldsymbol{\gamma}_r), \\
\mathrm{E}\frac{\partial^2 \ln L}{\partial\lambda_r\partial\sigma^2} &= -\frac{1}{\sigma^2}\mathrm{tr}[\mathbf{R}\mathbf{W}_r\mathbf{S}^{-1}\mathbf{R}^{-1}],
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{E}\frac{\partial^2 \ln L}{\partial\boldsymbol{\gamma}_r\partial\boldsymbol{\gamma}_s'} &= -\frac{1}{\sigma^2}\mathbf{X}'\mathbf{W}_r'\mathbf{R}'\mathbf{R}\mathbf{W}_s\mathbf{X}, \\
\mathrm{E}\frac{\partial^2 \ln L}{\partial\boldsymbol{\beta}\partial\boldsymbol{\gamma}_r'} &= -\frac{1}{\sigma^2}\mathbf{X}'\mathbf{R}'\mathbf{R}\mathbf{W}_r\mathbf{X}, \\
\mathrm{E}\frac{\partial^2 \ln L}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} &= -\frac{1}{\sigma^2}\mathbf{X}'\mathbf{R}'\mathbf{R}\mathbf{X}, \\
\mathrm{E}\frac{\partial^2 \ln L}{\partial\boldsymbol{\gamma}_r\partial\sigma^2} &= \mathrm{E}\frac{\partial^2 \ln L}{\partial\boldsymbol{\beta}\partial\sigma^2} = \mathbf{0}, \\
\mathrm{E}\frac{\partial^2 \ln L}{\partial(\sigma^2)^2} &= -\frac{n}{2\sigma^4}.
\end{aligned}
$$

Observe that when $\boldsymbol{\lambda} = \boldsymbol{\rho} = \mathbf{0}$, $\mathbf{S} = \mathbf{R} = \mathbf{I}_n$ and $\frac{\partial}{\partial\rho_r}\mathbf{R} = -\mathbf{W}_r$ for $r = 1, \cdots, q$. Evaluated at the restricted ML estimator $\widehat{\boldsymbol{\theta}}$, we have $\frac{\partial \ln L}{\partial\rho_r}|_{\widehat{\boldsymbol{\theta}}} = \frac{1}{\widehat{\sigma}^2}\widehat{\mathbf{u}}'W_r\widehat{\mathbf{u}}$, $\frac{\partial \ln L}{\partial\lambda_r}|_{\widehat{\boldsymbol{\theta}}} = \frac{1}{\widehat{\sigma}^2}\mathbf{y}'W_r'\widehat{\mathbf{u}}$, $\frac{\partial \ln L}{\partial\boldsymbol{\gamma}_r}|_{\widehat{\boldsymbol{\theta}}} = \frac{1}{\widehat{\sigma}^2}\mathbf{X}'W_r'\widehat{\mathbf{u}}$,

$\frac{\partial \ln L}{\partial \boldsymbol{\beta}}|_{\widehat{\boldsymbol{\theta}}} = \mathbf{0}$, $\frac{\partial \ln L}{\partial \sigma^2}|_{\widehat{\boldsymbol{\theta}}} = 0$, and

$$\begin{aligned}
\mathrm{E}\frac{\partial^2 \ln L}{\partial \rho_r \partial \rho_s}|_{\widehat{\boldsymbol{\theta}}} &= \mathrm{E}\frac{\partial^2 \ln L}{\partial \rho_r \partial \lambda_s}|_{\widehat{\boldsymbol{\theta}}} = -2\mathrm{tr}(\overline{\mathbf{W}}_r \overline{\mathbf{W}}_s),\\
\mathrm{E}\frac{\partial^2 \ln L}{\partial \lambda_r \partial \lambda_s}|_{\widehat{\boldsymbol{\theta}}} &= -2\mathrm{tr}(\overline{\mathbf{W}}_r \overline{\mathbf{W}}_s) - \frac{1}{\widehat{\sigma}^2}\widehat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{W}_r'\mathbf{W}_s\mathbf{X}\widehat{\boldsymbol{\beta}},\\
\mathrm{E}\frac{\partial^2 \ln L}{\partial \lambda_r \partial \boldsymbol{\gamma}_s}|_{\widehat{\boldsymbol{\theta}}} &= -\frac{1}{\widehat{\sigma}^2}\mathbf{X}'\mathbf{W}_s'\mathbf{W}_r\mathbf{X}\widehat{\boldsymbol{\beta}},\\
\mathrm{E}\frac{\partial^2 \ln L}{\partial \lambda_r \partial \boldsymbol{\beta}}|_{\widehat{\boldsymbol{\theta}}} &= -\frac{1}{\widehat{\sigma}^2}\mathbf{X}'\mathbf{W}_r\mathbf{X}\widehat{\boldsymbol{\beta}},\\
\mathrm{E}\frac{\partial^2 \ln L}{\partial \boldsymbol{\gamma}_r \partial \boldsymbol{\gamma}_s'}|_{\widehat{\boldsymbol{\theta}}} &= -\frac{1}{\widehat{\sigma}^2}\mathbf{X}'\mathbf{W}_r'\mathbf{W}_s\mathbf{X},\\
\mathrm{E}\frac{\partial^2 \ln L}{\partial \boldsymbol{\beta} \partial \boldsymbol{\gamma}_r'}|_{\widehat{\boldsymbol{\theta}}} &= -\frac{1}{\widehat{\sigma}^2}\mathbf{X}'\mathbf{W}_r\mathbf{X},\\
\mathrm{E}\frac{\partial^2 \ln L}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}|_{\widehat{\boldsymbol{\theta}}} &= -\frac{1}{\widehat{\sigma}^2}\mathbf{X}'\mathbf{X},\\
\mathrm{E}\frac{\partial^2 \ln L}{\partial (\sigma^2)^2}|_{\widehat{\boldsymbol{\theta}}} &= -\frac{n}{2(\widehat{\sigma}^2)^2},
\end{aligned}$$

$$\mathrm{E}\frac{\partial^2 \ln L}{\partial \rho_r \partial \gamma_{k,s}}|_{\widehat{\boldsymbol{\theta}}} = \mathrm{E}\frac{\partial^2 \ln L}{\partial \rho_r \partial \beta_k}|_{\widehat{\boldsymbol{\theta}}} = \mathrm{E}\frac{\partial^2 \ln L}{\partial \rho_r \partial \sigma^2}|_{\widehat{\boldsymbol{\theta}}} = \mathrm{E}\frac{\partial^2 \ln L}{\partial \lambda_r \partial \sigma^2}|_{\widehat{\boldsymbol{\theta}}} = \mathrm{E}\frac{\partial^2 \ln L}{\partial \gamma_{k,r} \partial \sigma^2}|_{\widehat{\boldsymbol{\theta}}} = \mathrm{E}\frac{\partial^2 \ln L}{\partial \beta_k \partial \sigma^2}|_{\widehat{\boldsymbol{\theta}}} = 0.$$

Let

$$\begin{aligned}
\mathbf{A}_{11} &= -\begin{bmatrix}
\mathrm{E}\frac{\partial^2 \ln L}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}'}|_{\widehat{\boldsymbol{\theta}}} & \mathrm{E}\frac{\partial^2 \ln L}{\partial \boldsymbol{\lambda} \partial \partial \boldsymbol{\gamma}_1'}|_{\widehat{\boldsymbol{\theta}}} & \cdots & \mathrm{E}\frac{\partial^2 \ln L}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\gamma}_q'}|_{\widehat{\boldsymbol{\theta}}} & \mathrm{E}\frac{\partial^2 \ln L}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\rho}'}|_{\widehat{\boldsymbol{\theta}}}\\
\mathrm{E}\frac{\partial^2 \ln L}{\partial \partial \boldsymbol{\gamma}_1 \partial \boldsymbol{\lambda}'}|_{\widehat{\boldsymbol{\theta}}} & \mathrm{E}\frac{\partial^2 \ln L}{\partial \boldsymbol{\gamma}_1 \partial \boldsymbol{\gamma}_1'}|_{\widehat{\boldsymbol{\theta}}} & \cdots & \mathrm{E}\frac{\partial^2 \ln L}{\partial \boldsymbol{\gamma}_1 \partial \boldsymbol{\gamma}_q'}|_{\widehat{\boldsymbol{\theta}}} & \mathbf{0}\\
\vdots & \vdots & \ddots & \vdots & \vdots\\
\mathrm{E}\frac{\partial^2 \ln L}{\partial \boldsymbol{\gamma}_q \partial \boldsymbol{\lambda}'}|_{\widehat{\boldsymbol{\theta}}} & \mathrm{E}\frac{\partial^2 \ln L}{\partial \boldsymbol{\gamma}_q \partial \boldsymbol{\gamma}_1'}|_{\widehat{\boldsymbol{\theta}}} & \cdots & \mathrm{E}\frac{\partial^2 \ln L}{\partial \boldsymbol{\gamma}_q \partial \boldsymbol{\gamma}_q'}|_{\widehat{\boldsymbol{\theta}}} & \mathbf{0}\\
\mathrm{E}\frac{\partial^2 \ln L}{\partial \boldsymbol{\rho} \partial \boldsymbol{\lambda}'}|_{\widehat{\boldsymbol{\theta}}} & \mathbf{0} & \cdots & \mathbf{0} & \mathrm{E}\frac{\partial^2 \ln L}{\partial \boldsymbol{\rho} \partial \boldsymbol{\rho}'}|_{\widehat{\boldsymbol{\theta}}}
\end{bmatrix}\\
\mathbf{A}_{21} &= -[\mathrm{E}\frac{\partial^2 \ln L}{\partial \boldsymbol{\beta} \partial \boldsymbol{\lambda}'}|_{\widehat{\boldsymbol{\theta}}}, \mathrm{E}\frac{\partial^2 \ln L}{\partial \boldsymbol{\beta} \partial \boldsymbol{\gamma}_1'}|_{\widehat{\boldsymbol{\theta}}}, \cdots, \mathrm{E}\frac{\partial^2 \ln L}{\partial \boldsymbol{\beta} \partial \boldsymbol{\gamma}_q'}|_{\widehat{\boldsymbol{\theta}}}, \mathbf{0}]\\
\mathbf{A}_{22} &= -\mathrm{E}\frac{\partial^2 \ln L}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}|_{\widehat{\boldsymbol{\theta}}},
\end{aligned}$$

then, the LM test statistic is given by

$$\begin{aligned}
\mathrm{LM}_y &= [\frac{\partial \ln L}{\partial \boldsymbol{\theta}'}]_{\widehat{\boldsymbol{\theta}}}[-\mathrm{E}\frac{\partial^2 \ln L}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}]^{+}_{\widehat{\boldsymbol{\theta}}}[\frac{\partial \ln L}{\partial \boldsymbol{\theta}}]_{\widehat{\boldsymbol{\theta}}} \\
&= \begin{bmatrix} \frac{\partial \ln L}{\partial \lambda}|_{\widehat{\boldsymbol{\theta}}} \\ \frac{\partial \ln L}{\partial \boldsymbol{\gamma}_1}|_{\widehat{\boldsymbol{\theta}}} \\ \vdots \\ \frac{\partial \ln L}{\partial \boldsymbol{\gamma}_q}|_{\widehat{\boldsymbol{\theta}}} \\ \frac{\partial \ln L}{\partial \boldsymbol{\rho}}|_{\widehat{\boldsymbol{\theta}}} \\ \mathbf{0}_{K_x \times 1} \\ 0 \end{bmatrix}' \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{21}' & 0 \\ \mathbf{A}_{21} & \mathbf{A}_{22} & 0 \\ 0 & 0 & -\mathrm{E}\frac{\partial^2 \ln L}{\partial(\sigma^2)^2}|_{\widehat{\boldsymbol{\theta}}} \end{bmatrix}^{+} \begin{bmatrix} \frac{\partial \ln L}{\partial \lambda}|_{\widehat{\boldsymbol{\theta}}} \\ \frac{\partial \ln L}{\partial \boldsymbol{\gamma}_1}|_{\widehat{\boldsymbol{\theta}}} \\ \vdots \\ \frac{\partial \ln L}{\partial \boldsymbol{\gamma}_q}|_{\widehat{\boldsymbol{\theta}}} \\ \frac{\partial \ln L}{\partial \boldsymbol{\rho}}|_{\widehat{\boldsymbol{\theta}}} \\ \mathbf{0}_{K_x \times 1} \\ 0 \end{bmatrix}.
\end{aligned}$$

Let $\widehat{\mathbf{V}}$ and $\widehat{\boldsymbol{\Phi}}$ be defined as in (8), then using results on the generalized inverse of partitioned matrices given in, e.g., Trenkler and Schipp (1993), we have

$$\mathrm{LM}_y = \begin{bmatrix} \frac{\partial \ln L}{\partial \lambda}|_{\widehat{\boldsymbol{\theta}}} \\ \frac{\partial \ln L}{\partial \boldsymbol{\gamma}_1}|_{\widehat{\boldsymbol{\theta}}} \\ \vdots \\ \frac{\partial \ln L}{\partial \boldsymbol{\gamma}_q}|_{\widehat{\boldsymbol{\theta}}} \\ \frac{\partial \ln L}{\partial \boldsymbol{\rho}}|_{\widehat{\boldsymbol{\theta}}} \end{bmatrix}' \begin{bmatrix} (\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{+} \end{bmatrix} \begin{bmatrix} \frac{\partial \ln L}{\partial \lambda}|_{\widehat{\boldsymbol{\theta}}} \\ \frac{\partial \ln L}{\partial \boldsymbol{\gamma}_1}|_{\widehat{\boldsymbol{\theta}}} \\ \vdots \\ \frac{\partial \ln L}{\partial \boldsymbol{\gamma}_q}|_{\widehat{\boldsymbol{\theta}}} \\ \frac{\partial \ln L}{\partial \boldsymbol{\rho}}|_{\widehat{\boldsymbol{\theta}}} \end{bmatrix} = \widehat{\mathbf{V}}'\widehat{\boldsymbol{\Phi}}^{+}\widehat{\mathbf{V}},$$

observing that $\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} = \widehat{\sigma}^{-4}\widehat{\boldsymbol{\Phi}}$. By Proposition 1 we have $\widehat{\mathbf{V}}'\widehat{\boldsymbol{\Phi}}^{+}\widehat{\mathbf{V}} = (\mathrm{L}\widehat{\mathbf{V}})'(\mathrm{L}\widehat{\boldsymbol{\Phi}}\mathrm{L}')^{-1}(\mathrm{L}\widehat{\mathbf{V}})$, which proves the claim. □

Proof of Proposition 4. Under the maintained assumptions, $\mathrm{E}[n^{-1/2}\mathbf{X}'\mathbf{u}] = \mathbf{0}$, $\mathrm{cov}[n^{-1/2}\mathbf{X}'\mathbf{u}] = n^{-1}\mathbf{X}'\boldsymbol{\Omega}_u\mathbf{X} = O(1)$, which implies $n^{-1/2}\mathbf{X}'\mathbf{u} = O_p(1)$. Furthermore, as $(n^{-1}\mathbf{X}'\mathbf{X})^{-1} = O(1)$, we have $n^{1/2}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) = (n^{-1}\mathbf{X}'\mathbf{X})^{-1}n^{-1/2}\mathbf{X}'\mathbf{u} = O_p(1)$. Let $\mathbf{A}$ be some $n \times n$ matrix whose row and column sums are uniformly bounded in absolute value by some finite constant. Then,

$$\begin{aligned}
n^{-1}\widetilde{\mathbf{u}}'\mathbf{A}\widetilde{\mathbf{u}} &= n^{-1}\mathbf{u}'\mathbf{A}\mathbf{u} - n^{-1}\mathbf{u}'\mathbf{A}\mathbf{X}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) - (\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})'n^{-1}\mathbf{X}'\mathbf{A}\mathbf{u} + (\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta})'n^{-1}\mathbf{X}'\mathbf{A}\mathbf{X}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\
&= n^{-1}\mathbf{u}'\mathbf{A}\mathbf{u} + o_p(1)
\end{aligned}$$

since by standard argumentation $n^{-1}\mathbf{X}'\mathbf{A}\mathbf{u} = o_p(1)$ and $n^{-1}\mathbf{X}'\mathbf{A}\mathbf{X} = O(1)$. Next observe that $n^{-1}\mathrm{E}[\mathbf{u}'\mathbf{A}\mathbf{u}] = n^{-1}\mathrm{tr}[\mathbf{A}\boldsymbol{\Omega}_u]$ is bounded by a finite constant under the maintained assumptions, and

$$\mathrm{cov}[n^{-1}\mathbf{u}'\mathbf{A}\mathbf{u}] = n^{-2}2\mathrm{tr}[\mathbf{C}^2] + n^{-2}\sum_{i=1}^{n} c_{ii}^2[\mathrm{E}(\varepsilon_i/\sigma)^4 - 3]$$

with $\mathbf{C} = [c_{ij}] = \sigma^2 \mathbf{R}'^{-1}\overline{\mathbf{A}}\mathbf{R}^{-1}$. Since the elements of $\mathbf{C}$ and $\mathbf{C}^2$ are uniformly bounded in absolute value under the maintained assumptions we have $\text{cov}[n^{-1}\mathbf{u}'\mathbf{A}\mathbf{u}] = o(1)$, and hence by Chebychev's inequality $n^{-1}\widetilde{\mathbf{u}}'\mathbf{A}\widetilde{\mathbf{u}} = n^{-1}\text{E}[\mathbf{u}'\mathbf{A}\mathbf{u}] + o_p(1)$. Using this result with $\mathbf{A} = \mathbf{W}_r$ shows that $n^{-1}\widetilde{\mathbf{V}}^U = n^{-1}\boldsymbol{\mu}^U + o_p(1)$, and using this result with $\mathbf{A} = \mathbf{I}$ shows that $\widetilde{\sigma}^2 = \overline{\sigma}^2 + o_p(1)$. Hence, under the maintained assumptions, $(n^{-1}\widetilde{\boldsymbol{\Phi}}^{UU})^{-1} - (n^{-1}\boldsymbol{\Phi}^{UU})^{-1} = o_p(1)$ and $(n^{-1}\widetilde{\boldsymbol{\Phi}}^{UU})^{-1} = O_p(1)$. Consequently, $n^{-1}\mathcal{I}_u^2(q) = n^{-1}\boldsymbol{\mu}^{U\prime}(\boldsymbol{\Phi}^{UU})^{-1}\boldsymbol{\mu}^U + \nu_n$ where $\nu_n = o_p(1)$. Let $a = c_\mu^2/C_\phi$ and observe that $n^{-1}\boldsymbol{\mu}^{U\prime}(\boldsymbol{\Phi}^{UU})^{-1}\boldsymbol{\mu}^U \geq a > 0$. Since $\nu_n = o_p(1)$, there exists an n_ε such that $P(|\nu_n| \geq a/2) \leq \varepsilon$ for all $n \geq n_\varepsilon$. Consequently for all $n \geq n_\varepsilon$,

$$
\begin{aligned}
\Pr(\mathcal{I}_u^2(q) \leq \gamma) &= \Pr(\boldsymbol{\mu}^{U\prime}(\boldsymbol{\Phi}^{UU})^{-1}\boldsymbol{\mu}^U + n\nu_n \leq \gamma) \\
&\leq \Pr(na + n\nu_n \leq \gamma) \\
&= \Pr(na + n\nu_n \leq \gamma, |\nu_n| < a/2) + \Pr(na + n\nu_n \leq \gamma, |\nu_n| \geq a/2) \\
&\leq \Pr(na/2 \leq \gamma) + \Pr(|\nu_n| \geq a/2) \\
&\leq \Pr(na/2 \leq \gamma) + \varepsilon
\end{aligned}
$$

Now let n_γ be such that $n_\gamma a/2 > \gamma$, then, for all $n \geq \max(n_\varepsilon, n_\gamma)$, $\Pr(na/2 \leq \gamma) = 0$ and thus $\Pr(\mathcal{I}_u^2(q) \leq \gamma) \leq \varepsilon$, which proves that $\lim_{n\to\infty} \Pr(\mathcal{I}_u^2(q) \leq \gamma) = 0$ for any $\gamma > 0$ as claimed. □

Proof of Proposition 5. Under the alternative hypothesis, $\widehat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{d} + \mathbf{M}_X\mathbf{S}^{-1}\mathbf{u}$. Under the maintained assumptions, $n^{-1}\widehat{\mathbf{V}}^X = n^{-1}\underline{\boldsymbol{\mu}}^X + o_p(1)$ and $n^{-1}\underline{\boldsymbol{\mu}}^X = O(1)$ by argumentation as used in the proof of Proposition 4. Let $\mathbf{A}$ be some $n \times n$ matrix whose row and column sums are uniformly bounded in absolute value by some finite constant. Then,

$$
\begin{aligned}
n^{-1}\widehat{\mathbf{u}}'\mathbf{A}\widehat{\mathbf{u}} &= n^{-1}\mathbf{d}'\mathbf{A}\mathbf{d} + 2n^{-1}\mathbf{d}'\overline{\mathbf{A}}\mathbf{M}_X\mathbf{S}^{-1}\mathbf{u} + n^{-1}\mathbf{u}'\mathbf{S}'^{-1}\mathbf{M}_X\mathbf{A}\mathbf{M}_X\mathbf{S}^{-1}\mathbf{u} \\
&= n^{-1}\mathbf{d}'\mathbf{A}\mathbf{d} + 2n^{-1}\mathbf{d}'\overline{\mathbf{A}}\mathbf{M}_X\mathbf{S}^{-1}\mathbf{u} \\
&\quad + n^{-1}\mathbf{u}'\mathbf{S}'^{-1}\mathbf{A}\mathbf{S}^{-1}\mathbf{u} + (n^{-1}\mathbf{u}'\mathbf{S}'^{-1}\mathbf{X})(n^{-1}\mathbf{X}'\mathbf{X})^{-1}(n^{-1}\mathbf{X}'\mathbf{A}\mathbf{X})(n^{-1}\mathbf{X}'\mathbf{X})^{-\mathbf{1}}(n^{-1}\mathbf{X}'\mathbf{S}^{-1}\mathbf{u}) \\
&\quad + (n^{-1}\mathbf{u}'\mathbf{S}'^{-1}\mathbf{X})(n^{-1}\mathbf{X}'\mathbf{X})^{-1}(n^{-1}\mathbf{X}'\mathbf{A}\mathbf{S}^{-1}\mathbf{u}) + (n^{-1}\mathbf{u}'\mathbf{S}'^{-1}\mathbf{A}\mathbf{X})(n^{-1}\mathbf{X}'\mathbf{X})^{-\mathbf{1}}(n^{-1}\mathbf{X}'\mathbf{S}^{-1}\mathbf{u}) \\
&= n^{-1}\mathbf{d}'\mathbf{A}\mathbf{d} + n^{-1}\mathbf{u}'\mathbf{S}'^{-1}\mathbf{A}\mathbf{S}^{-1}\mathbf{u} + o_p(1).
\end{aligned}
$$

since $n^{-1}\mathbf{d}'\overline{\mathbf{A}}\mathbf{M}_X\mathbf{S}^{-1}\mathbf{u} = o_p(1)$, $n^{-1}\mathbf{X}'\mathbf{S}^{-1}\mathbf{u} = o_p(1)$, $n^{-1}\mathbf{X}'\mathbf{A}\mathbf{S}^{-1}\mathbf{u} = o_p(1)$ by argumentation as used in the proof of Proposition 4, and $(n^{-1}\mathbf{X}'\mathbf{X})^{-1} = O(1)$. Under the maintained assumptions, $n^{-1}\mathbf{d}'\mathbf{A}\mathbf{d} = O(1)$. Let $\mathbf{B} = \mathbf{S}'^{-1}\mathbf{A}\mathbf{S}^{-1}$, and observe that $n^{-1}\text{E}[\mathbf{u}'\mathbf{B}\mathbf{u}] = n^{-1}\text{tr}(\mathbf{B}\boldsymbol{\Omega}_u) = O(1)$ under the maintained assumptions, and

$$
\text{cov}[n^{-1}\mathbf{u}'\mathbf{B}\mathbf{u}] = n^{-2}2\text{tr}[\mathbf{C}^2] + n^{-2}\sum_{i=1}^{n} c_{ii}^2[\text{E}(\varepsilon_i/\sigma)^4 - 3]
$$

with $\mathbf{C} = [c_{ij}] = \sigma^2\mathbf{R}'^{-1}\overline{\mathbf{B}}\mathbf{R}^{-1}$. Since the elements of $\mathbf{C}$ and $\mathbf{C}^2$ are uniformly bounded in absolute value under the maintained assumptions we have $\mathrm{cov}[n^{-1}\mathbf{u}'\mathbf{B}\mathbf{u}] = o(1)$, and hence by Chebychev's inequality $n^{-1}\widehat{\mathbf{u}}'\mathbf{A}\widehat{\mathbf{u}} = n^{-1}\mathbf{d}'\mathbf{A}\mathbf{d} + n^{-1}\mathrm{tr}(\mathbf{B}\boldsymbol{\Omega}_u) + o_p(1)$. Using this result with $\mathbf{A} = \mathbf{W}_r$ shows that $n^{-1}\widehat{\mathbf{V}}^U = n^{-1}\underline{\boldsymbol{\mu}}^U + o_p(1)$, and using this result with $\mathbf{A} = \mathbf{I}_n$ shows that $\widehat{\sigma}^2 = \underline{\sigma}^2 + o_p(1)$. Hence, under the maintained assumptions, $n^{-1}\mathrm{L}\widehat{\mathbf{V}} = n^{-1}\underline{\boldsymbol{\mu}} + o_p(1)$, $(n^{-1}\mathrm{L}\widehat{\boldsymbol{\Phi}}\mathrm{L}')^{-1} - (n^{-1}\underline{\boldsymbol{\Phi}})^{-1} = o_p(1)$, and $(n^{-1}\mathrm{L}\widehat{\boldsymbol{\Phi}}\mathrm{L}')^{-1} = O_p(1)$. Consequently, $n^{-1}\mathcal{I}_y^2(q) = n^{-1}\underline{\boldsymbol{\mu}}'(\underline{\boldsymbol{\Phi}})^{-1}\underline{\boldsymbol{\mu}} + \nu_n$ where $\nu_n = o_p(1)$. The rest of the proof follows by the same argument as that in the proof of Proposition 4. □

Proof of Proposition 6. The proposition is proven by adapting the argumentation of Lieberman (1994). In the following let $k = p + q$ and $t = t_1 + \cdots + t_k$. It is readily checked that

$$\mathrm{E}\left[\left(\frac{\mathbf{u}'\mathbf{A}\mathbf{u} + \mathbf{a}'\mathbf{u}}{\mathbf{u}'\mathbf{S}\mathbf{u}}\right)^p \left(\frac{\mathbf{u}'\mathbf{B}\mathbf{u} + \mathbf{b}'\mathbf{u}}{\mathbf{u}'\mathbf{S}\mathbf{u}}\right)^q\right] = \int_{-\infty}^{0} \cdots \int_{-\infty}^{0} \left.\frac{\partial^k M(t_a, t_b, t_1 + \cdots + t_k)}{\partial t_a^p \partial t_b^q}\right|_{t_a, t_b = 0} dt_1 \cdots dt_k$$

Observe that

$$\left.\frac{\partial^k M(t_a, t_b, t_1 + \cdots + t_k)}{\partial t_a^p \partial t_b^q}\right|_{t_a, t_b = 0} = M^{(k)}(0, 0, t_1 + \cdots + t_k)\exp[h(0, 0, t_1 + \cdots + t_k)],$$

where

$$\begin{aligned} M^{(k)}(0, 0, t_1 + \cdots + t_k) &= \left[\left.\frac{\partial^k M(t_a, t_b, t_1 + \cdots + t_k)}{\partial t_a^p \partial t_b^q}\right|_{t_a, t_b = 0}\right] \Big/ M(0, 0, t_1 + \cdots + t_k), \\ h(0, 0, t_1 + \cdots + t_k) &= \log M(0, 0, t_1 + \cdots + t_k). \end{aligned}$$

Observe that

$$h_l(0, 0, t_1 + \cdots + t_k) = \frac{\partial h(0, 0, t_1 + \cdots + t_k)}{\partial t_l} = \frac{\mathrm{E}\left\{\mathbf{u}'\mathbf{S}\mathbf{u}\exp\left[(t_1 + \cdots + t_k)\mathbf{u}'\mathbf{S}\mathbf{u}\right]\right\}}{\mathrm{E}\left\{\exp\left[(t_1 + \cdots + t_k)\mathbf{u}'\mathbf{S}\mathbf{u}\right]\right\}} > 0$$

since $\mathbf{S}$ is positive definite. Thus, over the range of integration $(-\infty, 0]^k$, the function $h(0, 0, t_1 + \cdots + t_k)$ attains its maximum at the boundary point $t_1 = \cdots = t_k = 0$. Using the Laplace approximation of integrals (see, e.g., Olver, 1997) yields

$$\begin{aligned} & \int_{-\infty}^{0} \cdots \int_{-\infty}^{0} \left.\frac{\partial^k M(t_a, t_b, t_1 + \cdots + t_k)}{\partial t_a^p \partial t_b^q}\right|_{t_a, t_b = 0} dt_1 \cdots dt_k \\ = & \int_{-\infty}^{0} \cdots \int_{-\infty}^{0} M^{(k)}(0, 0, t_1 + \cdots + t_k)\exp[h(0, 0, t_1 + \cdots + t_k)]dt_1 \cdots dt_k \\ \simeq & \; M^{(k)}(0, 0, 0)\exp[h(0, 0, 0)]/\Pi_{l=1}^k h_l(0, 0, 0) \end{aligned}$$

with $h_l(0,0,t_1+\cdots+t_k)=\partial h(0,0,t_1+\cdots+t_k)/\partial t_l$. Next observe that

$$
\begin{aligned}
M^{(k)}(0,0,0) &= \left[\left.\frac{\partial^k M(0,0,t_1+\cdots+t_k)}{\partial t_a^p \partial t_b^q}\right|_{t_a,t_b=0}\right] \Big/ M(0,0,0) = \mathrm{E}\left[(\mathbf{u}'\mathbf{A}\mathbf{u}+\mathbf{a}'\mathbf{u})^p(\mathbf{u}'\mathbf{B}\mathbf{u}+\mathbf{b}'\mathbf{u})^q\right],\\
h(0,0,0) &= 0,\\
h_l(0,0,0) &= \left[\left.\frac{\partial M(0,0,t_1+\cdots+t_k)}{\partial t_l}\right|_{t_1=\cdots=t_k=0}\right] \Big/ M(0,0,0) = \mathrm{E}\mathbf{u}'\mathbf{S}\mathbf{u},
\end{aligned}
$$

and thus

$$
\mathrm{E}\left[\left(\frac{\mathbf{u}'\mathbf{A}\mathbf{u}+\mathbf{a}'\mathbf{u}}{\mathbf{u}'\mathbf{S}\mathbf{u}}\right)^p\left(\frac{\mathbf{u}'\mathbf{B}\mathbf{u}+\mathbf{b}'\mathbf{u}}{\mathbf{u}'\mathbf{S}\mathbf{u}}\right)^q\right] \simeq \frac{\mathrm{E}\left[(\mathbf{u}'\mathbf{A}\mathbf{u}+\mathbf{a}'\mathbf{u})^p(\mathbf{u}'\mathbf{B}\mathbf{u}+\mathbf{b}'\mathbf{u})^q\right]}{[\mathrm{E}(\mathbf{u}'\mathbf{S}\mathbf{u})]^{p+q}}.
$$

□

B.2 Derivation of Laplace Approximated Moments

Corresponding to the partitioning of $\widehat{\sigma}_{\mathrm{u}}^{-2}\widehat{\mathbf{V}}=[\widehat{\sigma}_{\mathrm{u}}^{-2}\widehat{\mathbf{V}}^{Y\prime},\widehat{\sigma}_{\mathrm{u}}^{-2}\widehat{\mathbf{V}}^{X\prime},\widehat{\sigma}_{\mathrm{u}}^{-2}\widehat{\mathbf{V}}^{U\prime}]'$ consider the following partitioning of $\boldsymbol{\mu}_L=\mathrm{E}_L[\widehat{\sigma}_{\mathrm{u}}^{-2}\widehat{\mathbf{V}}]$ and $\boldsymbol{\Phi}_L=\mathrm{E}_L[\widehat{\sigma}_{\mathrm{u}}^{-4}\widehat{\mathbf{V}}\widehat{\mathbf{V}}']$:

$$
\boldsymbol{\mu}_L=\begin{bmatrix}\boldsymbol{\mu}_L^Y\\ \boldsymbol{\mu}_L^X\\ \boldsymbol{\mu}_L^U\end{bmatrix},\qquad \boldsymbol{\Phi}_L=\begin{bmatrix}\boldsymbol{\Phi}_L^{YY} & \boldsymbol{\Phi}_L^{YX} & \boldsymbol{\Phi}_L^{YU}\\ \boldsymbol{\Phi}_L^{XY} & \boldsymbol{\Phi}_L^{XX} & \boldsymbol{\Phi}_L^{XU}\\ \boldsymbol{\Phi}_L^{UY} & \boldsymbol{\Phi}_L^{UX} & \boldsymbol{\Phi}_L^{UU}\end{bmatrix}.
$$

The elements of $\widehat{\sigma}_{\mathrm{u}}^{-2}\widehat{\mathbf{V}}^Y,\widehat{\sigma}_{\mathrm{u}}^{-2}\widehat{\mathbf{V}}^X$, and $\widehat{\sigma}_{\mathrm{u}}^{-2}\widehat{\mathbf{V}}^U$ corresponding to $\mathbf{W}_r$ are given by

$$
\widehat{\sigma}_{\mathrm{u}}^{-2}\mathbf{y}'\mathbf{W}_r'\widehat{\mathbf{u}}=\frac{\mathbf{u}'\mathbf{A}_{y,r}\mathbf{u}+\mathbf{a}_{y,r}'\mathbf{u}}{\mathbf{u}'\mathbf{S}\mathbf{u}},\quad \widehat{\sigma}_{\mathrm{u}}^{-2}\mathbf{x}_k'\mathbf{W}_r'\widehat{\mathbf{u}}=\frac{\mathbf{a}_{k,r}'\mathbf{u}}{\mathbf{u}'\mathbf{S}\mathbf{u}},\quad \widehat{\sigma}_{\mathrm{u}}^{-2}\widehat{\mathbf{u}}'\mathbf{W}_r'\widehat{\mathbf{u}}=\frac{\mathbf{u}'\mathbf{A}_{u,r}\mathbf{u}}{\mathbf{u}'\mathbf{S}\mathbf{u}},
$$

with $\mathbf{A}_{y,r}=(\mathbf{M}_X\mathbf{W}_r+\mathbf{W}_r'\mathbf{M}_X)/2$, $\mathbf{a}_{y,r}=\mathbf{M}_X\mathbf{W}_r\mathbf{X}\boldsymbol{\beta}$, $\mathbf{a}_{k,r}=\mathbf{M}_X\mathbf{W}_r\mathbf{x}_k$, $\mathbf{A}_{u,r}=\mathbf{M}_X\overline{W}_r\mathbf{M}_X$, and $\mathbf{S}=(n-K_x)^{-1}\mathbf{M}_X$. By Lemma A.1 in Kelejian and Prucha (2010) we have for any conformably symmetric matrices $\mathbf{A}$ and $\mathbf{B}$ and vectors $\mathbf{a}$ and $\mathbf{b}$:

$$
\begin{aligned}
&\mathrm{E}(\mathbf{u}'\mathbf{A}\mathbf{u}+\mathbf{a}'\mathbf{u})=\sigma^2\mathrm{tr}(\mathbf{A})\\
&\mathrm{E}\left[(\mathbf{u}'\mathbf{A}\mathbf{u}+\mathbf{a}'\mathbf{u})(\mathbf{u}'\mathbf{B}\mathbf{u}+\mathbf{b}'\mathbf{u})\right]=2\sigma^4\mathrm{tr}(\mathbf{A}\mathbf{B})+\sigma^2\mathbf{a}'\mathbf{b}\\
&+\left[\mu^{(4)}-3\sigma^4\right]\mathrm{vec}_D(\mathbf{A})'\mathrm{vec}_D(\mathbf{B})+\mu^{(3)}\left[\mathbf{a}'\mathrm{vec}_D(\mathbf{B})+\mathbf{b}'\mathrm{vec}_D(\mathbf{A})\right]+\sigma^4\mathrm{tr}(\mathbf{A})\mathrm{tr}(\mathbf{B})
\end{aligned}
$$

Using Proposition 6 and observing furthermore that $\mathrm{E}(\mathbf{u}'\mathbf{S}\mathbf{u}) = \sigma^2$ it is then readily seen that

$$\boldsymbol{\mu}_L^Y = \left[\mathrm{tr}(\overline{\mathbf{W}}_r \mathbf{M}_X)\right]_{r=1,\cdots,q}, \quad \boldsymbol{\mu}_L^X = \mathbf{0}, \quad \boldsymbol{\mu}_L^U = \left[\mathrm{tr}(\overline{\mathbf{W}}_r \mathbf{M}_X)\right]_{r=1,\cdots,q},$$

and

$$\begin{aligned}
\boldsymbol{\Phi}_L^{YY} &= \big[\mathrm{tr}(\mathbf{W}_r \mathbf{M}_X \mathbf{W}_s \mathbf{M}_X + \mathbf{W}_r' \mathbf{M}_X \mathbf{W}_s \mathbf{M}_X) + \sigma^{-2} \boldsymbol{\beta}' \mathbf{X}' \mathbf{W}_r' \mathbf{M}_X \mathbf{W}_s \mathbf{X} \boldsymbol{\beta} \\
&\quad + (\sigma^{-4}\mu^{(4)} - 3)\mathrm{vec}_D(\mathbf{W}_r' \mathbf{M}_X)' \mathrm{vec}_D(\mathbf{W}_s' \mathbf{M}_X) + \sigma^{-4}\mu^{(3)}[\boldsymbol{\beta}' \mathbf{X}' \mathbf{W}_r' \mathbf{M}_X \mathrm{vec}_D(\mathbf{W}_s' \mathbf{M}_X) \\
&\quad + \boldsymbol{\beta}' \mathbf{X}' \mathbf{W}_s' \mathbf{M}_X \mathrm{vec}_D(\mathbf{W}_r' \mathbf{M}_X)] + \mathrm{tr}(\overline{\mathbf{W}}_r \mathbf{M}_X)\mathrm{tr}(\overline{\mathbf{W}}_s \mathbf{M}_X)\big]_{r,s=1,\cdots,q}, \\
\boldsymbol{\Phi}_L^{YX} &= \left[\sigma^{-2} \boldsymbol{\beta}' \mathbf{X}' \mathbf{W}_r' \mathbf{M}_X \mathbf{W}_s \mathbf{X} + \sigma^{-4}\mu^{(3)} \mathrm{vec}_D(\mathbf{W}_r' \mathbf{M}_X) \mathbf{M}_X \mathbf{W}_s \mathbf{X}\right]_{r,s=1,\cdots,q}, \\
\boldsymbol{\Phi}_L^{YU} &= \Big[2\mathrm{tr}(\overline{\mathbf{W}}_r \mathbf{M}_X \overline{\mathbf{W}}_s \mathbf{M}_X) + (\sigma^{-4}\mu^{(4)} - 3)\mathrm{vec}_D(\mathbf{W}_r' \mathbf{M}_X)\mathrm{vec}_D(\mathbf{M}_X \overline{\mathbf{W}}_s \mathbf{M}_X) \\
&\quad + \sigma^{-4}\mu^{(3)} \boldsymbol{\beta}' \mathbf{X}' \mathbf{W}_r' \mathbf{M}_X \mathrm{vec}_D(\mathbf{M}_X \overline{\mathbf{W}}_s \mathbf{M}_X) + \mathrm{tr}(\overline{\mathbf{W}}_r \mathbf{M}_X)\mathrm{tr}(\overline{\mathbf{W}}_s \mathbf{M}_X)\Big]_{r,s=1,\cdots,q},
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{\Phi}_L^{XX} &= \left[\sigma^{-2} \mathbf{X}' \mathbf{W}_r' \mathbf{M}_X \mathbf{W}_s \mathbf{X}\right]_{r,s=1,\cdots,q}, \\
\boldsymbol{\Phi}_L^{XU} &= \left[\sigma^{-4}\mu^{(3)} \mathbf{X}' \mathbf{W}_r' \mathbf{M}_X \mathrm{vec}_D(\mathbf{M}_X \overline{\mathbf{W}}_s \mathbf{M}_X)\right]_{r,s=1,\cdots,q}, \\
\boldsymbol{\Phi}_L^{UU} &= [2\mathrm{tr}(\overline{\mathbf{W}}_r \mathbf{M}_X \overline{\mathbf{W}}_s \mathbf{M}_X) + (\sigma^{-4}\mu^{(4)} - 3)\mathrm{vec}_D(\mathbf{M}_X \overline{\mathbf{W}}_r \mathbf{M}_X)' \mathrm{vec}_D(\mathbf{M}_X \overline{\mathbf{W}}_s \mathbf{M}_X) \\
&\quad + \mathrm{tr}(\overline{\mathbf{W}}_r \mathbf{M}_X)\mathrm{tr}(\overline{\mathbf{W}}_s \mathbf{M}_X)]_{r,s=1,\cdots,q}.
\end{aligned}$$

Note that $\boldsymbol{\mu}_L$ does not depend on any unknown parameters, and hence $\widehat{\boldsymbol{\mu}}_L = \boldsymbol{\mu}_L$. The estimator for $\widehat{\boldsymbol{\Phi}}_L$ is obtained by replacing $\boldsymbol{\beta}$ by the OLS estimator and σ^2 by $\widehat{\sigma}^2$ or $\widehat{\sigma}_{\mathrm{u}}^2$, and $\mu^{(3)}$ and $\mu^{(4)}$ by $n^{-1}\sum_{i=1}^n \widehat{u}_i^3$ and $n^{-1}\sum_{i=1}^n \widehat{u}_i^3$, respectively.

Remark B.1. Cliff and Ord (1981) gives results on the exact mean and variance (and higher moments) of ratios of quadratic forms under normality. Drukker and Prucha (2013) also give results on the covariance of ratios of quadratic forms under normality. We now can use those results to check on the approximation error for the Laplace approximation when the disturbances are normally distributed. In particular, for spatial weight matrices $\mathbf{W}_r$ and $\mathbf{W}_s$, let

$$Q_r = \widehat{\sigma}_{\mathrm{u}}^{-2} \widehat{\mathbf{u}}' \mathbf{W}_r \widehat{\mathbf{u}} = \frac{\mathbf{u}' \mathbf{A}_r \mathbf{u}}{\mathbf{u}' \mathbf{S} \mathbf{u}}, \qquad \text{and} \qquad Q_s = \widehat{\sigma}_{\mathrm{u}}^{-2} \widehat{\mathbf{u}}' \mathbf{W}_s \widehat{\mathbf{u}} = \frac{\mathbf{u}' \mathbf{A}_s \mathbf{u}}{\mathbf{u}' \mathbf{S} \mathbf{u}},$$

where $\mathbf{A}_r = \mathbf{M}_X \overline{\mathbf{W}}_r \mathbf{M}_X$, $\mathbf{A}_s = \mathbf{M}_X \overline{\mathbf{W}}_s \mathbf{M}_X$ and $\mathbf{S} = (n - K_x)^{-1} \mathbf{M}_X$. Provided $\mathbf{u} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ it follows from Drukker and Prucha (2013) that

$$\mathrm{E}\left[Q_r^p Q_s^q\right] = \frac{\mathrm{E}\left[(\mathbf{u}' \mathbf{A}_r \mathbf{u})^p (\mathbf{u}' \mathbf{A}_s \mathbf{u})^q\right]}{\mathrm{E}\left[(\mathbf{u}' \mathbf{S} \mathbf{u})^{p+q}\right]},$$

for $p, q = 0, 1$. Let E_L denote the Laplace approximation of the expected value, then in light of Proposition 6,

$$\begin{aligned}
\frac{\mathrm{E}_L\left[Q_r\right]}{\mathrm{E}\left[Q_r\right]} &= \frac{\mathrm{E}(\mathbf{u}'\mathbf{S}\mathbf{u})}{\mathrm{E}(\mathbf{u}'\mathbf{S}\mathbf{u})} = 1, \\
\frac{\mathrm{E}_L\left[Q_r Q_s\right]}{\mathrm{E}\left[Q_r Q_s\right]} &= \frac{\mathrm{E}\left[(\mathbf{u}'\mathbf{S}\mathbf{u})^2\right]}{\left[\mathrm{E}(\mathbf{u}'\mathbf{S}\mathbf{u})\right]^2} = 1 + 2(n - K_x)^{-1} = 1 + O(n^{-1})
\end{aligned}$$

observing that $\mathrm{E}(\mathbf{u}'\mathbf{S}\mathbf{u}) = \sigma^2$ and $\mathrm{E}\left[(\mathbf{u}'\mathbf{S}\mathbf{u})^2\right] = 2\sigma^4 \mathrm{tr}(\mathbf{S}^2) + \sigma^4 \mathrm{tr}(\mathbf{S}) = [2(n - K_x)^{-1} + 1]\sigma^4$.

Table 1. Rejection Rates for $I_u^2(q)$ Tests ($\mathrm{n}=500, \mathrm{m}=10$, exogenous z, and homoskedastic errors)

						Standardized Tests			
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	Bonferroni Test	$I_u^2(1)$ Test with W_1	$I_u^2(1)$ Test with W_2	$I_u^2(2)$ Test with W_1, W_2	Bonferroni Test
		Normal Errors							
0	0	0.0453	0.0490	0.0457	0.0392	0.0465	0.0477	0.0476	0.0420
.2	0	0.5615	0.2631	0.5001	0.5021	0.5761	0.2821	0.5171	0.5204
.4	0	0.9902	0.7956	0.9857	0.9869	0.9913	0.8100	0.9874	0.9877
0	.2	0.1770	0.3543	0.2952	0.3075	0.1902	0.3754	0.3132	0.3258
0	.4	0.6065	0.9013	0.8587	0.8695	0.6228	0.9108	0.8697	0.8803
.2	.2	0.8834	0.8498	0.8854	0.9001	0.8923	0.8621	0.8961	0.9089
.4	.4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		Log-normal Errors							
0	0	0.0381	0.0355	0.0404	0.0368	0.0392	0.0374	0.0427	0.0390
.2	0	0.5362	0.2311	0.4542	0.4652	0.5542	0.2507	0.4699	0.4866
.4	0	0.9959	0.7996	0.9900	0.9922	0.9965	0.8135	0.9912	0.9931
0	.2	0.1582	0.3230	0.2624	0.2747	0.1705	0.3428	0.2775	0.2920
0	.4	0.5829	0.9091	0.8494	0.8658	0.6024	0.9184	0.8607	0.8785
.2	.2	0.9060	0.8551	0.8894	0.9073	0.9141	0.8685	0.9003	0.9164
.4	.4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Nominal size is 0.05

Table 2a. Rejection Rates for $I_y^2(q)$ Tests ($n = 500, m = 10$, exogenous z, and homoskedastic normal errors)

										Standardized Tests			
λ_1	λ_2	ρ_1	ρ_2	γ_1	γ_2	$I_y^2(1)$ Test with W_1	$I_y^2(1)$ Test with W_2	$I_y^2(2)$ Test with W_1, W_2	Bonferroni Test	$I_y^2(1)$ Test with W_1	$I_y^2(1)$ Test with W_2	$I_y^2(2)$ Test with W_1, W_2	Bonferroni Test
0	0	0	0	0	0	0.0465	0.0460	0.0502	0.0391	0.0478	0.0459	0.0494	0.0406
.1	0	0	0	0	0	0.4171	0.2321	0.3409	0.3561	0.4263	0.2381	0.3469	0.3648
.2	0	0	0	0	0	0.9688	0.7690	0.9391	0.9523	0.9697	0.7759	0.9417	0.9540
0	.1	0	0	0	0	0.1760	0.2852	0.2271	0.2436	0.1791	0.2931	0.2336	0.2502
0	.2	0	0	0	0	0.6036	0.8656	0.7851	0.8206	0.6087	0.8705	0.7912	0.8261
.1	.1	0	0	0	0	0.8454	0.8195	0.8171	0.8576	0.8498	0.8244	0.8208	0.8618
.2	.2	0	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0	0	.2	0	0	0	0.4874	0.2206	0.4281	0.4274	0.5047	0.2343	0.4419	0.4428
0	0	.4	0	0	0	0.9831	0.7365	0.9730	0.9761	0.9846	0.7518	0.9756	0.9787
0	0	0	.2	0	0	0.1504	0.2987	0.2470	0.2553	0.1580	0.3138	0.2616	0.2693
0	0	0	.4	0	0	0.5291	0.8597	0.8024	0.8190	0.5448	0.8715	0.8133	0.8317
0	0	.2	.2	0	0	0.8382	0.7944	0.8395	0.8608	0.8482	0.8069	0.8511	0.8725
0	0	.4	.4	0	0	1.0000	0.9998	1.0000	1.0000	1.0000	0.9998	1.0000	1.0000
.1	.1	.2	.2	0	0	0.9989	0.9969	0.9991	0.9994	0.9990	0.9973	0.9991	0.9994
.2	.2	.4	.4	0	0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0	0	0	0	.2	0	0.8149	0.5631	0.7128	0.7568	0.8131	0.5639	0.7105	0.7561
0	0	0	0	.5	0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0	0	0	0	0	.2	0.4265	0.6628	0.5418	0.5847	0.4249	0.6630	0.5396	0.5842
0	0	0	0	0	.5	0.9976	1.0000	1.0000	1.0000	0.9976	1.0000	1.0000	1.0000
0	0	0	0	.2	.2	0.9983	0.9962	0.9968	0.9988	0.9983	0.9962	0.9966	0.9988
0	0	0	0	.5	.5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Nominal size is 0.05

Table 2b. Rejection Rates for $I_y^2(q)$ Tests ($\mathrm{n} = 500, \mathrm{m} = 10$, exogenous z, and homoskedastic log-normal errors)

										Standardized Tests			
λ_1	λ_2	ρ_1	ρ_2	γ_1	γ_2	$I_y^2(1)$ Test with W_1	$I_y^2(1)$ Test with W_2	$I_y^2(2)$ Test with W_1, W_2	Bonferroni Test	$I_y^2(1)$ Test with W_1	$I_y^2(1)$ Test with W_2	$I_y^2(2)$ Test with W_1, W_2	Bonferroni Test
0	0	0	0	0	0	0.0456	0.0442	0.0469	0.0420	0.0473	0.0451	0.0474	0.0433
.1	0	0	0	0	0	0.3985	0.2076	0.3112	0.3354	0.4087	0.2129	0.3196	0.3453
.2	0	0	0	0	0	0.9737	0.7738	0.9406	0.9582	0.9755	0.7813	0.9436	0.9611
0	.1	0	0	0	0	0.1577	0.2623	0.2037	0.2193	0.1615	0.2717	0.2087	0.2288
0	.2	0	0	0	0	0.6003	0.8696	0.7761	0.8205	0.6073	0.8747	0.7835	0.8264
.1	.1	0	0	0	0	0.8494	0.8273	0.8085	0.8607	0.8547	0.8330	0.8184	0.8680
.2	.2	0	0	0	0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0	0	.2	0	0	0	0.4615	0.2007	0.3914	0.4013	0.4765	0.2113	0.4058	0.4169
0	0	.4	0	0	0	0.9907	0.7343	0.9798	0.9848	0.9917	0.7515	0.9824	0.9860
0	0	0	.2	0	0	0.1374	0.2744	0.2219	0.2345	0.1454	0.2903	0.2363	0.2495
0	0	0	.4	0	0	0.5068	0.8600	0.7807	0.8127	0.5245	0.8712	0.7956	0.8258
0	0	.2	.2	0	0	0.8498	0.7977	0.8379	0.8607	0.8608	0.8117	0.8487	0.8698
0	0	.4	.4	0	0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
.1	.1	.2	.2	0	0	0.9995	0.9989	0.9995	0.9998	0.9995	0.9990	0.9997	0.9999
.2	.2	.4	.4	0	0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0	0	0	0	.2	0	0.8285	0.5727	0.7248	0.7700	0.8290	0.5707	0.7233	0.7690
0	0	0	0	.5	0	1.0000	0.9990	0.9997	0.9999	1.0000	0.9990	0.9997	0.9999
0	0	0	0	0	.2	0.4204	0.6739	0.5404	0.5976	0.4199	0.6738	0.5383	0.5970
0	0	0	0	0	.5	0.9949	0.9994	0.9988	0.9993	0.9950	0.9994	0.9988	0.9993
0	0	0	0	.2	.2	0.9958	0.9959	0.9936	0.9961	0.9956	0.9958	0.9936	0.9961
0	0	0	0	.5	.5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Nominal size is 0.05

Table 3. Rejection Rates for $I_u^2(q)$ Tests ($\mathrm{n} = 500, \mathrm{m} = 10$, exogenous z, and homoskedastic normal errors)

			Standardized Tests	
ρ_1	$I_u^2(q)$ Test with $W_1, \dots, W_q$	Bonferroni Test	$I_u^2(q)$ Test with $W_1, \dots, W_q$	Bonferroni Test
	$q = 5$			
0	0.0479	0.0400	0.0498	0.0414
.2	0.3891	0.4312	0.4041	0.4515
.4	0.9627	0.9765	0.9669	0.9799
.6	1.0000	1.0000	1.0000	1.0000
.8	1.0000	1.0000	1.0000	1.0000
	$q = 10$			
0	0.0514	0.0374	0.0524	0.0390
.2	0.3056	0.3938	0.3180	0.4138
.4	0.9290	0.9673	0.9322	0.9713
.6	1.0000	1.0000	1.0000	1.0000
.8	1.0000	1.0000	1.0000	1.0000

Nominal size is 0.05

Table 4. Rejection Rates for $I_y^2(q)$ Tests ($\mathrm{n} = 500, \mathrm{m} = 10$, exogenous z, and homoskedastic normal errors)

					Standardized Tests	
λ_1	ρ_1	γ_1	$I_y^2(q)$ Test with $W_1, \dots, W_q$	Bonferroni Test	$I_y^2(q)$ Test with $W_1, \dots, W_q$	Bonferroni Test
			$q = 5$			
0	0	0	0.0497	0.0366	0.0499	0.0377
.1	0	0	0.2625	0.3225	0.2681	0.3305
.2	0	0	0.8856	0.9475	0.8880	0.9499
0	.2	0	0.3218	0.3570	0.3339	0.3767
0	.4	0	0.9342	0.9602	0.9382	0.9644
.1	.2	0	0.7501	0.8169	0.7607	0.8293
.2	.4	0	1.0000	1.0000	1.0000	1.0000
0	0	.2	0.5861	0.7269	0.5836	0.7261
0	0	.5	1.0000	1.0000	1.0000	1.0000
			$q = 10$			
0	0	0	0.0481	0.0333	0.0487	0.0349
.1	0	0	0.1915	0.2835	0.1947	0.2931
.2	0	0	0.7817	0.9189	0.7851	0.9222
0	.2	0	0.2518	0.3301	0.2616	0.3483
0	.4	0	0.8869	0.9476	0.8928	0.9543
.1	.2	0	0.6455	0.7842	0.6554	0.7954
.2	.4	0	0.9994	1.0000	0.9995	1.0000
0	0	.2	0.4088	0.6545	0.4054	0.6533
0	0	.5	1.0000	1.0000	1.0000	1.0000

Nominal size is 0.05

Table 5. Rejection Rates for $I_u^2(q)$ Tests ($\mathrm{n} = 500, \mathrm{m} = 10$, endogenous z, and heteroskedastic normal errors)

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	Bonferroni Test
		$\lambda_1 = \lambda_2 = 0$			
0	0	0.0485	0.0475	0.0478	0.0404
.2	0	0.4699	0.2427	0.4077	0.4095
.4	0	0.9728	0.7690	0.9579	0.9619
0	.2	0.1541	0.3502	0.2897	0.2875
0	.4	0.5009	0.8973	0.8533	0.8538
.2	.2	0.8092	0.8348	0.8329	0.8539
.4	.4	1.0000	1.0000	1.0000	1.0000
		$\lambda_1 = \lambda_2 = .2$			
0	0	0.0433	0.0433	0.0433	0.0389
.2	0	0.3551	0.1903	0.2928	0.3067
.4	0	0.9169	0.6319	0.8754	0.8926
0	.2	0.1145	0.2689	0.2138	0.2145
0	.4	0.3926	0.7946	0.7229	0.7312
.2	.2	0.6826	0.7134	0.7030	0.7387
.4	.4	0.9998	0.9999	0.9998	0.9998

Nominal size is 0.05

Table 6. Rejection Rates for $I_y^2(q)$ Tests ($\mathrm{n} = 500, \mathrm{m} = 10$, endogenous z, and heteroskedastic normal errors)

λ_1	λ_2	ρ_1	ρ_2	γ_1	γ_2	$I_y^2(1)$ Test with W_1	$I_y^2(1)$ Test with W_2	$I_y^2(2)$ Test with W_1, W_2	Bonferroni Test
0	0	0	0	0	0	0.0478	0.0501	0.0491	0.0433
.1	0	0	0	0	0	0.7039	0.4319	0.6043	0.6353
.2	0	0	0	0	0	0.9999	0.9761	0.9990	0.9994
0	.1	0	0	0	0	0.3041	0.5374	0.4300	0.4636
0	.2	0	0	0	0	0.9038	0.9922	0.9846	0.9887
.1	.1	0	0	0	0	0.9922	0.9869	0.9887	0.9924
.2	.2	0	0	0	0	1.0000	1.0000	1.0000	1.0000
0	0	.2	0	0	0	0.4042	0.2084	0.3545	0.3547
0	0	.4	0	0	0	0.9572	0.7075	0.9368	0.9436
0	0	0	.2	0	0	0.1314	0.2949	0.2398	0.2392
0	0	0	.4	0	0	0.4358	0.8579	0.8001	0.8056
0	0	.2	.2	0	0	0.7506	0.7836	0.7793	0.8076
0	0	.4	.4	0	0	1.0000	1.0000	1.0000	1.0000
.1	.1	.2	.2	0	0	0.9998	0.9999	1.0000	1.0000
.2	.2	.4	.4	0	0	1.0000	1.0000	1.0000	1.0000
0	0	0	0	.2	0	0.9951	0.9106	0.9843	0.9904
0	0	0	0	.5	0	1.0000	1.0000	1.0000	1.0000
0	0	0	0	0	.2	0.7937	0.9570	0.9187	0.9363
0	0	0	0	0	.5	1.0000	1.0000	1.0000	1.0000
0	0	0	0	.2	.2	1.0000	1.0000	1.0000	1.0000
0	0	0	0	.5	.5	1.0000	1.0000	1.0000	1.0000

Nominal size is 0.05

Table 7. Rejection Rates for $I_u^2(q)$ Tests ($\text{n} = 500, \text{m} = 10$, exogenous z, and homoskedastic errors)

ρ	$I_u^2(1)$ Test with W_0	$I_u^2(1)$ Test with W_1	$I_u^2(1)$ Test with W_2	$I_u^2(2)$ Test with W_1, W_2
	Exogenous W_0			
0	0.0535	0.0521	0.0512	0.0502
.2	0.3394	0.2775	0.2523	0.2683
.4	0.8744	0.7744	0.7324	0.7753
.6	0.9958	0.9833	0.9752	0.9852
.8	1.0000	0.9998	0.9996	0.9997
	Endogenous W_0			
0	0.0657	0.0453	0.0490	0.0457
.2	0.4136	0.2229	0.2286	0.2286
.4	0.9277	0.7055	0.7156	0.7260
.6	0.9993	0.9772	0.9768	0.9828
.8	1.0000	0.9998	0.9997	0.9997

Nominal size is 0.05

Table 8. Rejection Rates for $I_y^2(q)$ Tests ($\text{n} = 500, \text{m} = 10$, exogenous z, and homoskedastic errors)

λ	ρ	γ	$I_y^2(1)$ Test with W_0	$I_y^2(1)$ Test with W_1	$I_y^2(1)$ Test with W_2	$I_y^2(2)$ Test with W_1, W_2
			Exogenous W_0			
0	0	0	0.0505	0.0494	0.0506	0.0527
.1	0	0	0.2852	0.2297	0.2127	0.2016
.2	0	0	0.8371	0.7182	0.6882	0.6827
0	.2	0	0.2794	0.2293	0.2114	0.2295
0	.4	0	0.8271	0.7182	0.6654	0.7160
.1	.2	0	0.6732	0.5519	0.5072	0.5353
.2	.4	0	0.9978	0.9867	0.9797	0.9879
0	0	.2	0.6538	0.5187	0.4999	0.4520
0	0	.5	1.0000	0.9996	0.9986	0.9987
			Endogenous W_0			
0	0	0	1.0000	0.0465	0.0460	0.0502
.1	0	0	1.0000	0.2506	0.2395	0.2279
.2	0	0	1.0000	0.7830	0.7547	0.7518
0	.2	0	1.0000	0.1956	0.1991	0.1999
0	.4	0	1.0000	0.6625	0.6691	0.6784
.1	.2	0	1.0000	0.6083	0.5927	0.5974
.2	.4	0	1.0000	0.9963	0.9951	0.9961
0	0	.2	1.0000	0.4422	0.4228	0.3781
0	0	.5	1.0000	0.9933	0.9901	0.9899

Nominal size is 0.05