

Simultaneous Equations Models with Higher-Order Spatial or Social Network Interactions

David M. Drukker¹, Peter H. Egger² and Ingmar R. Prucha³

December 27, 2021 (This Version)

¹Department of Economics and International Business, Sam Houston State University, Huntsville, TX 77351, Tel.: +1-936-294-4791, e-mail: dmetrics@gmail.com

²KOF, ETH Zurich, Leonhardstrasse 21, 8092 Zurich, Switzerland; Tel.: +41-44-632-4108, e-mail: egger@kof.ethz.ch

³Department of Economics, University of Maryland, College Park, MD 20742, Tel.: +1-301-405-3499, e-mail: prucha@umd.edu

Abstract

This paper develops an estimation methodology for network data generated from a system of simultaneous equations, which allows for network interdependencies via spatial lags in the endogenous and exogenous variables, as well as in the disturbances. By allowing for higher-order spatial lags our specification provides important flexibility in modeling network interactions. The estimation methodology builds, among others, on the two-step generalized method of moments (GMM) estimation approach introduced in Kelejian and Prucha (1998, 1999, 2004). The paper considers limited and full information estimators, and one- and two-step estimators, and establishes their asymptotic properties. In contrast to some of the earlier two-step estimation literature our asymptotic results facilitate joint tests for the absence of all forms of network spillovers.

Key Words: Cliff-Ord spatial model; Limited information estimation; Full information estimation; Two-stage least squares estimation; Three-stage least squares estimation; Generalized method of moments estimation

JEL Classification: C21, C31

1 Introduction¹

In this paper we develop a generalized estimation theory for simultaneous equation systems for cross sectional data with possible network interactions in the dependent variables, the exogenous variables and the disturbances. A leading application will be spatial networks. However, since network interdependencies are modeled only to relate to a measure of proximity, without assuming that observations are indexed by location, the developed methodology can be of interest to the estimation of a much wider class of networks, including social networks.

There is substantial empirical evidence of cross sectional interdependence among observations in many areas of economics both at the macro level, where cross sectional units may, e.g., be countries, states or counties, as well as at the micro level where cross sectional units may, e.g., be industries, firms or individuals.²

An important class of models for spatial networks originates from the seminal work by Whittle (1954) and Cliff and Ord (1973, 1981). In those models cross-sectional interactions are modeled through spatial lags, where the weights used in forming the spatial lags are reflective of the relative importance of the links between neighbors for the generation of spillovers. In a spatial setting the relative importance would typically be taken to be inversely related to a measure of distance. The usefulness of those models for the analysis of a wide class of networks beyond spatial networks stems from the recognition that the notion of distance is not confined to geographic distance.³

¹We are grateful to the editor, a guest editor, and three anonymous referees for their very helpful comments. We gratefully acknowledge financial support from the National Institute of Health through SBIR grants R43 AG027622 and R44 AG027622. We also thank the CESifo for their hospitality.

²For instance, Conley and Ligon (2002) or Ertur and Koch (2007) document spatial spillovers in economic growth. Holtz-Eakin (1994) or Audretsch and Feldmann (1996) put forward evidence for spatial spillovers in productivity. Egger, Pfaffermayr, and Winner (2005) provide evidence for spatial interdependencies of value added tax rates, and Devereux, Lockwood, and Redoano (2008) for spatial corporate tax rates. Case, Hines, and Rosen (1993) report on spatial budget spillovers. The results in Behrens, Ertur, and Koch (2012) suggest that bilateral trade flows exhibit spatial interdependence, and Baltagi, Egger, and Pfaffermayr (2007, 2008) and Blonigen, Davies, Waddell, and Naughton (2007) illustrate that the same holds true for bilateral foreign direct investment.

Pinkse, Slade, and Brett (2002) provide evidence for spatial price competition among wholesale-gasoline terminals. For contributions to the literature on social interactions see, e.g., Ballester, Calvó-Armengol and Zenou (2006), Lee (2007b), Calvó-Armengol, Patacchini and Zenou (2009), Blume, Bock, Durlauf and Ioannides (2011), Cohen-Cole, Liu and Zenou (2018), and Liu (2014).

³As stated, we think Cliff-Ord type models provide important tools for analyzing networks. However, we also want to point out an important literature in statistics where spatial depen-

Spatial econometrics has a long history in geography, regional science and urban economics; see, e.g., Anselin (1988). For the last two decades the development of econometric methods of inference for Cliff-Ord type models has also been an active area of research in economics.⁴ Most of the literature focused on single-equation models where a single dependent variable, say, y_i , is determined for units $i = 1, \dots, n$. However, in economics it is frequent that the outcomes for several dependent variables, say, y_{i1}, \dots, y_{iG} , are determined jointly by a system of equations for units $i = 1, \dots, n$. In this case the simultaneous nature of the outcomes can stem from two sources, interactions between different economic variables as well as interactions between cross sectional units.

Surprisingly, the literature on the estimation of simultaneous systems of spatially interrelated cross sectional equations has been quite limited until recently. Kelejian and Prucha (2004) provide, by extending the methodology developed in Kelejian and Prucha (1998, 1999) for single equations, an early development of generalized method of moments (GMM) estimators for such models. However, as discussed in more detail below, they do not provide a full asymptotic theory for all considered estimators and their setup only covers first-order spatial lags. Liu (2014, 2019, 2020), Cohen-Cole, Liu, and Zenou (2018), and Liu and Saraiva (2019) build and extend the methodology of Kelejian and Prucha (2004) within the context of social interaction models with first order spatial lags, and cross sectionally independent disturbances. Their contributions include one-step GMM estimation methods that utilize both linear and quadratic moment conditions, identification conditions, bias correction procedures for many instruments, heteroskedasticity, and an estimation methodology for a simultaneous system of equations with binary outcomes generated from an incomplete information network game. Other recent contributions to the literature on spatial simultaneous equation models are Baltagi and Deng (2015), who consider an extension of a two-equation system with first-order spatial lags to panels. Wang, Li, and Wang (2014) analyze the quasi maximum likelihood estimator for such a system in the cross section. Yang and Lee (2017) consider identification and quasi maximum likelihood estimation for a multi-equation system with a first-order spatial lag in the dependent variable. Yang and Lee (2019) provide an extension to dynamic panel data models allowing for multiple weights matrices. In contrast to the current paper, the above cited literature only considers first-order spatial

dence is molded via random Markov fields and conditional autoregressive models; see e.g., Cressie (1993).

⁴See, e.g., Anselin (2010) for a review of the development of spatial econometric methods.

lags in the dependent variable, and does not also consider spatial spillovers in the disturbance process. Those papers also differ in terms of the considered estimation methodology. In particular, a methodological focus of this paper is on two-step estimation, while also covering one-step estimation.

Within the context of Cliff-Ord type models two-step procedures are generally less efficient than one-step procedures. However, they are attractive, especially for situations where the instruments are strong and the efficiency loss is small, because of their relative computational simplicity. An important limitation of the two-step estimation methodology developed in Kelejian and Prucha (2004) is that the paper only establishes the consistency of the estimator for the spatial autoregressive parameter in the disturbance process, but not its asymptotic normality. As a result, the methodology does not facilitate a joint test for the absence of spatial interactions in the dependent variables, the exogenous variables and the disturbances. Closely related to this is that Kelejian and Prucha (2004) do not consider efficiently weighted GMM estimators for the spatial autoregressive parameters, since that paper lacked the knowledge of the limiting distribution for those estimators.

Another important limitation of the earlier paper is that it only allowed for first-order spatial lags. Allowing for higher-order spatial lags is important for at least two reasons. First, the researcher may not be sure about the channel through which interactions occur – e.g., through geographic proximity, or technological proximity, or both. By allowing for higher-order spatial lags the researcher can consult the data on this issue. Second, as argued below, higher-order spatial lags can be used to partially relax the requirement regarding a priori knowledge of what weights should be assigned to different units in the construction of a spatial lag.⁵ Given the complexity of our systems specification, we do not pursue nonparametric estimation here.

The paper is organized as follows: Section 2 specifies the considered simultaneous equation system with spatial/cross sectional network interactions. In this section we also discuss two exemplary applications. The first example highlights how higher-order spatial lags can be used to achieve a more flexible specification of the spatial weights. The second example considers a social interaction model where individuals make interdependent choices on the level of effort for multi-

⁵The term higher-order spatial lags seems to have been imported from the time series literature. However, in contrast to the time series literature there is not a natural ordering of the lags. Different spatial lags model different pathways for network dependence, and the corresponding spatial autoregressive parameters reflect their importance.

ple activities. In Section 3, we discuss the moment conditions underlying the considered GMM estimators for the regression parameters and the parameters of the disturbance process. The paper focuses on two-step estimation procedures. It turns out that the distribution of the GMM estimator for the spatial autoregressive parameters of the disturbance process depends on the estimator of regression parameters. Section 4 is hence devoted to give generic results concerning the consistency and asymptotic normality of two-step GMM estimators. In particular, we give generic results concerning the joint limiting distribution of estimators for all model parameters of interest, which can be utilized in the usual way to form general Wald tests regarding the model parameters. Results for one-step estimators are in essence delivered as a special case of two-step estimation. In Section 5 we consider specific limited and full information two-step estimators, and provide specific expressions for consistent estimators of the associated asymptotic variance-covariance (VC) matrices of those estimators. In Section 6 we consider limited and full information one-step estimators that combine the linear and quadratic moment conditions used by the two-step estimators. The last section concludes with a summary of our findings and possible directions for future research. All technical derivations are given in appendices and in an Online Supplementary Appendix. In the Online Supplementary Appendix we also report on a Monte Carlo study of the small sample properties of the various estimators and test statistics.

Throughout the paper we adopt the following notations and conventions. Let $(\mathbf{A}_n)_{n \in N}$ be some sequence of matrices, then we denote the (i, j) -th element of \mathbf{A}_n with $a_{ij,n}$. If \mathbf{A}_n is nonsingular, then we denote its inverse with \mathbf{A}_n^{-1} , and the (i, j) -th element of \mathbf{A}_n^{-1} with a_n^{ij} . Let \mathbf{A}_n be of dimension $p_n \times p_n$, then the maximum column sum and row sum matrix norms of \mathbf{A}_n are, respectively, defined as

$$\|\mathbf{A}_n\|_1 = \max_{1 \leq j \leq p_n} \sum_{i=1}^{p_n} |a_{ij,n}| \quad \text{and} \quad \|\mathbf{A}_n\|_\infty = \max_{1 \leq i \leq p_n} \sum_{j=1}^{p_n} |a_{ij,n}|.$$

If $\|\mathbf{A}_n\|_1 \leq c$ and $\|\mathbf{A}_n\|_\infty \leq c$ for some finite constant c which does not depend on n , then we say that the row and column sums of the sequence of matrices \mathbf{A}_n are uniformly bounded in absolute value. We note that if the row and column sums of the matrices \mathbf{A}_n and \mathbf{B}_n are uniformly bounded in absolute value, then so are the row and column sums of $\mathbf{A}_n + \mathbf{B}_n$ and $\mathbf{A}_n \mathbf{B}_n$; cf., e.g., Kapoor, Kelejian and Prucha (2007, Remark A2). For any square matrix \mathbf{A}_n , $\bar{\mathbf{A}}_n = (\mathbf{A}_n + \mathbf{A}_n')/2$, and for any vector or matrix \mathbf{A}_n , $\|\mathbf{A}_n\| = [\text{tr}(\mathbf{A}_n' \mathbf{A}_n)]^{1/2}$,

where tr denotes the trace operator. Let $\mathbf{A}_{n,g}$, $g = 1, \dots, G$, be a sequence of matrices, then $diag_{g=1}^G(\mathbf{A}_{n,g})$ denotes the block diagonal matrix, where $\mathbf{A}_{n,g}$ is the g -th diagonal block.

2 Model

In the following, we specify our simultaneous system of G equations for G endogenous variables observed for n cross sectional units.

2.1 Structural Form Model

In specifying the system we allow for two sources of simultaneity. First, the observations for the g -th endogenous variable for the i -th unit may depend on observations of the other endogenous variables for the i -th unit, as in the classical text book simultaneous equation system.⁶ Second, simultaneity may stem from Cliff and Ord (1973, 1981) type higher-order cross-sectional network interactions, where spatial interactions represent a leading application.

As remarked, the model specification will be fairly general and allows for network interactions modeled by, possibly, higher-order spatial lags in the dependent variables, the exogenous variables and the disturbances. More specifically, let g denote the equation index, then we assume that the cross-sectional data are generated by the following system ($g = 1, \dots, G$):

$$\begin{aligned} \mathbf{y}_{g,n} &= \sum_{l=1}^G b_{lg,n} \mathbf{y}_{l,n} + \sum_{k=1}^K c_{kg,n} \mathbf{x}_{k,n} + \sum_{l=1}^G \left[\sum_{s=1}^p \lambda_{lg,s,n} \mathbf{W}_{s,n} \right] \mathbf{y}_{l,n} + \mathbf{u}_{g,n}, \\ \mathbf{u}_{g,n} &= \left[\sum_{r=1}^q \rho_{g,r,n} \mathbf{M}_{r,n} \right] \mathbf{u}_{g,n} + \boldsymbol{\varepsilon}_{g,n}, \end{aligned} \quad (1)$$

where $\mathbf{y}_{g,n}$ is the $n \times 1$ vector of cross-sectional observations on the dependent variable in the g -th equation, $\mathbf{x}_{k,n}$ is the $n \times 1$ vector of cross-sectional observations on the k -th exogenous variable, which is taken to be nonstochastic,⁷ $\mathbf{u}_{g,n}$ is the $n \times 1$ disturbance vector in the g -th equation, $\mathbf{W}_{s,n}$ and $\mathbf{M}_{r,n}$ are $n \times n$ weights matrices, $\boldsymbol{\varepsilon}_{g,n}$ is the $n \times 1$ vector of innovations entering the disturbance

⁶Our specification differs from, e.g., Anselin (1988) and Wang, Lee and Bao (2018) who consider systems for one variable.

⁷In treating the exogenous variables as non-stochastic, the analysis may be viewed as conditional on the exogenous variables.

process for the g -th equation, and n denotes the sample size. With $b_{lg,n}$ and $c_{kg,n}$ we denote the (scalar) parameters corresponding to the l -th endogenous and k -th exogenous variables, respectively. Of course, the structural model parameters are not identified without certain restrictions. Those restrictions will be introduced below.

Consistent with the usual terminology for Cliff-Ord type network interactions, we refer to $\mathbf{W}_{s,n}$ and $\mathbf{M}_{r,n}$ as spatial weights matrices, to

$$\bar{\mathbf{y}}_{l,s,n} = \mathbf{W}_{s,n} \mathbf{y}_{l,n} \quad \text{and} \quad \bar{\mathbf{u}}_{g,r,n} = \mathbf{M}_{r,n} \mathbf{u}_{g,n}$$

as spatial lags, and to the (real scalar) parameters $\lambda_{lg,s,n}$ and $\rho_{g,r,n}$ as spatial autoregressive parameters. The weights matrices carry the information on the links between units and on the relative weight of those links, and the spatial autoregressive parameters describe the strength of the spillovers. Although originally introduced for spatial networks, Cliff-Ord type interaction models do not require the indexing of observations by location. In general they only rely on a measure of distance in the formation of the spatial weights. Since the notions of space and distance or proximity are not confined to geographic space, these models have, as discussed in the introduction, also been applied in various other settings. This includes social interaction models, where one considered specification has been to assign to each of the i -th individual's friends a weight of $1/n_i$, where n_i denotes the total number of friends of i , while assigning zero weights to individuals not belonging to the circle of friends. In the following, we continue to refer to $\bar{\mathbf{y}}_{l,s,n}$ and $\bar{\mathbf{u}}_{g,r,n}$ and $\lambda_{lg,s,n}$ and $\rho_{g,r,n}$ as spatial lags and spatial autoregressive parameters, but note the wider applicability.

The reason for allowing the elements of the spatial weights matrices to depend on the sample size is to permit – as is frequent practice in applications – normalizations of these matrices where the normalization factor(s) depend on the sample size.⁸ The i -th element of $\bar{\mathbf{y}}_{l,s,n}$ is given by $\bar{y}_{il,s,n} = \sum_{j=1}^n w_{ij,s,n} y_{jl,n}$. We note that even if the elements of the spatial weights matrices do not depend on the sample size, the elements of the spatial lag $\bar{\mathbf{y}}_{l,s,n}$ and, analogously, the elements of $\bar{\mathbf{u}}_{g,r,n}$ will generally depend on the sample size. This in turn implies that also the elements of $\mathbf{y}_{g,n}$ and $\mathbf{u}_{g,n}$ will generally depend on the sample size, i.e., form triangular arrays. In allowing the elements of $\mathbf{x}_{k,n}$ to depend on the

⁸The normalizing factors may in turn affect the parameters of the spatial lags, which is the reason for allowing the parameters in (1) to depend on the sample size; see, e.g., Kelejian and Prucha (2010) for further discussions regarding normalizations.

sample size, we implicitly also allow for some of the exogenous variables to be spatial lags of exogenous variables. For example, the elements of $\mathbf{x}_{k,n}$ could be of the form $x_{ik,n} = \sum_{j=1}^n w_{ij,s,n} \xi_j$ where ξ_j is some basic exogenous variable. Thus the model accommodates, as remarked above, cross-sectional interactions in the endogenous variables, the exogenous variables, and the disturbances.

The above model generalizes the spatial simultaneous equation model considered in Kelejian and Prucha (2004) in allowing for higher-order spatial lags.⁹ Consistent with the terminology introduced by Anselin and Florax (1995) in a single-equation context, we refer to the above model as a simultaneous spatial autoregressive model of order p with spatially autoregressive disturbances of order q , for short, a simultaneous SARAR(p,q) model.¹⁰ One reason for allowing for multiple spatial weights matrices is that they can capture different forms of proximity between units. For example, within the context of R&D spillovers between firms, one matrix may refer to geographic proximity between firms, and the other may correspond to a measure of proximity in the product space. As another example, as discussed in more detail below, within the context of a social interaction model different matrices may refer to different circles of friends, e.g., one matrix may identify the very close friends, and a second matrix the other friends. Additionally, as discussed below, an estimation theory that allows for multiple spatial weights matrices can also be used to accommodate certain parameterizations of the spatial weights.

Model (1) can be written more compactly as

$$\begin{aligned} \mathbf{Y}_n &= \mathbf{Y}_n \mathbf{B}_n + \mathbf{X}_n \mathbf{C}_n + \bar{\mathbf{Y}}_n \mathbf{\Lambda}_n + \mathbf{U}_n, \\ \mathbf{U}_n &= \bar{\mathbf{U}}_n \mathbf{R}_n + \mathbf{E}_n \end{aligned} \quad (2)$$

with

$$\begin{aligned} \mathbf{Y}_n &= (\mathbf{y}_{1,n}, \dots, \mathbf{y}_{G,n})_{n \times G}, & \mathbf{X}_n &= (\mathbf{x}_{1,n}, \dots, \mathbf{x}_{K,n})_{n \times K}, \\ \mathbf{U}_n &= (\mathbf{u}_{1,n}, \dots, \mathbf{u}_{G,n})_{n \times G}, & \mathbf{E}_n &= (\boldsymbol{\varepsilon}_{1,n}, \dots, \boldsymbol{\varepsilon}_{G,n})_{n \times G}, \\ \bar{\mathbf{Y}}_n &= (\bar{\mathbf{y}}_{1,1,n}, \dots, \bar{\mathbf{y}}_{1,p,n}, \dots, \bar{\mathbf{y}}_{G,1,n}, \dots, \bar{\mathbf{y}}_{G,p,n})_{n \times pG}, \\ \bar{\mathbf{U}}_n &= (\bar{\mathbf{u}}_{1,1,n}, \dots, \bar{\mathbf{u}}_{1,q,n}, \dots, \bar{\mathbf{u}}_{G,1,n}, \dots, \bar{\mathbf{u}}_{G,q,n})_{n \times qG}, \end{aligned}$$

and where the parameter matrices $\mathbf{B}_n = (b_{lg,n})_{G \times G}$, $\mathbf{C}_n = (c_{lg,n})_{K \times G}$, $\mathbf{\Lambda}_n =$

⁹Extensions of the estimation methodology will be discussed later.

¹⁰For single equations higher order SAR models have been considered by Blommestein (1983, 1985) and Huang (1984), among others, and more recently by Bell and Bockstael (2000), Cohen and Morrison Paul (2007), Badinger and Egger (2010), and Lee and Liu (2010).

$(\lambda_{lg,s,n})_{pG \times G}$, and $\mathbf{R}_n = (\rho_{g,r,n})_{qG \times G}$ are defined conformably.¹¹

2.2 Exemplary Applications

We next motivate the importance of considering higher-order spatial lags with two examples. The first example illustrates how higher-order spatial lags can be useful for certain parameterizations of the spatial weights. The second example formulates a social interaction model where the utility maximizing solution is described by a system of equations with higher-order spillovers as defined in (1).

2.2.1 Parameterized Spatial Weights

As part of the specification of the model in (1) the researcher has to specify the elements of the spatial weights matrices. When those elements are specified incorrectly, the model is misspecified and the estimates will generally be inconsistent. Allowing for higher-order spatial lags provides important flexibility and robustness in modeling network interactions. However, while adding higher order spatial lags helps to address potential specification errors, it does not, of course, eliminate the possibility; for a recent contribution on an omnibus test for weights matrix misspecification see Lee, Phillips and Rossi (2021).

In the following we discuss exemplarily how higher-order spatial lags can be used to allow for certain flexible parameterizations of the spatial weights. For simplicity of notation we drop subscripts n . Spatial weights are often specified as a function of some distance measure, possibly combined with some contiguity measure. Let $\mathbf{W} = (w_{ij})$ be the basic spatial weights matrix, let d_{ij} denote some distance measure between units i and j , and let d_{ij}^* be some contiguity measure taking on values of one or zero. Then, the researcher may specify the weights as the product of the contiguity measure and a polynomial in d_{ij} , treating the coefficients of the polynomial as unknown parameters:¹²

$$w_{ij} = d_{ij}^* [\lambda_1 d_{ij} + \dots + \lambda_p d_{ij}^p].$$

Now, suppose the researcher models \mathbf{y}_g as a function of, say, $\lambda_{lg} \mathbf{W} \mathbf{y}_l$, then clearly

$$\lambda_{lg} \mathbf{W} \mathbf{y}_l = \left[\lambda_{lg} \sum_{s=1}^p \lambda_s \mathbf{W}_s \right] \mathbf{y}_l = \left[\sum_{s=1}^p \lambda_{lg,s} \mathbf{W}_s \right] \mathbf{y}_l$$

¹¹For clarity, we note that the g -th column of $\mathbf{\Lambda}_n$ and \mathbf{R}_n are, respectively, given by $[\lambda_{1g,1,n}, \dots, \lambda_{1g,p,n}, \dots, \lambda_{Gg,1,n}, \dots, \lambda_{Gg,p,n}]'$ and $[0, \dots, 0, \rho_{g,1,n}, \dots, \rho_{g,q,n}, 0, \dots, 0]'$.

¹²Alternatively the researcher could specify a polynomial in $1/d_{ij}$.

with $\lambda_{lg,s} = \lambda_{lg}\lambda_s$, $\mathbf{W}_s = (w_{ij,s})$, and $w_{ij,s} = d_{ij}^* d_{ij}^s$. In allowing for higher-order spatial lags, model (1) covers this specification as a special case. Of course, the above specification of spatial weights is entirely illustrative, and model (1) will cover many other specifications, including specifications with alternate basis functions instead of power functions, and more general measures of distance and contiguity. The same ideas also apply to the modeling of the disturbance process.¹³

2.2.2 Social Interactions in Multiple Activities

The following example extends Cohen-Cole, Liu and Zenou (2018). We follow their basic setup, but allow for more flexible peer effects. More specifically, consider a model where n individuals choose effort levels for G activities, say y_{i1}, \dots, y_{iG} , allowing for peer effects among p groups of peers, e.g., for $p = 2$ we may distinguish between very close friends and other friends. Now let $w_{ij,s}^*$ be one or zero depending on whether individual j belongs to the s -th peer group of individual i , and let n_{is} be the size of that peer group. Let $w_{ij,s} = w_{ij,s}^*/n_{is}$ denote the corresponding normalized weights, let $\bar{y}_{ig,s} = \sum_{j=1}^n w_{ij,s} y_{jg}$ denote the average effort level for activity g by the s -th group of peers, and assume that the utility of individual i is of the following linear-quadratic form:

$$\begin{aligned} u(y_{i1}, \dots, y_{iG}) &= v(y_{i1}, \dots, y_{iG}, \bar{y}_{i1,1}, \dots, \bar{y}_{i1,p}, \dots, \bar{y}_{iG,1}, \dots, \bar{y}_{iG,p}) \quad (3) \\ &= \underbrace{\sum_{g=1}^G \pi_{ig}^* y_{ig} + \sum_{g=1}^G y_{ig} \sum_{l=1}^G \sum_{s=1}^p \lambda_{lg,s}^* \bar{y}_{il,s}}_{\text{payoff}} - \underbrace{\frac{1}{2} \sum_{g=1}^G b_g^* y_{ig}^2 - \sum_{g=1}^G \sum_{l=1, l \neq g}^G b_{lg}^* y_{ig} y_{il}}_{\text{cost}}. \end{aligned}$$

The specification considered in Cohen-Cole, Liu and Zenou (2018) corresponds to $p = 1$. The first-order conditions for the maximum of $u(y_{i1}, \dots, y_{iG})$ yield

$$y_{ig} = \pi_{ig} + \sum_{l=1, l \neq g}^G b_{lg} y_{il} + \sum_{l=1}^G \sum_{s=1}^p \lambda_{lg,s} \bar{y}_{il,s}$$

¹³The above observations are related to Pinkse and Slade (1998), who estimate, in a single-equation context, the spatial weights corresponding to the dependent variable nonparametrically. Given the complexity of our systems specification, we do not pursue nonparametric estimation here.

with $\pi_{ig} = \pi_{ig}^*/b_g^*$, $b_{lg} = -b_{lg}^*/b_g^*$, $\lambda_{lg,s} = \lambda_{lg,s}^*/b_g^*$, or in matrix notation

$$\mathbf{y}_g = \boldsymbol{\pi}_g + \sum_{l=1, l \neq g}^G b_{lg} \mathbf{y}_l + \sum_{l=1}^G \sum_{s=1}^p \lambda_{lg,s} \mathbf{W}_s \mathbf{y}_l \quad (4)$$

with $\boldsymbol{\pi}_g = [\pi_{1g}, \dots, \pi_{ng}]'$. Similar to Cohen-Cole, Liu and Zenou (2018) assume that $\boldsymbol{\pi}_g$ can be modeled as

$$\boldsymbol{\pi}_g = \sum_{k=1}^K c_{kg} \mathbf{x}_k + \mathbf{u}_g. \quad (5)$$

Substituting (5) into (4) then shows that the utility maximizing vectors of effort for the G activities are defined as the solution of a model of the form specified in (1).

We note that by specifying the \mathbf{W}_s to be block diagonal, we can accommodate situations where the individuals $i = 1, \dots, n$ belong to, say, C groups which, e.g., represent class rooms. Also, some of the \mathbf{x}_k covariates may represent group indicator variables, and others may be spatial lags of some basic covariates.

2.3 Reduced Form and Structural Model with Exclusion Restrictions

Towards computing the reduced form of the above model, let

$$\mathbf{y}_n = \text{vec}(\mathbf{Y}_n), \quad \mathbf{x}_n = \text{vec}(\mathbf{X}_n), \quad \mathbf{u}_n = \text{vec}(\mathbf{U}_n), \quad \boldsymbol{\varepsilon}_n = \text{vec}(\mathbf{E}_n),$$

and let

$$\mathbf{W}_n = [\mathbf{W}'_{1,n}, \dots, \mathbf{W}'_{p,n}]', \quad \mathbf{M}_n = [\mathbf{M}'_{1,n}, \dots, \mathbf{M}'_{q,n}]'$$

Observing that $\text{vec}(\overline{\mathbf{Y}}_n) = (\mathbf{I}_G \otimes \mathbf{W}_n) \mathbf{y}_n$ and $\text{vec}(\overline{\mathbf{U}}_n) = (\mathbf{I}_G \otimes \mathbf{M}_n) \mathbf{u}_n$, and that $\text{vec}(\mathbf{A}_1 \mathbf{A}_2) = (\mathbf{A}_2' \otimes \mathbf{I}) \text{vec}(\mathbf{A}_1)$ for any two conformable matrices \mathbf{A}_1 and \mathbf{A}_2 , it is readily seen that the spatial simultaneous equation system (2) can be re-written in stacked notation as

$$\begin{aligned} \mathbf{y}_n &= \mathbf{B}_n^* \mathbf{y}_n + \mathbf{C}_n^* \mathbf{x}_n + \mathbf{u}_n, \\ \mathbf{u}_n &= \mathbf{R}_n^* \mathbf{u}_n + \boldsymbol{\varepsilon}_n, \end{aligned} \quad (6)$$

where $\mathbf{B}_n^* = [(\mathbf{B}'_n \otimes \mathbf{I}_n) + (\mathbf{\Lambda}'_n \otimes \mathbf{I}_n)(\mathbf{I}_G \otimes \mathbf{W}_n)]$, $\mathbf{C}_n^* = (\mathbf{C}'_n \otimes \mathbf{I}_n)$, and $\mathbf{R}_n^* = (\mathbf{R}'_n \otimes \mathbf{I}_n)(\mathbf{I}_G \otimes \mathbf{M}_n)$. Assuming invertability of $\mathbf{I}_{nG} - \mathbf{B}_n^*$ and $\mathbf{I}_{nG} - \mathbf{R}_n^*$, the reduced form of the system is now given by

$$\begin{aligned} \mathbf{y}_n &= (\mathbf{I}_{nG} - \mathbf{B}_n^*)^{-1} [\mathbf{C}_n^* \mathbf{x}_n + \mathbf{u}_n], \\ \mathbf{u}_n &= (\mathbf{I}_{nG} - \mathbf{R}_n^*)^{-1} \boldsymbol{\varepsilon}_n. \end{aligned} \quad (7)$$

As remarked, the structural parameters of the spatial simultaneous equation system (1) and (2) are not identified unless we impose exclusion restrictions. Let $\boldsymbol{\beta}_{g,n}$, $\boldsymbol{\gamma}_{g,n}$, $\boldsymbol{\lambda}_{g,n}$, and $\boldsymbol{\rho}_{g,n}$ denote the $G_g \times 1$, $K_g \times 1$, $p_g \times 1$ and $q_g \times 1$ vectors of non-zero elements of the g -th column of \mathbf{B}_n , \mathbf{C}_n , $\mathbf{\Lambda}_n$, and \mathbf{R}_n , respectively, and let $\mathbf{Y}_{g,n}$, $\mathbf{X}_{g,n}$, $\bar{\mathbf{Y}}_{g,n}$, and $\bar{\mathbf{U}}_{g,n}$ be the corresponding matrices of observations on the endogenous variables, exogenous variables, spatially lagged endogenous variables, and spatially lagged disturbances appearing in the structural equation for the g -th endogenous variable. Then, system (2) can be expressed as ($g = 1, \dots, G$):

$$\begin{aligned} \mathbf{y}_{g,n} &= \mathbf{Z}_{g,n} \boldsymbol{\delta}_{g,n} + \mathbf{u}_{g,n}, \\ \mathbf{u}_{g,n} &= \bar{\mathbf{U}}_{g,n} \boldsymbol{\rho}_{g,n} + \boldsymbol{\varepsilon}_{g,n}, \end{aligned} \quad (8)$$

where $\mathbf{Z}_{g,n} = [\mathbf{Y}_{g,n}, \mathbf{X}_{g,n}, \bar{\mathbf{Y}}_{g,n}]$ and $\boldsymbol{\delta}_{g,n} = [\boldsymbol{\beta}'_{g,n}, \boldsymbol{\gamma}'_{g,n}, \boldsymbol{\lambda}'_{g,n}]'$.

2.4 Model Assumptions

We maintain the following assumptions regarding the spatial weights matrices and model parameters.

Assumption 1 For $s = 1, \dots, p$ and $r = 1, \dots, q$: (a) All diagonal elements of $\mathbf{W}_{s,n}$ and $\mathbf{M}_{r,n}$ are zero. (b) $\|\mathbf{W}_{s,n}\|_1 \leq c$, $\|\mathbf{M}_{r,n}\|_1 \leq c$ for some finite constant c which does not depend on n , and $\|\mathbf{W}_{s,n}\|_\infty = 1$, $\|\mathbf{M}_{r,n}\|_\infty = 1$.

Assumption 2 (a) The matrices $\mathbf{I}_{nG} - \mathbf{B}_n^*$ are nonsingular. (b) The spatial autoregressive parameters satisfy $\sup_n \sum_{r \in \mathbf{I}_{g,p}} |\rho_{g,r,n}| < 1$ for $g = 1, \dots, G$, where $\mathbf{I}_{g,p} = \{r_{g,1}, \dots, r_{g,q_g}\} \subseteq \{1, \dots, q\}$ denotes the set of indices associated with the elements of $\boldsymbol{\rho}_{g,n}$. (c) The row and column sums of the matrices $[\mathbf{I}_{nG} - \mathbf{B}_n^*]^{-1}$ are uniformly bounded in absolute value.

The above assumptions are in line with the recent spatial literature. Assumption 1(a) entails a normalization rule. Assumption 1(b) implies that the row and

column sums of the matrices $\mathbf{W}_{s,n}$ and $\mathbf{M}_{r,n}$ are uniformly bounded in absolute value. The assumption that $\|\mathbf{W}_{s,n}\|_\infty = 1$ and $\|\mathbf{M}_{r,n}\|_\infty = 1$ implies a normalization for the parameters. For interpretation, let $\mathbf{W}_{s,n}^*$ be some spatial weights matrix with $\|\mathbf{W}_{s,n}^*\|_\infty \neq 1$ and let $\lambda_{lg,s,n}^*$ be the corresponding spatial autoregressive parameter on $\mathbf{W}_{s,n}^* \mathbf{y}_{l,n}$. Now define $\mathbf{W}_{s,n} = \mathbf{W}_{s,n}^* / \|\mathbf{W}_{s,n}^*\|_\infty$ and $\lambda_{lg,s,n} = \lambda_{lg,s,n}^* \|\mathbf{W}_{s,n}^*\|_\infty$, then $\|\mathbf{W}_{s,n}\|_\infty = 1$ and $\lambda_{lg,s,n} \mathbf{W}_{s,n} = \lambda_{lg,s,n}^* \mathbf{W}_{s,n}^*$. Thus the normalizations $\|\mathbf{W}_{s,n}\|_\infty = 1$ and $\|\mathbf{M}_{r,n}\|_\infty = 1$ can always be achieved by appropriately re-scaling the elements of the spatial weights matrix, provided the corresponding spatial autoregressive parameter is correspondingly redefined; for further discussions see Kelejian and Prucha (2010).¹⁴

Assumption 2(a) ensures that the first equation of the expression for the reduced form in (7) is well defined. In allowing in Assumption 2(b) for the index set $\mathbf{I}_{g,\rho}$ to vary with g we allow for different orders of spatial lags in the disturbance process of different equations. Next observe that $\mathbf{R}_n^* = \text{diag}_{g=1}^G [\mathbf{R}_{g,n}^*]$ with $\mathbf{R}_{g,n}^* = \mathbf{R}_{g,n}^*(\boldsymbol{\rho}_{g,n}) = \sum_{r \in \mathbf{I}_{g,\rho}} \rho_{g,r,n} \mathbf{M}_{r,n}$. In light of this it follows from Assumptions 1(b) and 2(b) that $\|\mathbf{R}_n^*\|_\infty \leq \max_g \sum_{r \in \mathbf{I}_{g,\rho}} |\rho_{g,r,n}| < 1$, which in turn implies that $\mathbf{I}_{nG} - \mathbf{R}_n^*$ is nonsingular. Consequently, also the second equation of the expression for the reduced form in (7) is well defined, and thus \mathbf{y}_n is uniquely defined by the model. Assumptions 1(b) and 2(b) imply even that $\sup_n \|\mathbf{R}_{g,n}^*\|_\infty < 1$, which implies that the row sums of the matrices $[\mathbf{I}_n - \mathbf{R}_{g,n}^*]^{-1}$ are uniformly bounded in absolute value. To see this observe that $\|[\mathbf{I}_n - \mathbf{R}_{g,n}^*]^{-1}\|_\infty \leq 1 / [1 - \|\mathbf{R}_{g,n}^*\|_\infty] \leq 1 / [1 - \sup_n \|\mathbf{R}_{g,n}^*\|_\infty] < \infty$. The above arguments use results in Horn and Johnson (1985, p. 301).

Assumption 3 (a) *The matrix of (nonstochastic) exogenous regressors \mathbf{X}_n in (2) has full column rank (for n sufficiently large). Furthermore, the elements of \mathbf{X}_n are uniformly bounded in absolute value by some finite constant.* (b) *The elements of the parameter matrices \mathbf{B}_n , \mathbf{C}_n and $\boldsymbol{\Lambda}_n$ are uniformly bounded in absolute value.*

An assumption such as Assumption 3(a) is common in the spatial literature. In treating \mathbf{X}_n as nonstochastic, our analysis should be viewed as conditional on

¹⁴The suggested re-scaling is simple and practically implementable even if n is large. In situations where n is sufficiently small such that the eigenvalues of the spatial weights matrices are computable, one could alternatively normalize each spatial weights matrix by its spectral radius, which would in conjunction with the next assumptions entail an expansion of the admissible parameter space.

\mathbf{X}_n . Note that in part (b) of Assumption 3 uniformity refers to n . Assumption 3(b) is trivially satisfied, if the parameters do not depend on the sample size, since any real number is finite.

We next state the assumptions maintained w.r.t. $\boldsymbol{\varepsilon}_n$. In the following let $\mathbf{V}_n = [\mathbf{v}_{1,n}, \dots, \mathbf{v}_{G,n}]$ be an $n \times G$ matrix of basic innovations and let $\mathbf{v}_n = \text{vec}(\mathbf{V}_n)$.

Assumption 4 *The innovations $\boldsymbol{\varepsilon}_n$ are generated as follows:*

$$\boldsymbol{\varepsilon}_n = (\boldsymbol{\Sigma}'_{\star} \otimes \mathbf{I}_n) \mathbf{v}_n, \quad (9)$$

where $\boldsymbol{\Sigma}'_{\star}$ is a nonsingular $G \times G$ matrix and the random variables $\{v_{ig,n} : i = 1, \dots, n, g = 1, \dots, G\}$ are, for each n , identically and independently distributed with zero mean, unitary variance, and finite $4 + \nu$ moments for some $\nu > 0$, and their distribution does not depend on n . Furthermore, let $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}'_{\star} \boldsymbol{\Sigma}_{\star}$.

The above assumption on the innovation process is in line with the specification of the disturbance terms for a classical simultaneous equation system. Let $\boldsymbol{\varepsilon}_n(i)$ denote the i -th row of \mathbf{E}_n , then, observing that $\mathbf{E}_n = \mathbf{V}_n \boldsymbol{\Sigma}_{\star}$, it is readily seen that the innovation vectors $\{\boldsymbol{\varepsilon}_n(i) : 1 \leq i \leq n\}$ are i.i.d. with zero mean and VC matrix $\boldsymbol{\Sigma}$. With respect to the stacked innovation vector, the assumption implies that $E\boldsymbol{\varepsilon}_n = 0$ and $E\boldsymbol{\varepsilon}_n \boldsymbol{\varepsilon}_n' = \boldsymbol{\Sigma} \otimes \mathbf{I}_n$.

Given (7), we note that Assumption 4 implies furthermore that $E\mathbf{u}_n = 0$ and $E\mathbf{y}_n = (\mathbf{I}_{nG} - \mathbf{B}_n^*)^{-1} \mathbf{C}_n^* \mathbf{x}_n$, and that the VC matrices of \mathbf{u}_n and \mathbf{y}_n are given by, respectively,

$$\begin{aligned} \boldsymbol{\Omega}_{u,n} &= (\mathbf{I}_{nG} - \mathbf{R}_n^*)^{-1} (\boldsymbol{\Sigma} \otimes \mathbf{I}_n) (\mathbf{I}_{nG} - \mathbf{R}_n^{*'})^{-1}, \\ \boldsymbol{\Omega}_{y,n} &= (\mathbf{I}_{nG} - \mathbf{B}_n^*)^{-1} \boldsymbol{\Omega}_{u,n} (\mathbf{I}_{nG} - \mathbf{B}_n^{*'})^{-1}. \end{aligned}$$

Assumptions 2 and 4 imply that the row and column sums of the VC matrix of \mathbf{u}_n (and similarly those of \mathbf{y}_n) are uniformly bounded in absolute value, thus limiting the degree of correlation between, respectively, the elements of \mathbf{u}_n and of \mathbf{y}_n .

Remark: Under the above assumptions it is shown in Lemmata A.2 and A.3 that all random variables in $[\mathbf{Y}_n, \mathbf{X}_n, \bar{\mathbf{Y}}_n]$ have uniformly bounded finite fourth

moments, and that

$$n^{-1}\mathbf{Z}'_{g,n}\mathbf{A}_n\mathbf{u}_{g,n} - n^{-1}E\mathbf{Z}'_{g,n}\mathbf{A}_n\mathbf{u}_{g,n} = o_p(1)$$

for any $n \times n$ real matrix \mathbf{A}_n whose row and column sums are bounded uniformly in absolute value.

For purposes of estimation it proves helpful to apply a spatial Cochrane-Orcutt transformation to the model. In particular, premultiplying (8) by $\mathbf{I}_n - \mathbf{R}_{g,n}^*(\boldsymbol{\rho}_{g,n})$ yields

$$\mathbf{y}_{*g,n} = \mathbf{Z}_{*g,n}\boldsymbol{\delta}_{g,n} + \varepsilon_{g,n}, \quad (10)$$

with $\mathbf{y}_{*g,n} = \mathbf{y}_{*g,n}(\boldsymbol{\rho}_{g,n}) = [\mathbf{I}_n - \mathbf{R}_{g,n}^*(\boldsymbol{\rho}_{g,n})]\mathbf{y}_{g,n}$ and $\mathbf{Z}_{*g,n} = \mathbf{Z}_{*g,n}(\boldsymbol{\rho}_{g,n}) = [\mathbf{I}_n - \mathbf{R}_{g,n}^*(\boldsymbol{\rho}_{g,n})]\mathbf{Z}_{g,n}$. Stacking the transformed equations yields

$$\mathbf{y}_{*n} = \mathbf{Z}_{*n}\boldsymbol{\delta}_n + \varepsilon_n, \quad (11)$$

with $\mathbf{y}_{*n} = [\mathbf{y}'_{*1,n}, \dots, \mathbf{y}'_{*G,n}]'$, $\mathbf{Z}_{*n} = \text{diag}_{g=1}^G[\mathbf{Z}_{*g,n}]$, $\boldsymbol{\delta}_n = [\boldsymbol{\delta}'_{1,n}, \dots, \boldsymbol{\delta}'_{G,n}]'$ and where ε_n is as defined above.

3 Moment Conditions

Recall that from the Cochrane-Orcutt transformed form of the model (10) we have

$$\varepsilon_{g,n} = \varepsilon_{g,n}(\boldsymbol{\rho}_{g,n}, \boldsymbol{\delta}_{g,n}) = [\mathbf{I}_n - \mathbf{R}_{g,n}^*(\boldsymbol{\rho}_{g,n})][\mathbf{y}_{g,n} - \mathbf{Z}_{g,n}\boldsymbol{\delta}_{g,n}].$$

The estimators for the model parameters $\boldsymbol{\rho}_{g,n}$ and $\boldsymbol{\delta}_{g,n}$ considered in this paper will utilize a set of linear and quadratic moment conditions of the form ($g = 1, \dots, G$)

$$E\mathbf{m}_{g,n}^\delta(\boldsymbol{\rho}_{g,n}, \boldsymbol{\delta}_{g,n}) = En^{-1}\mathbf{H}'_n\varepsilon_{g,n} = \mathbf{0}, \quad (12)$$

$$E\mathbf{m}_{g,n}^\rho(\boldsymbol{\rho}_{g,n}, \boldsymbol{\delta}_{g,n}) = E \begin{bmatrix} n^{-1}\boldsymbol{\varepsilon}'_{g,n}\mathbf{A}_{1,n}\varepsilon_{g,n} \\ \vdots \\ n^{-1}\boldsymbol{\varepsilon}'_{g,n}\mathbf{A}_{S,n}\varepsilon_{g,n} \end{bmatrix} = \mathbf{0}, \quad (13)$$

where the $n \times p_H$ instrument matrix \mathbf{H}_n in the linear form and the $n \times n$ weighting matrices $\mathbf{A}_{s,n}$ in the quadratic forms are non-stochastic. In the following we

will also simply write $\mathbf{m}_{g,n}^\delta$ and $\mathbf{m}_{g,n}^\rho$ for the sample moment vector at the true parameter values.¹⁵

We maintain the following assumptions regarding the instruments \mathbf{H}_n . Specific choices of instruments and a discussion of how the nonlinearity of $E\mathbf{y}_n$ in the parameters generates instruments from within the model are given after the assumptions.

Assumption 5 : *The instrument matrices \mathbf{H}_n are nonstochastic and have full column rank $p_H \geq G_g + K_g + p_g$ (for all n large enough). Furthermore, the elements of the matrices \mathbf{H}_n are uniformly bounded in absolute value. Additionally \mathbf{H}_n is assumed to contain, at least, the linearly independent columns of $\underline{\mathbf{H}}_n = [\mathbf{X}_n, \mathbf{M}_{1,n}\mathbf{X}_n, \dots, \mathbf{M}_{p,n}\mathbf{X}_n]$.*

The inclusion of $\underline{\mathbf{H}}_n$ in \mathbf{H}_n ensures that the exogenous variables on the right hand side (r.h.s.) of the Cochrane-Orcutt transformed model serve as their own best instruments. For limited information estimators it suffices to postulate that \mathbf{H}_n is assumed to contain, at least, the linearly independent columns of $[\mathbf{X}_{g,n}, \mathbf{M}_{1,n}\mathbf{X}_{g,n}, \dots, \mathbf{M}_{p,n}\mathbf{X}_{g,n}]$.

Assumption 6 *The instruments \mathbf{H}_n satisfy furthermore:*

- (a) $\mathbf{Q}_{HH} = \lim_{n \rightarrow \infty} n^{-1} \mathbf{H}'_n \mathbf{H}_n$ is finite and nonsingular.
- (b) $\mathbf{Q}_{HZ,g} = \text{plim}_{n \rightarrow \infty} n^{-1} \mathbf{H}'_n \mathbf{Z}_{g,n}$ and $\mathbf{Q}_{HMZ,r,g} = \text{plim}_{n \rightarrow \infty} n^{-1} \mathbf{H}'_n \mathbf{M}_{r,n} \mathbf{Z}_{g,n}$ are finite and have full column rank.
- (c) Let $\mathbf{Q}_{HZ,g^*}(\boldsymbol{\rho}_{g,n}) = \mathbf{Q}_{HZ,g} - \sum_{r \in \mathbf{I}_{g,\rho}} \rho_{g,r,n} \mathbf{Q}_{HMZ,r,g}$, then $\lambda_{\min} [\mathbf{Q}_{HZ,g^*}(\boldsymbol{\rho}_{g,n})' \mathbf{Q}_{HH}^{-1} \mathbf{Q}_{HZ,g^*}(\boldsymbol{\rho}_{g,n})] \geq c$ for some $c > 0$.

The above assumptions are in the spirit of those maintained, e.g., in Kelejian and Prucha (1998, 2004, 2010) and Lee (2003). We first discuss Assumption 5. The best instruments for $\mathbf{Y}_{g,n}$ and $\bar{\mathbf{Y}}_{g,n}$ are given by their conditional means. Observe that in light of (6) we have

$$E\mathbf{y}_n = (\mathbf{I}_{nG} - \mathbf{B}_n^*)^{-1} \mathbf{C}_n^* \mathbf{x}_n$$

with $\mathbf{B}_n^* = [(\mathbf{B}'_n \otimes \mathbf{I}_n) + (\boldsymbol{\Lambda}'_n \otimes \mathbf{I}_n) (\mathbf{I}_G \otimes \mathbf{W}_n)]$. For large n the accurate computation of the inverse of $\mathbf{I}_{nG} - \mathbf{B}_n^*$, which is of dimension $nG \times nG$ and depends

¹⁵We note that our setup could be readily modified to accommodate for \mathbf{H}_n and for the $\mathbf{A}_{s,n}$ to vary with g at the expense of further complicating the notation.

on unknown parameters, will be challenging if not impossible, unless the weights matrices are sparse. Furthermore, even in the single equation case existing results on the asymptotic properties of GMM estimators based on the best instruments have so far only been obtained by restricting the parameter space to a compact interval in, say, $(-1, 1)$. To avoid those difficulties and limitations we employ an approximation of the best instruments, which is consistent in spirit with the approach adopted in the above cited literature.

Given $\|\mathbf{B}_n^*\| < 1$ we have $(\mathbf{I}_{nG} - \mathbf{B}_n^*)^{-1} = \sum_{d=0}^{\infty} (\mathbf{B}_n^*)^d$ and thus $E\mathbf{y}_n = \sum_{d=0}^{\infty} (\mathbf{B}_n^*)^d \mathbf{C}_n^* \mathbf{x}_n$. In light of the structure of \mathbf{B}_n^* it is not difficult to see that the blocks of $(\mathbf{I}_{nG} - \mathbf{B}_n^*)^{-1}$ can be expressed as infinite weighted sums of the matrices \mathbf{I}_n , $\{\mathbf{W}_{j_1, n}\}_{j_1=1}^p$, $\{\mathbf{W}_{j_1, n} \mathbf{W}_{j_2, n}\}_{j_1, j_2=1}^p$, $\{\mathbf{W}_{j_1, n} \mathbf{W}_{j_2, n} \mathbf{W}_{j_3, n}\}_{j_1, j_2, j_3=1}^p$, \dots . Adopting the notation $[\mathbf{E}_j]_{j=1}^m := [\mathbf{E}_1, \dots, \mathbf{E}_m]$ for any set of conformable matrices $\mathbf{E}_1, \dots, \mathbf{E}_m$, define

$$\begin{aligned} \bar{\mathbf{X}}_{1, n} &= [\mathbf{W}_{j_1, n} \mathbf{X}_n]_{j_1=1}^p, \\ \bar{\mathbf{X}}_{2, n} &= [\mathbf{W}_{j_1, n} \mathbf{W}_{j_2, n} \mathbf{X}_n]_{j_1, j_2=1}^p, \\ &\vdots \\ \bar{\mathbf{X}}_{R, n} &= [\mathbf{W}_{j_1, n} \mathbf{W}_{j_2, n} \dots \mathbf{W}_{j_R, n} \mathbf{X}_n]_{j_1, j_2, \dots, j_R=1}^p, \end{aligned}$$

and let $\mathbf{H}_{R, n} = [\mathbf{X}_n, \bar{\mathbf{X}}_{1, n}, \dots, \bar{\mathbf{X}}_{R, n}]$. Now suppose $\mathbf{W}_{s_1}, \dots, \mathbf{W}_{s_g}$ are the spatial weights matrices appearing in $\bar{\mathbf{Y}}_{g, n}$, then by including in \mathbf{H}_n the linearly independent columns of $[\mathbf{H}_{R, n}, \mathbf{W}_{s_1} \mathbf{H}_{R, n}, \dots, \mathbf{W}_{s_g} \mathbf{H}_{R, n}]$, we may view the fitted values of $\mathbf{Z}_{g, n}$ as computationally simple approximations of the best instruments $E\mathbf{Z}_{g, n}$. Suppose further that $\mathbf{M}_{r_1}, \dots, \mathbf{M}_{r_g}$ are the spatial weights matrices in the disturbance process of the g -th equation, then by including in \mathbf{H}_n also the linearly independent columns of $\mathbf{M}_{r_1} [\mathbf{H}_{R, n}, \mathbf{W}_{s_1} \mathbf{H}_{R, n}, \dots, \mathbf{W}_{s_g} \mathbf{H}_{R, n}], \dots, \mathbf{M}_{r_g} [\mathbf{H}_{R, n}, \mathbf{W}_{s_1} \mathbf{H}_{R, n}, \dots, \mathbf{W}_{s_g} \mathbf{H}_{R, n}]$ we may view the fitted values of $\mathbf{Z}_{*g, n}$ as computationally simple approximations of the best instruments $E\mathbf{Z}_{*g, n}$. The Monte Carlo results presented in the Online Supplementary Appendix suggest that in many situations relatively low values of R are sufficient for providing a good approximation. A discussion, in a simplified context, as to how instruments are generated from within the model is given in Appendix F in the Online Supplementary Appendix.

Assumption 6(a) is standard. Assumption 6(b) is a sufficient condition that ensures the identification of $\boldsymbol{\delta}_{g, n}$ from linear moment conditions corresponding to the untransformed model (8). Assumption 6(c) is used to ensure identification

from linear moment conditions corresponding to the transformed model (10). Those assumptions are crucial for the consistency of the first-step estimators of $\delta_{g,n}$ from linear moment conditions only. A more detailed discussion of those assumptions, including scenarios where those conditions will not hold, are provided in Appendix F in the Online Supplementary Appendix. Of course, within the context of one-step estimation identification is still possible with the use of the quadratic moment conditions, even if identification by the linear moment conditions fails. More detailed remarks and references are provided in Appendix F in the Online Supplementary Appendix.

We will maintain the following assumptions regarding the matrices $\mathbf{A}_{s,n}$ in the quadratic moment conditions (12).

Assumption 7 : *The row and column sums of the matrices $\mathbf{A}_{s,n}$, $s = 1, \dots, S$, are bounded uniformly in absolute value by some finite constant and, furthermore, all diagonal elements of $\mathbf{A}_{s,n}$ are zero for any $s = 1, \dots, S$.*

The assumptions that the diagonal elements of $\mathbf{A}_{s,n}$ are zero ensures that the moment conditions are robust against heteroskedasticity. Exemplary specifications for $\mathbf{A}_{s,n}$ include

$$\begin{aligned} \mathbf{M}_{r,n}, & \quad \mathbf{M}'_{r,n} \mathbf{M}_{r,n} - \text{diag}(\mathbf{M}'_{r,n} \mathbf{M}_{r,n}), \\ \mathbf{W}_{s,n}, & \quad \mathbf{W}'_{s,n} \mathbf{W}_{s,n} - \text{diag}(\mathbf{W}'_{s,n} \mathbf{W}_{s,n}), \quad \mathbf{M}'_{s,n} \mathbf{W}_{s,n} - \text{diag}(\mathbf{M}'_{s,n} \mathbf{W}_{s,n}). \end{aligned}$$

For computational purposes and for proving consistency, it is convenient to re-write the moment conditions in (13) as

$$\boldsymbol{\gamma}_{g,n} - \boldsymbol{\Gamma}_{g,n} \mathbf{r}_{g,n}(\boldsymbol{\rho}_{g,n}) = \mathbf{0}, \quad (14)$$

where

$$\boldsymbol{\gamma}_{g,n} = \begin{bmatrix} \gamma_{1,g,n} \\ \vdots \\ \gamma_{S,g,n} \end{bmatrix}_{S \times 1}, \quad \boldsymbol{\Gamma}_{g,n} = \begin{bmatrix} \boldsymbol{\Gamma}_{11,g,n} & \boldsymbol{\Gamma}_{12,g,n} & \boldsymbol{\Gamma}_{13,g,n} \\ \vdots & \vdots & \vdots \\ \boldsymbol{\Gamma}_{S1,g,n} & \boldsymbol{\Gamma}_{S2,g,n} & \boldsymbol{\Gamma}_{S3,g,n} \end{bmatrix}_{S \times q_g^*}, \quad \mathbf{r}_{g,n}(\boldsymbol{\rho}_{g,n}) = \begin{bmatrix} \mathbf{r}_{1,g,n} \\ \mathbf{r}_{2,g,n} \\ \mathbf{r}_{3,g,n} \end{bmatrix}_{q_g^* \times 1}$$

with

$$\begin{aligned}
\gamma_{s,g,n} &= n^{-1} E \mathbf{u}'_{g,n} \bar{\mathbf{A}}_{s,n} \mathbf{u}_{g,n}, \\
\Gamma_{s1,g,n} &= n^{-1} (2E \mathbf{u}'_{g,n} \mathbf{M}'_{r_{g,1},n} \bar{\mathbf{A}}_{s,n} \mathbf{u}_{g,n}, \dots, 2E \mathbf{u}'_{g,n} \mathbf{M}'_{r_{g,q_g},n} \bar{\mathbf{A}}_{s,n} \mathbf{u}_{g,n}), \\
\Gamma_{s2,g,n} &= -n^{-1} (E \mathbf{u}'_{g,n} \mathbf{M}'_{r_{g,1},n} \bar{\mathbf{A}}_{s,n} \mathbf{M}_{r_{g,1},n} \mathbf{u}_{g,n}, \dots, E \mathbf{u}'_{g,n} \mathbf{M}'_{r_{g,q_g},n} \bar{\mathbf{A}}_{s,n} \mathbf{M}_{r_{g,q_g},n} \mathbf{u}_{g,n}), \\
\Gamma_{s3,g,n} &= -n^{-1} (2E \mathbf{u}'_{g,n} \mathbf{M}'_{r_{g,1},n} \bar{\mathbf{A}}_{s,n} \mathbf{M}_{r_{g,2},n} \mathbf{u}_{g,n}, \dots, 2E \mathbf{u}'_{g,n} \mathbf{M}'_{r_{g,q_g-1},n} \bar{\mathbf{A}}_{s,n} \mathbf{M}_{r_{g,q_g},n} \mathbf{u}_{g,n}), \\
\mathbf{r}_{1,g,n} &= (\rho_{g,r_{g,1},n}, \dots, \rho_{g,r_{g,q_g},n})', \\
\mathbf{r}_{2,g,n} &= (\rho_{g,r_{g,1},n}^2, \dots, \rho_{g,r_{g,q_g},n}^2)', \\
\mathbf{r}_{3,g,n} &= (\rho_{g,r_{g,1},n} \rho_{g,r_{g,2},n}, \dots, \rho_{g,r_{g,1},n} \rho_{g,r_{g,q_g},n}, \dots, \rho_{g,r_{g,q_g-1},n} \rho_{g,r_{g,q_g},n})',
\end{aligned}$$

recalling the definition of the index set $\mathbf{I}_{g,\rho} = \{r_{g,1}, \dots, r_{g,q_g}\} \subseteq \{1, \dots, q\}$ and where $q_g^* = 2q_g + q_g(q_g - 1)/2$.¹⁶

For the case of two-step estimation, let $\tilde{\boldsymbol{\delta}}_{g,n}$ be some estimator for $\boldsymbol{\delta}_{g,n}$, let $\tilde{\mathbf{u}}_{g,n} = \mathbf{y}_{g,n} - \mathbf{Z}_{g,n} \tilde{\boldsymbol{\delta}}_{g,n}$, and let $\tilde{\Gamma}_{g,n}$ and $\tilde{\gamma}_{g,n}$ denote the corresponding estimators of $\Gamma_{g,n}$ and $\gamma_{g,n}$, respectively, which are obtained by suppressing the expectations operator and replacing $\mathbf{u}_{g,n}$ by $\tilde{\mathbf{u}}_{g,n}$ in the above expressions. Then

$$\mathbf{m}_{g,n}^\rho(\boldsymbol{\rho}_g, \tilde{\boldsymbol{\delta}}_{g,n}) = \tilde{\gamma}_{g,n} - \tilde{\Gamma}_{g,n} \mathbf{r}_{g,n}(\boldsymbol{\rho}_g). \quad (15)$$

4 Generic Asymptotic Properties

In this section we give a generic discussion of the asymptotic properties of GMM estimators for $\boldsymbol{\rho}_{g,n}$ and $\boldsymbol{\delta}_{g,n}$ corresponding to the moment conditions in (12) and (13). A main focus is on two-step estimators. Two-step estimators are appealing, since they are computationally simple, and since in a single-equation context the loss of efficiency has been found to be small under various reasonable scenarios, provided that the instruments are not weak; see, e.g. Das et al. (2003) and the Monte Carlo results presented in the Online Supplementary Appendix. A two-step procedure may also provide some robustness for the estimation of $\boldsymbol{\delta}_{g,n}$ against misspecification of the disturbance process. As will be seen below, the limiting distribution of the GMM estimators for $\boldsymbol{\rho}_{g,n}$ will depend on the

¹⁶To clarify the double indexing notation, consider the example where $\mathbf{I}_{g,\rho} = \{2, 4, 5\}$ and thus $q_g = 3$. In this case, and dropping the subscript n , $\mathbf{r}_{1,g} = (\rho_{g,2}, \rho_{g,4}, \rho_{g,5})'$, $\mathbf{r}_{2,g} = (\rho_{g,2}^2, \rho_{g,4}^2, \rho_{g,5}^2)'$ and $\mathbf{r}_{3,g} = (\rho_{g,2}\rho_{g,4}, \rho_{g,2}\rho_{g,5}, \rho_{g,4}\rho_{g,5})'$. The weight matrices in $\Gamma_{s1,g}$, $\Gamma_{s2,g}$, and $\Gamma_{s3,g}$ are indexed analogously.

estimator of $\boldsymbol{\delta}_{g,n}$ used in computing the estimated residual. That is, $\boldsymbol{\delta}_{g,n}$ is not a nuisance parameter for $\boldsymbol{\rho}_{g,n}$ (although the reverse is true). As a consequence, the derivation of the limiting distribution of the two-step estimators is technically more challenging. We will discuss one-step estimators later on in the paper. In essence the results in this section also deliver the limiting distribution of one-step estimators as a special case.

The GMM estimators for the autoregressive parameter vectors $\boldsymbol{\rho}_n$ introduced below generalize the GMM estimators in Kelejian and Prucha (2004). In contrast to Kelejian and Prucha (2004), we not only accommodate higher-order spatial disturbance processes, but we also provide for a full asymptotic theory for those estimators. As a result, we are able to introduce more efficient estimators, and provide results on the joint limiting distribution of the estimators for all model parameters.

4.1 Consistency of GMM Estimator for $\boldsymbol{\rho}$

Let $\tilde{\boldsymbol{\Upsilon}}_{gg,n}$ be some $S \times S$ symmetric positive semidefinite (moments) weighting matrix. Then a corresponding GMM estimator for $\boldsymbol{\rho}_{g,n}$ can be defined as

$$\tilde{\boldsymbol{\rho}}_{g,n} = \tilde{\boldsymbol{\rho}}_{g,n}(\tilde{\boldsymbol{\Upsilon}}_{gg,n}) = \underset{\bar{\boldsymbol{\rho}}_g: \sum_{r \in \mathbf{I}_{g,\rho}} |\bar{\rho}_{g,r}| \in [-a_\rho, a_\rho]}{\operatorname{argmin}} \quad \mathbf{m}_{g,n}^\rho(\bar{\boldsymbol{\rho}}_g, \tilde{\boldsymbol{\delta}}_{g,n})' \tilde{\boldsymbol{\Upsilon}}_{gg,n} \mathbf{m}_{g,n}^\rho(\bar{\boldsymbol{\rho}}_g, \tilde{\boldsymbol{\delta}}_{g,n}) \quad (16)$$

with $a_\rho \geq 1$. We note that the objective function for $\tilde{\boldsymbol{\rho}}_{g,n}$ remains well defined even for values of $\bar{\boldsymbol{\rho}}_g$ for which $\mathbf{I}_n - \mathbf{R}_{g,n}^*(\bar{\boldsymbol{\rho}}_g)$ is singular, which allows us to take as the optimization space a compact set containing the true parameter space.

We postulate the following additional assumption to establish consistency of $\tilde{\boldsymbol{\rho}}_{g,n}$.

Assumption 8 *The smallest eigenvalue of $\boldsymbol{\Gamma}'_{g,n} \boldsymbol{\Gamma}_{g,n}$ is uniformly bounded away from zero.*

Assumption 9 *$\tilde{\boldsymbol{\Upsilon}}_{g,n} - \boldsymbol{\Upsilon}_{g,n} = o_p(1)$, where $\boldsymbol{\Upsilon}_{g,n}$ is an $S \times S$ non-stochastic symmetric positive definite matrix. The largest eigenvalues of $\boldsymbol{\Upsilon}_{g,n}$ are bounded uniformly from above, and the smallest eigenvalues of $\boldsymbol{\Upsilon}_{g,n}$ are bounded uniformly away from zero (and, thus, by the equivalence of matrix norms, $\boldsymbol{\Upsilon}_{g,n}$ and $\boldsymbol{\Upsilon}_{g,n}^{-1}$ are $O(1)$).*

Assumption 8 requires $\boldsymbol{\Gamma}'_{g,n} \boldsymbol{\Gamma}_{g,n}$ to be nonsingular and in conjunction with Assumption 9 ensures that the smallest eigenvalue of $\boldsymbol{\Gamma}'_{g,n} \boldsymbol{\Upsilon}_{g,n} \boldsymbol{\Gamma}_{g,n}$ is uniformly

bounded away from zero. This will be sufficient to demonstrate that $\rho_{g,n}$ is identifiable unique w.r.t. the nonstochastic analogue of the objective function of the GMM estimator¹⁷

$$\begin{aligned} & \mathbf{E} \mathbf{m}_{g,n}^\rho(\bar{\rho}_g, \delta_{g,n})' \Upsilon_{g,n} \mathbf{E} \mathbf{m}_{g,n}^\rho(\bar{\rho}_g, \delta_{g,n}) \\ &= [\gamma_{g,n} - \Gamma_{g,n} \mathbf{r}_{g,n}(\bar{\rho}_{g,n})]' \Upsilon_{g,n} [\gamma_{g,n} - \Gamma_{g,n} \mathbf{r}_{g,n}(\bar{\rho}_{g,n})]. \end{aligned}$$

Some additional remarks on this assumption and connections to the GMM literature are provided in Appendix F in the Online Supplementary Appendix. Assumption 9 ensures that $\tilde{\Upsilon}_{g,n}$ is positive definite with probability trending to one. Of course, Assumption 9 is satisfied for $\tilde{\Upsilon}_{g,n} = \Upsilon_{g,n} = \mathbf{I}_S$. In this case, the estimator defined by (16) reduces to a nonlinear least squares estimator. Choices of $\tilde{\Upsilon}_{g,n}$ which result in efficient estimates of $\rho_{g,n}$ will be discussed below in conjunction with the asymptotic normality result.

Our basic consistency result for $\tilde{\rho}_{g,n}$ is given by the next theorem.

Theorem 1 (*Consistency*) *Let $\tilde{\rho}_{g,n} = \tilde{\rho}_{g,n}(\tilde{\Upsilon}_{g,n})$ denote the GMM estimator for $\rho_{g,n}$ defined by (16). Suppose Assumptions 1-9 hold, and suppose that $n^{1/2}(\tilde{\delta}_{g,n} - \delta_{g,n}) = O_p(1)$, then,*

$$\tilde{\rho}_{g,n} - \rho_{g,n} \xrightarrow{p} \mathbf{0} \text{ as } n \rightarrow \infty.$$

4.2 Asymptotic Distribution of GMM Estimator for ρ

The limiting distribution of the GMM estimator $\tilde{\rho}_{g,n}$ will generally depend on the estimator $\tilde{\delta}_{g,n}$ used to compute estimated disturbances. To define GMM estimators for $\delta_{g,n}$ we can employ the moment conditions (12). Leading examples for limited and full information GMM estimators for $\delta_{g,n}$ will be the spatial 2SLS and 3SLS estimators defined in the next section. To keep the discussion general we maintain the following assumption regarding $\tilde{\delta}_{g,n}$.

Assumption 10 *The estimator $\tilde{\delta}_{g,n}$ is asymptotically linear in ε_n in the sense that*

$$n^{1/2}(\tilde{\delta}_{g,n} - \delta_{g,n}) = n^{-1/2} \sum_{h=1}^G \mathbf{T}'_{gh,n} \varepsilon_{h,n} + o_p(1)$$

with $\mathbf{T}_{gh,n} = \mathbf{F}_{gh,n} \mathbf{P}_{gh,n}$, where $\mathbf{F}_{gh,n}$ and $\mathbf{P}_{gh,n}$ are, respectively, $n \times p_F$ and $p_F \times p_{\delta_g}$ real non-stochastic matrices whose elements are uniformly bounded

¹⁷For a definition of identifiable uniqueness, see, e.g., Definition 3.1 in Pötscher and Prucha (1997).

in absolute value by a finite constant, and where p_{δ_g} is the dimension of the parameter vector $\delta_{g,n}$. (We note that under the maintained assumptions the elements of $\mathbf{T}_{gh,n}$ are again uniformly bounded in absolute value by a finite constant).

In the Appendix A we show that our spatial 2SLS and 3SLS estimators satisfy this assumption. For 2SLS estimators of the parameters of the, say, first equation, $\sum_{h=1}^G \mathbf{T}'_{1h,n} \boldsymbol{\varepsilon}_{h,n}$ reduces to $\mathbf{T}'_{11,n} \boldsymbol{\varepsilon}_{1,n}$. Also note that under Assumptions 4 and 10 we have $n^{1/2}(\tilde{\boldsymbol{\delta}}_{g,n} - \boldsymbol{\delta}_{g,n}) = O_p(1)$, as is assumed by Theorem 1.

In preparation of our result concerning the asymptotic distribution of the GMM estimator we next define the matrices that will compose the limiting VC matrix. In particular, consider the $S \times q_g$ matrix

$$\mathbf{J}_{g,n} = -E \frac{\partial \mathbf{m}_{g,n}}{\partial \boldsymbol{\rho}_{g,n}} = \boldsymbol{\Gamma}_{g,n} \frac{\partial \mathbf{r}_{g,n}(\boldsymbol{\rho}_{n,g})}{\partial \boldsymbol{\rho}_{g,n}}, \quad (17)$$

and the $S \times S$ matrix $\boldsymbol{\Psi}_{gg,n}^{\rho\rho} = (\psi_{rs,gg,n}^{\rho\rho})$ where

$$\begin{aligned} \psi_{rs,gg,n}^{\rho\rho} &= \sigma_{gg}^2 (2n)^{-1} \text{tr} [(\mathbf{A}_{r,n} + \mathbf{A}'_{r,n})(\mathbf{A}_{s,n} + \mathbf{A}'_{s,n})] \\ &\quad + n^{-1} \boldsymbol{\alpha}'_{g,r,n} \left[\sum_{h=1}^G \sum_{l=1}^G \sigma_{hl} \mathbf{T}'_{gh,n} \mathbf{T}_{gl,n} \right] \boldsymbol{\alpha}_{g,s,n}, \end{aligned} \quad (18)$$

with

$$\boldsymbol{\alpha}_{g,r,n} = -n^{-1} E [\mathbf{Z}'_{g,n} (\mathbf{I}_n - \mathbf{R}_{g,n}^*(\boldsymbol{\rho}_{g,n})) (\mathbf{A}_{r,n} + \mathbf{A}'_{r,n}) (\mathbf{I}_n - \mathbf{R}_{g,n}^*(\boldsymbol{\rho}_{g,n})) \mathbf{u}_{g,n}].$$

As shown in the proof of the subsequent theorem, $\boldsymbol{\Psi}_{gg,n}^{\rho\rho}$ is the asymptotic VC matrix of the sample moment vector $\mathbf{m}_{g,n}^\rho(\boldsymbol{\rho}_g, \tilde{\boldsymbol{\delta}}_{g,n})$. The second term in (18) stems from the fact that the sample moment vector depends on estimated residuals. If the true residuals could be observed, we could take $\tilde{\boldsymbol{\delta}}_{g,n} = \boldsymbol{\delta}_{g,n}$ or $\mathbf{T}_{gh,n} = \mathbf{0}$, in which case the second term in (18) is zero.

Theorem 2 (Asymptotic Normality) *Let $\tilde{\boldsymbol{\rho}}_{g,n} = \tilde{\boldsymbol{\rho}}_{g,n}(\tilde{\boldsymbol{\Upsilon}}_{g,n})$ denote the GMM estimator for $\boldsymbol{\rho}_{g,n}$ defined by (16). Then given Assumptions 1-10, and given that $\lambda_{\min}(\boldsymbol{\Psi}_{gg,n}^{\rho\rho}) \geq c_\Psi^* > 0$, we have*

$$n^{1/2}(\tilde{\boldsymbol{\rho}}_{g,n} - \boldsymbol{\rho}_{g,n}) = [\mathbf{J}'_{g,n} \boldsymbol{\Upsilon}_{g,n} \mathbf{J}_{g,n}]^{-1} \mathbf{J}'_{g,n} \boldsymbol{\Upsilon}_{g,n} \boldsymbol{\xi}_{g,n} + o_p(1), \quad (19)$$

where

$$\boldsymbol{\xi}_{g,n} = \begin{bmatrix} \xi_{g1,n} \\ \vdots \\ \xi_{gS,n} \end{bmatrix} = -n^{-1/2} \begin{bmatrix} \frac{1}{2}\boldsymbol{\varepsilon}'_{g,n}(\mathbf{A}_{1,n} + \mathbf{A}'_{1,n})\boldsymbol{\varepsilon}_{g,n} + \boldsymbol{\alpha}'_{g,1,n} \sum_{h=1}^G \mathbf{T}'_{gh,n} \varepsilon_{h,n} \\ \vdots \\ \frac{1}{2}\boldsymbol{\varepsilon}'_{g,n}(\mathbf{A}_{S,n} + \mathbf{A}'_{S,n})\boldsymbol{\varepsilon}_{g,n} + \boldsymbol{\alpha}'_{g,S,n} \sum_{h=1}^G \mathbf{T}'_{gh,n} \varepsilon_{h,n} \end{bmatrix} \quad (20)$$

and $(\boldsymbol{\Psi}_{gg,n}^{\rho\rho})^{-1/2}\boldsymbol{\xi}_{g,n} \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_S)$. Furthermore $n^{1/2}(\tilde{\boldsymbol{\rho}}_{g,n} - \boldsymbol{\rho}_{g,n}) = O_p(1)$ and $\lambda_{\min}[\boldsymbol{\Omega}_{gg,n}^{\rho\rho}(\boldsymbol{\Upsilon}_{g,n})] \geq \text{const} > 0$ for

$$\boldsymbol{\Omega}_{gg,n}^{\rho\rho}(\boldsymbol{\Upsilon}_{g,n}) = (\mathbf{J}'_{g,n} \boldsymbol{\Upsilon}_{g,n} \mathbf{J}_{g,n})^{-1} \mathbf{J}'_{g,n} \boldsymbol{\Upsilon}_{g,n} \boldsymbol{\Psi}_{gg,n}^{\rho\rho} \boldsymbol{\Upsilon}_{g,n} \mathbf{J}_{g,n} (\mathbf{J}'_{g,n} \boldsymbol{\Upsilon}_{g,n} \mathbf{J}_{g,n})^{-1}. \quad (21)$$

The above theorem implies that the difference between the cumulative distribution function of $n^{1/2}(\tilde{\boldsymbol{\rho}}_{g,n} - \boldsymbol{\rho}_{g,n})$ and that of $N[\mathbf{0}, \boldsymbol{\Omega}_{\tilde{\boldsymbol{\rho}}_{g,n}}]$ converges pointwise to zero, which justifies the use of the latter distribution as an approximation of the former.¹⁸

Remark. Clearly, $\boldsymbol{\Omega}_{gg,n}^{\rho\rho}((\boldsymbol{\Psi}_{gg,n}^{\rho\rho})^{-1}) = [\mathbf{J}'_{g,n}(\boldsymbol{\Psi}_{gg,n}^{\rho\rho})^{-1}\mathbf{J}_{g,n}]^{-1}$ and $\boldsymbol{\Omega}_{gg,n}^{\rho\rho}(\boldsymbol{\Upsilon}_{g,n}) - \boldsymbol{\Omega}_{gg,n}^{\rho\rho}((\boldsymbol{\Psi}_{gg,n}^{\rho\rho})^{-1})$ is positive semi-definite for any $\boldsymbol{\Upsilon}_{g,n}$. Thus, choosing $\tilde{\boldsymbol{\Upsilon}}_{g,n}$ as a consistent estimator for $(\boldsymbol{\Psi}_{gg,n}^{\rho\rho})^{-1}$ leads to the efficient GMM estimator. As discussed in the proof of the above theorem, the elements of $\boldsymbol{\Psi}_{gg,n}^{\rho\rho}$ are uniformly bounded in absolute value and, hence, $\lambda_{\max}(\boldsymbol{\Psi}_{gg,n}^{\rho\rho}) \leq c_{\Psi}^{**}$ for some $c_{\Psi}^{**} < \infty$. Since by assumption also $0 < c_{\Psi}^* \leq \lambda_{\min}(\boldsymbol{\Psi}_{g,n}^{\rho\rho})$, it follows that the conditions on the eigenvalues of $\boldsymbol{\Upsilon}_{g,n}$ postulated in Assumption 9 are automatically satisfied by $(\boldsymbol{\Psi}_{gg,n}^{\rho\rho})^{-1}$.

We next define a consistent estimator for $\boldsymbol{\Omega}_{gg,n}^{\rho\rho}(\boldsymbol{\Upsilon}_{g,n})$. As a preliminary result we have the following lemma.

Lemma 1 : *Suppose Assumptions 1-4 hold. For $g, h = 1, \dots, G$ define*

$$\tilde{\sigma}_{gh,n} = n^{-1} \tilde{\boldsymbol{\varepsilon}}'_{g,n} \tilde{\boldsymbol{\varepsilon}}_{h,n} \quad (22)$$

with $\tilde{\boldsymbol{\varepsilon}}_{g,n} = \mathbf{y}_{*g,n}(\tilde{\boldsymbol{\rho}}_{g,n}) - \mathbf{Z}_{*g,n}(\tilde{\boldsymbol{\rho}}_{g,n})\tilde{\boldsymbol{\delta}}_{g,n}$ and assume $\tilde{\boldsymbol{\delta}}_{g,n} - \boldsymbol{\delta}_{g,n} = o_p(1)$ and $\tilde{\boldsymbol{\rho}}_{g,n} - \boldsymbol{\rho}_{g,n} = o_p(1)$, then $\tilde{\sigma}_{gh,n} - \sigma_{gh} = o_p(1)$.

We note that in the above lemma $\tilde{\boldsymbol{\rho}}_{g,n}$ and $\tilde{\boldsymbol{\delta}}_{g,n}$ can be any consistent esti-

¹⁸This follows from Corollary F4 in Pötscher and Prucha (1997). Compare also the discussion on pp. 86-87 in that reference.

mators. Now let $\tilde{\Gamma}_{g,n}$ be defined as in (15) and corresponding to (17) define

$$\tilde{\mathbf{J}}_{g,n} = \tilde{\Gamma}_{g,n} \frac{\partial \mathbf{r}_{g,n}(\tilde{\boldsymbol{\rho}}_{n,g})}{\partial \boldsymbol{\rho}_{g,n}}. \quad (23)$$

Furthermore, corresponding to (18) consider the following estimator $\tilde{\Psi}_{gg,n}^{\rho\rho} = \left(\tilde{\psi}_{rs,gg,n}^{\rho\rho} \right)$ for $\Psi_{gg,n}^{\rho\rho}$, where

$$\begin{aligned} \tilde{\psi}_{rs,gg,n}^{\rho\rho} &= \tilde{\sigma}_{gg,n}^2 (2n)^{-1} \text{tr} [(\mathbf{A}_{r,n} + \mathbf{A}'_{r,n})(\mathbf{A}_{s,n} + \mathbf{A}'_{s,n})] \\ &+ n^{-1} \tilde{\boldsymbol{\alpha}}'_{g,r,n} \left[\sum_{h=1}^G \sum_{l=1}^G \tilde{\sigma}_{hl,n} \tilde{\mathbf{T}}'_{gh,n} \tilde{\mathbf{T}}_{gl,n} \right] \tilde{\boldsymbol{\alpha}}_{g,s,n}, \end{aligned} \quad (24)$$

with

$$\tilde{\boldsymbol{\alpha}}_{g,r,n} = -n^{-1} [\mathbf{Z}'_{g,n}(\mathbf{I}_n - \mathbf{R}_{g,n}^*(\tilde{\boldsymbol{\rho}}_{g,n}))(\mathbf{A}_{r,n} + \mathbf{A}'_{r,n})(\mathbf{I}_n - \mathbf{R}_{g,n}^*(\tilde{\boldsymbol{\rho}}_{g,n}))\tilde{\mathbf{u}}_{g,n}],$$

where $\tilde{\mathbf{T}}_{gh,n}$ is some estimator for $\mathbf{T}_{gh,n}$, and $\tilde{\mathbf{u}}_{g,n} = \mathbf{y}_{g,n} - \mathbf{Z}_{g,n}\tilde{\boldsymbol{\delta}}_{g,n}$. Given estimators for $\mathbf{J}_{g,n}$ and $\Psi_{gg,n}^{\rho\rho}$ we can now formulate the following estimator for $\Omega_{gg,n}^{\rho\rho}$:

$$\tilde{\Omega}_{gg,n}^{\rho\rho}(\tilde{\Upsilon}_{g,n}) = (\tilde{\mathbf{J}}'_{g,n} \tilde{\Upsilon}_{g,n} \tilde{\mathbf{J}}_{g,n})^{-1} \tilde{\mathbf{J}}'_{g,n} \tilde{\Upsilon}_{g,n} \tilde{\Psi}_{gg,n}^{\rho\rho} \tilde{\Upsilon}_{g,n} \tilde{\mathbf{J}}_{g,n} (\tilde{\mathbf{J}}'_{g,n} \tilde{\Upsilon}_{g,n} \tilde{\mathbf{J}}_{g,n})^{-1}. \quad (25)$$

The next theorem establishes the consistency of $\tilde{\Psi}_{g,n}^{\rho\rho}$ and $\tilde{\Omega}_{gg,n}^{\rho\rho}$.

Theorem 3 (VC Matrix Estimation) *Suppose all assumptions of Theorem 2 hold, except for Assumption 9, and suppose that $n^{-1}\tilde{\mathbf{T}}'_{gh,n} \tilde{\mathbf{T}}_{gl,n} - n^{-1}\mathbf{T}'_{gh,n} \mathbf{T}_{gl,n} = o_p(1)$. Then, provided that $\tilde{\boldsymbol{\rho}}_{g,n} - \boldsymbol{\rho}_{g,n} = o_p(1)$,*

$$\tilde{\Psi}_{gg,n}^{\rho\rho} - \Psi_{gg,n}^{\rho\rho} = o_p(1), (\tilde{\Psi}_{gg,n}^{\rho\rho})^{-1} - (\Psi_{gg,n}^{\rho\rho})^{-1} = o_p(1),$$

and $\tilde{\Psi}_{gg,n}^{\rho\rho} = O(1)$, $(\Psi_{gg,n}^{\rho\rho})^{-1} = O(1)$. *If furthermore Assumption 9 holds, then,*

$$\tilde{\Omega}_{gg,n}^{\rho\rho} - \Omega_{gg,n}^{\rho\rho} = o_p(1), (\tilde{\Omega}_{gg,n}^{\rho\rho})^{-1} - (\Omega_{gg,n}^{\rho\rho})^{-1} = o_p(1),$$

and $\tilde{\Omega}_{gg,n}^{\rho\rho} = O(1)$, $(\Omega_{gg,n}^{\rho\rho})^{-1} = O(1)$.

Note that for the first part of the above theorem $\tilde{\boldsymbol{\rho}}_{g,n}$ can be any consistent estimator.

4.3 Joint Asymptotic Distribution of Estimators for ρ and δ

In the following, let $\boldsymbol{\delta}_n = [\delta'_{1,n}, \dots, \delta'_{G,n}]'$ and $\boldsymbol{\rho}_n = [\rho'_{1,n}, \dots, \rho'_{G,n}]'$, and let $\tilde{\boldsymbol{\delta}}_n = [\tilde{\delta}'_{1,n}, \dots, \tilde{\delta}'_{G,n}]'$ and $\tilde{\boldsymbol{\rho}}_n = [\tilde{\rho}'_{1,n}, \dots, \tilde{\rho}'_{G,n}]'$ denote the corresponding estimators as defined in the previous subsections. In this section, we derive the joint limiting distribution of $\tilde{\boldsymbol{\delta}}_n$ and $\tilde{\boldsymbol{\rho}}_n$. Knowledge of the joint asymptotic distribution of all model parameters will then enable the researcher to test for the presence of network spillovers in any of the dependent variables, explanatory variables, or disturbances in the system.

As shown in the proof of the next theorem, the joint limiting distribution of the estimators will depend on the joint limiting distribution of the following vector of linear and quadratic forms $[\boldsymbol{\eta}'_n, \boldsymbol{\xi}'_n]'$ with $\boldsymbol{\eta}'_n = [\boldsymbol{\eta}'_{1,n}, \dots, \boldsymbol{\eta}'_{G,n}]'$ and $\boldsymbol{\xi}'_n = [\boldsymbol{\xi}'_{1,n}, \dots, \boldsymbol{\xi}'_{G,n}]'$, where $\boldsymbol{\eta}_{g,n} = n^{-1/2} \sum_{h=1}^G \mathbf{T}'_{gh,n} \boldsymbol{\varepsilon}_{h,n}$ and the $\boldsymbol{\xi}_{g,n}$ are defined in (20). Let

$$\mathbf{T}_n = \begin{bmatrix} \mathbf{T}_{11,n} & \cdots & \mathbf{T}_{G1,n} \\ \vdots & \ddots & \vdots \\ \mathbf{T}_{1G,n} & \cdots & \mathbf{T}_{GG,n} \end{bmatrix},$$

then the VC matrix of $[\boldsymbol{\eta}'_n, \boldsymbol{\xi}'_n]'$ is given by

$$\boldsymbol{\Psi}_n = \begin{bmatrix} \boldsymbol{\Psi}_n^{\delta\delta} & \boldsymbol{\Psi}_n^{\delta\rho} \\ \boldsymbol{\Psi}_n^{\delta\rho'} & \boldsymbol{\Psi}_n^{\rho\rho} \end{bmatrix}, \quad (26)$$

where

$$\begin{aligned} \boldsymbol{\Psi}_n^{\rho\rho} &= E\boldsymbol{\xi}_n\boldsymbol{\xi}'_n = \begin{bmatrix} \boldsymbol{\Psi}_{11,n}^{\rho\rho} & \cdots & \boldsymbol{\Psi}_{1G,n}^{\rho\rho} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Psi}_{G1,n}^{\rho\rho} & \cdots & \boldsymbol{\Psi}_{GG,n}^{\rho\rho} \end{bmatrix}, \\ \boldsymbol{\Psi}_n^{\delta\delta} &= E\boldsymbol{\eta}_n\boldsymbol{\eta}'_n = n^{-1}\mathbf{T}'_n(\boldsymbol{\Sigma} \otimes \mathbf{I}_n)\mathbf{T}_n, \\ \boldsymbol{\Psi}_n^{\delta\rho} &= E\boldsymbol{\eta}_n\boldsymbol{\xi}'_n = n^{-1}\mathbf{T}'_n(\boldsymbol{\Sigma} \otimes \mathbf{I}_n)\mathbf{T}_n \text{diag}_{g=1}^G [\boldsymbol{\alpha}_{g,1,n}, \dots, \boldsymbol{\alpha}_{g,S,n}], \end{aligned}$$

and where the (r, s) -th element of $\Psi_{gh,n}^{\rho\rho}$ is given by

$$\begin{aligned}\psi_{rs,gh,n}^{\rho\rho} &= \sigma_{gh,n}^2(2n)^{-1}tr [(\mathbf{A}_{r,n} + \mathbf{A}'_{r,n})(\mathbf{A}_{s,n} + \mathbf{A}'_{s,n})] \\ &\quad + n^{-1}\alpha'_{g,r,n} \left[\sum_{u=1}^G \sum_{v=1}^G \sigma_{uv,n} \mathbf{T}'_{gu,n} \mathbf{T}_{hv,n} \right] \alpha_{h,s,n}.\end{aligned}$$

Analogous to (24) consider the following estimator $\tilde{\Psi}_{gh,n}^{\rho\rho} = \left(\tilde{\psi}_{rs,gh,n}^{\rho\rho} \right)$ for $\Psi_{gh,n}^{\rho\rho}$, where

$$\begin{aligned}\tilde{\psi}_{rs,gh,n}^{\rho\rho} &= \tilde{\sigma}_{gh,n}^2(2n)^{-1}tr [(\mathbf{A}_{r,n} + \mathbf{A}'_{r,n})(\mathbf{A}_{s,n} + \mathbf{A}'_{s,n})] \\ &\quad + n^{-1}\tilde{\alpha}'_{g,r,n} \left[\sum_{u=1}^G \sum_{v=1}^G \tilde{\sigma}_{uv,n} \tilde{\mathbf{T}}'_{gu,n} \tilde{\mathbf{T}}_{hv,n} \right] \tilde{\alpha}_{h,s,n},\end{aligned}$$

with

$$\tilde{\alpha}_{g,r,n} = -n^{-1} [\mathbf{Z}'_{g,n}(\mathbf{I}_n - \mathbf{R}_{g,n}^*(\tilde{\rho}_{g,n}))(\mathbf{A}_{r,n} + \mathbf{A}'_{r,n})(\mathbf{I}_n - \mathbf{R}_{g,n}^*(\tilde{\rho}_{g,n}))\tilde{\mathbf{u}}_{g,n}],$$

with $\tilde{\mathbf{T}}_{gh,n}$ being some estimator for $\mathbf{T}_{gh,n}$, and $\tilde{\mathbf{u}}_{g,n} = \mathbf{y}_{g,n} - \mathbf{Z}_{g,n}\tilde{\delta}_{g,n}$. Next, consider the estimators

$$\begin{aligned}\tilde{\Psi}_n^{\delta\delta} &= n^{-1}\tilde{\mathbf{T}}'_n(\tilde{\Sigma}_n \otimes \mathbf{I}_n)\tilde{\mathbf{T}}_n, \\ \tilde{\Psi}_n^{\delta\rho} &= n^{-1}\tilde{\mathbf{T}}'_n(\tilde{\Sigma}_n \otimes \mathbf{I}_n)\tilde{\mathbf{T}}_n \text{diag}_{g=1}^G([\tilde{\alpha}_{g,1,n}, \dots, \tilde{\alpha}_{g,S,n}],\end{aligned}$$

then our estimator for Ψ_n can be formulated as

$$\tilde{\Psi}_n = \begin{bmatrix} \tilde{\Psi}_n^{\delta\delta} & \tilde{\Psi}_n^{\delta\rho} \\ \tilde{\Psi}_n^{\delta\rho'} & \tilde{\Psi}_n^{\rho\rho} \end{bmatrix}. \quad (27)$$

We now have the following theorem concerning the joint limiting distribution of $\tilde{\delta}_n - \delta_n$ and $\tilde{\rho}_n - \rho_n$.

Theorem 4 (Asymptotic Normality) Let $\tilde{\rho}_n = [\tilde{\rho}'_{1,n}, \dots, \tilde{\rho}'_{G,n}]'$ where $\tilde{\rho}_{g,n} = \tilde{\rho}_{g,n}(\tilde{\Upsilon}_{g,n})$ denotes the GMM estimator for $\rho_{g,n}$ defined by (16), and let $\tilde{\delta}_n = [\tilde{\delta}'_{1,n}, \dots, \tilde{\delta}'_{G,n}]'$ be an estimator for δ , where $\tilde{\delta}_{g,n}$ is asymptotically linear in ε_n . Then given Assumptions 1-10, $n^{-1}\tilde{\mathbf{T}}'_{gh,n}\tilde{\mathbf{T}}_{kl,n} - n^{-1}\mathbf{T}'_{gh,n}\mathbf{T}_{kl,n} = o_p(1)$, and given that $\lambda_{\min}(\Psi_n) \geq c$ for some $c > 0$, we have

$$n^{1/2} \begin{bmatrix} \tilde{\delta}_n - \delta_n \\ \tilde{\rho}_n - \rho_n \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \text{diag}_g[(\mathbf{J}'_{g,n}\Upsilon_{g,n}\mathbf{J}_{g,n})^{-1}\mathbf{J}'_{g,n}\Upsilon_{g,n}] \end{bmatrix} \begin{bmatrix} \boldsymbol{\eta}_n \\ \boldsymbol{\xi}_n \end{bmatrix} + o_p(1),$$

with

$$\Psi_n^{-1/2} \begin{bmatrix} \boldsymbol{\eta}_n \\ \boldsymbol{\xi}_n \end{bmatrix} \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_d),$$

where $d = \sum_{g=1}^G (G_g + K_g + p_g + q_g)$. Furthermore, let

$$\begin{aligned} \boldsymbol{\Omega}_n &= \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \text{diag}_{g=1}^G ((\mathbf{J}'_{g,n} \boldsymbol{\Upsilon}_{g,n} \mathbf{J}_{g,n})^{-1} \mathbf{J}'_{g,n} \boldsymbol{\Upsilon}_{g,n}) \end{bmatrix} \boldsymbol{\Psi}_n \\ &\times \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \text{diag}_{g=1}^G (\boldsymbol{\Upsilon}_{g,n} \mathbf{J}_{g,n} (\mathbf{J}'_{g,n} \boldsymbol{\Upsilon}_{g,n} \mathbf{J}_{g,n})^{-1}) \end{bmatrix}, \\ \tilde{\boldsymbol{\Omega}}_n &= \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \text{diag}_{g=1}^G ((\tilde{\mathbf{J}}'_{g,n} \tilde{\boldsymbol{\Upsilon}}_{g,n} \tilde{\mathbf{J}}_{g,n})^{-1} \tilde{\mathbf{J}}'_{g,n} \tilde{\boldsymbol{\Upsilon}}_{g,n}) \end{bmatrix} \tilde{\boldsymbol{\Psi}}_n \\ &\times \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \text{diag}_{g=1}^G (\tilde{\boldsymbol{\Upsilon}}_{g,n} \tilde{\mathbf{J}}_{g,n} (\tilde{\mathbf{J}}'_{g,n} \tilde{\boldsymbol{\Upsilon}}_{g,n} \tilde{\mathbf{J}}_{g,n})^{-1}) \end{bmatrix}. \end{aligned} \quad (28)$$

Then, $\tilde{\boldsymbol{\Psi}}_n - \boldsymbol{\Psi}_n = o_p(1)$, $\tilde{\boldsymbol{\Psi}}_n^{-1} - \boldsymbol{\Psi}_n^{-1} = o_p(1)$, $\boldsymbol{\Psi}_n = O(1)$, $\boldsymbol{\Psi}_n^{-1} = O(1)$ and $\tilde{\boldsymbol{\Omega}}_n - \boldsymbol{\Omega}_n = o_p(1)$, $\tilde{\boldsymbol{\Omega}}_n^{-1} - \boldsymbol{\Omega}_n^{-1} = o_p(1)$, $\boldsymbol{\Omega}_n = O(1)$, $\boldsymbol{\Omega}_n^{-1} = O(1)$.

Theorem 4 implies that the difference between the joint cumulative distribution function of the estimators of all model parameters $n^{1/2}[(\tilde{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_n)', (\tilde{\boldsymbol{\rho}}_n - \boldsymbol{\rho}_n)']'$ and that of $N(\mathbf{0}, \boldsymbol{\Omega}_n)$ converges pointwise to zero so that using the latter as an approximation of the former is justified.¹⁹ The theorem also states that $\tilde{\boldsymbol{\Omega}}_n$ is a consistent estimator of $\boldsymbol{\Omega}_n$. Of course, since the marginal distribution of a multivariate normal distribution is normal, the above theorem also establishes the limiting distribution of any subvector of $n^{1/2}[(\tilde{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_n)', (\tilde{\boldsymbol{\rho}}_n - \boldsymbol{\rho}_n)']'$.

5 Limited and Full Information Two-Step Estimators

In the previous section we developed generic results regarding the asymptotic properties of two-step GMM estimators for the parameters of model (1). The results show that the limiting distribution of GMM estimators for $\boldsymbol{\rho}_{g,n}$, which employ an initial estimator for $\boldsymbol{\delta}_{g,n}$ in computing estimated residuals, will generally depend on the limiting distribution of the latter. Thus, establishing the proper asymptotic theory for specific estimators is “delicate”. In this section,

¹⁹This follows from Corollary F4 in Pötscher and Prucha (1997). Compare also the discussion on pp. 86-87 in that reference.

we define specific limited and full information estimators and provide specific expressions for their asymptotic VC matrices.

5.1 Definition of Limited Information Estimators

In the following we define, in a sequence of steps, a specific generalized spatial two-stage least squares (GS2SLS) estimator of $\delta_{g,n}$ and a GMM estimator of $\rho_{g,n}$ based on GS2SLS residuals. WLOG we consider the estimation of the parameters of the g -th equation.

Step 1a: 2SLS estimator of $\delta_{g,n}$

As a first step, we apply 2SLS to the g -th equation of the untransformed model (8) using the instrument matrix \mathbf{H}_n in Assumptions 5 and 6 to estimate $\delta_{g,n}$. The 2SLS estimator, say $\tilde{\delta}_{g,n}$, is then defined as

$$\tilde{\delta}_{g,n} = (\tilde{\mathbf{Z}}'_{g,n} \mathbf{Z}_{g,n})^{-1} \tilde{\mathbf{Z}}'_{g,n} \mathbf{y}_{g,n}, \quad (29)$$

where $\tilde{\mathbf{Z}}_{g,n} = \mathbf{P}_{\mathbf{H}_n} \mathbf{Z}_{g,n} = (\tilde{\mathbf{Y}}_{g,n}, \mathbf{X}_{g,n}, \tilde{\mathbf{Y}}_{g,n})$, $\tilde{\mathbf{Y}}_{g,n} = \mathbf{P}_{\mathbf{H}_n} \mathbf{Y}_{g,n}$, $\tilde{\mathbf{Y}}_{g,n} = \mathbf{P}_{\mathbf{H}_n} \bar{\mathbf{Y}}_{g,n}$, and where $\mathbf{P}_{\mathbf{H}_n} = \mathbf{H}_n (\mathbf{H}'_n \mathbf{H}_n)^{-1} \mathbf{H}'_n$.²⁰

Step 1b: Initial GMM estimator of the vector $\rho_{g,n}$ based on 2SLS residuals

Let $\tilde{\mathbf{u}}_{g,n} = \mathbf{u}_{g,n}(\tilde{\delta}_{g,n}) = \mathbf{y}_{g,n} - \mathbf{Z}_{g,n} \tilde{\delta}_{g,n}$ denote the 2SLS residuals of the g -th equation, and let $\mathbf{m}_{g,n}^\rho(\bar{\rho}_{g,n}, \tilde{\delta}_{g,n})$ denote the corresponding sample moment vector as defined in (15). Our initial GMM estimator for $\rho_{g,n}$ is now defined as

$$\tilde{\rho}_{g,n} = \underset{\bar{\rho}_{g,n}: \sum_{r \in \mathbf{I}_{g,\rho}} |\bar{\rho}_{g,r}| \in [-a_\rho, a_\rho]}{\operatorname{argmin}} \mathbf{m}_{g,n}^\rho(\bar{\rho}_{g,n}, \tilde{\delta}_{g,n})' \mathbf{m}_{g,n}^\rho(\bar{\rho}_{g,n}, \tilde{\delta}_{g,n}) \quad (30)$$

with $a_\rho \geq 1$.

Step 2a: GS2SLS estimator of $\delta_{g,n}$

Analogous to Kelejian and Prucha (1998), we next compute a generalized spatial two-stage least squares (GS2SLS) estimator of $\delta_{g,n}$, $\hat{\delta}_{g,n}(\tilde{\rho}_{g,n})$. This estimator is defined as the 2SLS estimator of the g -th equation of the spatially Cochrane-Orcutt transformed model (10) with $\rho_{g,n}$ replaced by $\tilde{\rho}_{g,n}$, i.e.,

$$\hat{\delta}_{g,n} = \hat{\delta}_{g,n}(\tilde{\rho}_{g,n}) = [\hat{\mathbf{Z}}'_{*g,n}(\tilde{\rho}_{g,n})' \mathbf{Z}'_{*g,n}(\tilde{\rho}_{g,n})]^{-1} \hat{\mathbf{Z}}'_{*g,n}(\tilde{\rho}_{g,n})' \mathbf{y}_{*g,n}(\tilde{\rho}_{g,n}), \quad (31)$$

²⁰In the previous section we used tilde to denote generic estimators. In the following tilde is used to denote our initial 2SLS based estimators.

where $\mathbf{y}_{*g,n}(\tilde{\boldsymbol{\rho}}_{g,n}) = [\mathbf{I}_n - \mathbf{R}_{g,n}^*(\tilde{\boldsymbol{\rho}}_{g,n})] \mathbf{y}_{g,n}$, $\mathbf{Z}_{*g,n}(\tilde{\boldsymbol{\rho}}_{g,n}) = [\mathbf{I}_n - \mathbf{R}_{g,n}^*(\tilde{\boldsymbol{\rho}}_{g,n})] \mathbf{Z}_{g,n}$, and $\widehat{\mathbf{Z}}_{*g,n}(\tilde{\boldsymbol{\rho}}_{g,n}) = \mathbf{P}_{\mathbf{H}_n} \mathbf{Z}_{*g,n}(\tilde{\boldsymbol{\rho}}_{g,n})$. We shall also utilize the following estimator for the (g, g) -th block of $\boldsymbol{\Psi}_n^{\delta\delta}$ (corresponding to $\widehat{\boldsymbol{\delta}}_{g,n}$):

$$\widehat{\boldsymbol{\Psi}}_{gg,n}^{\delta\delta} = \widehat{\sigma}_{gg} [n^{-1} \widehat{\mathbf{Z}}_{*g,n}(\tilde{\boldsymbol{\rho}}_{g,n})' \widehat{\mathbf{Z}}_{*g,n}(\tilde{\boldsymbol{\rho}}_{g,n})]^{-1},$$

where $\widehat{\sigma}_{gg,n} = n^{-1} \widehat{\boldsymbol{\varepsilon}}'_{g,n} \widehat{\boldsymbol{\varepsilon}}_{g,n}$ with $\widehat{\boldsymbol{\varepsilon}}_{g,n} = \mathbf{y}_{*g,n}(\tilde{\boldsymbol{\rho}}_{g,n}) - \mathbf{Z}_{*g,n}(\tilde{\boldsymbol{\rho}}_{g,n}) \widehat{\boldsymbol{\delta}}_{g,n}$.

Step 2b: Efficient GMM estimator of $\boldsymbol{\rho}_{g,n}$ based on GS2SLS residuals

Let $\widehat{\mathbf{u}}_{g,n} = \mathbf{y}_{g,n} - \mathbf{Z}_{g,n} \widehat{\boldsymbol{\delta}}_{g,n}$ denote the GS2SLS residuals of the g -th equation, and let $\mathbf{m}_{g,n}^\rho(\bar{\boldsymbol{\rho}}_{g,n}, \widehat{\boldsymbol{\delta}}_{g,n})$ denote the corresponding sample moment vector as defined in (15). Then, the corresponding efficient GMM estimator for $\boldsymbol{\rho}_{g,n}$ based on GS2SLS residuals is given by

$$\widehat{\boldsymbol{\rho}}_{g,n} = \underset{\bar{\boldsymbol{\rho}}_g: \sum_{r \in \mathbf{I}_{g,\rho}} |\bar{\rho}_{g,r}| \in [-a_\rho, a_\rho]}{\operatorname{argmin}} \quad \mathbf{m}_{g,n}^\rho(\bar{\boldsymbol{\rho}}_{g,n}, \widehat{\boldsymbol{\delta}}_{g,n})' (\widehat{\boldsymbol{\Psi}}_{gg,n}^{\rho\rho})^{-1} \mathbf{m}_{g,n}^\rho(\bar{\boldsymbol{\rho}}_{g,n}, \widehat{\boldsymbol{\delta}}_{g,n}), \quad (32)$$

where $\widehat{\boldsymbol{\Psi}}_{gg,n}^{\rho\rho} = (\widehat{\psi}_{rs,gg,n}^{\rho\rho})$ is an estimator of the VC matrix of the limiting distribution of the normalized sample moments $n^{1/2} \mathbf{m}_{g,n}^\rho(\boldsymbol{\rho}_{g,n}, \widehat{\boldsymbol{\delta}}_{g,n})$. Specifically, we have

$$\begin{aligned} \widehat{\psi}_{rs,gg,n}^{\rho\rho} &= (2n)^{-1} \widehat{\sigma}_{gg,n}^2 \operatorname{tr} [(\mathbf{A}_{r,n} + \mathbf{A}'_{r,n})(\mathbf{A}_{s,n} + \mathbf{A}'_{s,n})] \\ &\quad + \widehat{\boldsymbol{\alpha}}'_{g,r,n} \widehat{\boldsymbol{\Psi}}_{gg,n}^{\delta\delta} \widehat{\boldsymbol{\alpha}}_{g,s,n}, \end{aligned}$$

with

$$\widehat{\boldsymbol{\alpha}}_{g,r,n} = -n^{-1} [\mathbf{Z}'_{*g,n}(\tilde{\boldsymbol{\rho}}_{g,n})(\mathbf{A}_{r,n} + \mathbf{A}'_{r,n})(\mathbf{I}_n - \mathbf{R}_{g,n}^*(\tilde{\boldsymbol{\rho}}_{g,n})) \widehat{\mathbf{u}}_{g,n}].$$

The claim that $(\widehat{\boldsymbol{\Psi}}_{gg,n}^{\rho\rho})^{-1}$ provides the efficient weighting of the sample moments will be established by Theorem 5 below.

5.2 Asymptotic Properties of Limited Information Estimators

In this subsection, we derive results concerning the joint limiting distribution of the GS2SLS estimators $\widehat{\boldsymbol{\rho}}_{g,n}$ and $\widehat{\boldsymbol{\delta}}_{g,n}$ by applying the generic limit theory developed in Theorem 4. In preparation we first specialize the expressions for

estimators of the (g, g) -th blocks of $\Psi_n^{\delta\rho}$, $\Psi_n^{\rho\rho}$ and $\Omega_n^{\delta\delta}$, $\Omega_n^{\delta\rho}$, $\Omega_n^{\rho\rho}$ as implied by the specific structure of $\widehat{\boldsymbol{\delta}}_{g,n}$ and $\widehat{\boldsymbol{\rho}}_{g,n}$. More specifically, let

$$\widehat{\Omega}_{gg,n}^{\delta\delta} = \widehat{\Psi}_{gg,n}^{\delta\delta}, \quad \widehat{\Omega}_{gg,n}^{\delta\rho} = \widehat{\Psi}_{gg,n}^{\delta\rho} \widehat{\mathbf{J}}_{g,n}, \quad \widehat{\Omega}_{gg,n}^{\rho\rho} = \left[\widetilde{\mathbf{J}}'_{g,n} \left(\widehat{\Psi}_{gg,n}^{\rho\rho} \right)^{-1} \widetilde{\mathbf{J}}_{g,n} \right]^{-1},$$

with

$$\begin{aligned} \widehat{\Psi}_{gg,n}^{\delta\rho} &= \widehat{\Psi}_{gg,n}^{\delta\delta} [\widehat{\boldsymbol{\alpha}}_{g,1,n}, \dots, \widehat{\boldsymbol{\alpha}}_{g,S,n}], \\ \widehat{\mathbf{J}}_{g,n} &= \left(\widehat{\Psi}_{gg,n}^{\rho\rho} \right)^{-1} \widetilde{\mathbf{J}}_{g,n} \left[\widetilde{\mathbf{J}}'_{g,n} \left(\widehat{\Psi}_{gg,n}^{\rho\rho} \right)^{-1} \widetilde{\mathbf{J}}_{g,n} \right]^{-1}, \end{aligned}$$

and $\widetilde{\mathbf{J}}_{g,n} = \mathbf{J}_{g,n}(\widetilde{\boldsymbol{\rho}}_{g,n})$. We now have the following result concerning the joint asymptotic distribution of $\widehat{\boldsymbol{\delta}}_{g,n}$ and $\widehat{\boldsymbol{\rho}}_{g,n}$.

Theorem 5 (*Joint Asymptotic Normality of $\widehat{\boldsymbol{\rho}}_{g,n}$ and $\widehat{\boldsymbol{\delta}}_{g,n}$*) *Suppose Assumptions 1-8 hold, and that the smallest eigenvalues of Ψ_n are bounded away from zero.²¹ Then, $\widehat{\boldsymbol{\rho}}_{g,n}$ is efficient among the class of GMM estimators based on GS2SLS residuals, and*

$$\begin{bmatrix} \widehat{\boldsymbol{\delta}}_{g,n} - \boldsymbol{\delta}_{g,n} \\ \widehat{\boldsymbol{\rho}}_{g,n} - \boldsymbol{\rho}_{g,n} \end{bmatrix} \sim AN \left[n^{-1} \begin{pmatrix} \widehat{\Omega}_{gg,n}^{\delta\delta} & \widehat{\Omega}_{gg,n}^{\delta\rho} \\ \widehat{\Omega}_{gg,n}^{\delta\rho'} & \widehat{\Omega}_{gg,n}^{\rho\rho} \end{pmatrix} \right].$$

In the above theorem the estimators for the asymptotic VC matrix of $\widehat{\boldsymbol{\rho}}_{g,n}$ and $\widehat{\boldsymbol{\delta}}_{g,n}$ employ $\widetilde{\boldsymbol{\rho}}_{g,n}$ as an estimator for $\boldsymbol{\rho}_{g,n}$. The theorem continues to hold if $\widetilde{\boldsymbol{\rho}}_{g,n}$ is replaced by $\widehat{\boldsymbol{\rho}}_{g,n}$ (or any other consistent estimator).

5.3 Definition of Full Information Estimators

In the previous section we discussed GS2SLS estimation, where the parameters of each equation are estimated separately from the spatially Cochrane-Orcutt transformed model (10). In the following, we consider full information estimation, where all parameters are estimated jointly from the stacked spatially Cochrane-Orcutt transformed model (11). In particular, we will consider a generalized spatial three-stage least squares (GS3SLS) estimator.

Step 3a: GS3SLS estimator of $\boldsymbol{\delta}_n$

²¹Explicit expressions for the sub-matrices composing Ψ_n specialized to $\widehat{\boldsymbol{\rho}}_{g,n}$ and $\widehat{\boldsymbol{\delta}}_{g,n}$ are given in the proof of the theorem.

The GS3SLS estimator of $\boldsymbol{\delta}_n$ based on the stacked spatially Cochrane-Orcutt transformed model (11) is given by

$$\widehat{\boldsymbol{\delta}}_n = [\widehat{\mathbf{Z}}'_{*n}(\widehat{\boldsymbol{\rho}}_n)(\widehat{\boldsymbol{\Sigma}}_n^{-1} \otimes \mathbf{I}_n)\mathbf{Z}_{*n}(\widehat{\boldsymbol{\rho}}_n)]^{-1}\widehat{\mathbf{Z}}'_{*n}(\widehat{\boldsymbol{\rho}}_n)(\widehat{\boldsymbol{\Sigma}}_n^{-1} \otimes \mathbf{I}_n)^{-1}\mathbf{y}_{*n}(\widehat{\boldsymbol{\rho}}_n), \quad (33)$$

where $\mathbf{y}_{*n}(\widehat{\boldsymbol{\rho}}_n) = [\mathbf{y}'_{*1,n}(\widehat{\boldsymbol{\rho}}_{1,n}), \dots, \mathbf{y}'_{*G,n}(\widehat{\boldsymbol{\rho}}_{G,n})]'$, $\mathbf{Z}_{*n}(\widehat{\boldsymbol{\rho}}_n) = \text{diag}_{g=1}^G [\mathbf{Z}_{*g,n}(\widehat{\boldsymbol{\rho}}_{g,n})]$, $\widehat{\mathbf{Z}}_{*n}(\widehat{\boldsymbol{\rho}}_n) = \text{diag}_{g=1}^G [\widehat{\mathbf{Z}}_{*g,n}(\widehat{\boldsymbol{\rho}}_{g,n})]$ with $\widehat{\mathbf{Z}}_{*g,n}(\widehat{\boldsymbol{\rho}}_{g,n}) = \mathbf{P}_{\mathbf{H}_n} \mathbf{Z}_{*g,n}(\widehat{\boldsymbol{\rho}}_{g,n})$, and where $\widehat{\boldsymbol{\Sigma}}_n = (\widehat{\sigma}_{gh,n})$ with $\widehat{\sigma}_{gh,n} = n^{-1}\widehat{\boldsymbol{\varepsilon}}'_{g,n}\widehat{\boldsymbol{\varepsilon}}_{h,n}$ and $\widehat{\boldsymbol{\varepsilon}}_{g,n} = \mathbf{y}_{*g,n}(\widehat{\boldsymbol{\rho}}_{g,n}) - \mathbf{Z}_{*g,n}(\widehat{\boldsymbol{\rho}}_{g,n})\widehat{\boldsymbol{\delta}}_{g,n}$. Below, we shall also utilize the following estimator for $\boldsymbol{\Psi}_n^{\delta\delta}$ (corresponding to $\widehat{\boldsymbol{\delta}}_n$):

$$\widehat{\boldsymbol{\Psi}}_n^{\delta\delta} = \left[n^{-1}\widehat{\mathbf{Z}}'_{*n}(\widehat{\boldsymbol{\rho}}_n)(\widehat{\boldsymbol{\Sigma}}_n^{-1} \otimes \mathbf{I}_n)\widehat{\mathbf{Z}}_{*n}(\widehat{\boldsymbol{\rho}}_n) \right]^{-1}$$

and we denote the (g, h) -th blocks of $\boldsymbol{\Psi}_n^{\delta\delta}$ and $\widehat{\boldsymbol{\Psi}}_n^{\delta\delta}$ with $\boldsymbol{\Psi}_{gh,n}^{\delta\delta}$ and $\widehat{\boldsymbol{\Psi}}_{gh,n}^{\delta\delta}$, respectively.

Step 3b: GMM estimator of $\boldsymbol{\rho}_{g,n}$ based on GS3SLS residuals

In a final step we compute a further GMM estimator of $\boldsymbol{\rho}_{g,n}$ based on the GS3SLS residuals $\widehat{\mathbf{u}}_{g,n} = \mathbf{y}_{g,n} - \mathbf{Z}_{g,n}\widehat{\boldsymbol{\delta}}_{g,n}$, where $\widehat{\boldsymbol{\delta}}_{g,n}$ denotes the g -th component of $\widehat{\boldsymbol{\delta}}_n$. Let $\mathbf{m}_{g,n}^\rho(\bar{\boldsymbol{\rho}}_{g,n}, \widehat{\boldsymbol{\delta}}_{g,n})$ denote the corresponding sample moment vector as defined in (15). Then, the corresponding GMM estimator for $\boldsymbol{\rho}_{g,n}$ based on GS3SLS residuals is given by

$$\widehat{\boldsymbol{\rho}}_{g,n} = \underset{\bar{\boldsymbol{\rho}}_g: \sum_{r \in \mathbf{I}_{g,\rho}} |\bar{\rho}_{g,r}| \in [-a_\rho, a_\rho]}{\text{argmin}} \mathbf{m}_{g,n}^\rho(\bar{\boldsymbol{\rho}}_{g,n}, \widehat{\boldsymbol{\delta}}_{g,n})' \left(\widehat{\boldsymbol{\Psi}}_{gg,n}^{\rho\rho} \right)^{-1} \mathbf{m}_{g,n}^\rho(\bar{\boldsymbol{\rho}}_{g,n}, \widehat{\boldsymbol{\delta}}_{g,n}), \quad (34)$$

where $\widehat{\boldsymbol{\Psi}}_{gg,n}^{\rho\rho}$ is an estimator of the VC matrix $\boldsymbol{\Psi}_{gg,n}^{\rho\rho}$ of the limiting distribution of the normalized sample moments $n^{1/2}\mathbf{m}_{g,n}^\rho(\boldsymbol{\rho}_{g,n}, \widehat{\boldsymbol{\delta}}_{g,n})$. Towards presenting the asymptotic distribution of $\widehat{\boldsymbol{\rho}}_{1,n}, \dots, \widehat{\boldsymbol{\rho}}_{G,n}$ we need estimators not only for the (g, g) -th block of $\boldsymbol{\Psi}_n^{\rho\rho}$, but more generally for the (g, h) -th block $\boldsymbol{\Psi}_{gh,n}^{\rho\rho}$. Let $\widehat{\boldsymbol{\Psi}}_n^{\rho\rho}$ and $\widehat{\boldsymbol{\Psi}}_{gh,n}^{\rho\rho}$ denote the estimators for $\boldsymbol{\Psi}_n^{\rho\rho}$ and $\boldsymbol{\Psi}_{gh,n}^{\rho\rho}$, respectively, then the (r, s) -th element of $\widehat{\boldsymbol{\Psi}}_{gh,n}^{\rho\rho}$ is defined as

$$\begin{aligned} \widehat{\psi}_{rs,gh,n}^{\rho\rho} &= (2n)^{-1}\widehat{\sigma}_{gh,n}^2 \text{tr} [(\mathbf{A}_{r,n} + \mathbf{A}'_{r,n})(\mathbf{A}_{s,n} + \mathbf{A}'_{s,n})] \\ &\quad + \widehat{\boldsymbol{\alpha}}'_{g,r,n} \widehat{\boldsymbol{\Psi}}_{gh,n}^{\delta\delta} \widehat{\boldsymbol{\alpha}}_{h,s,n}, \end{aligned} \quad (35)$$

with $\widehat{\boldsymbol{\alpha}}_{g,r,n} = -n^{-1} \left[\mathbf{Z}'_{g,n} (\mathbf{I}_n - \mathbf{R}_{g,n}^* (\widehat{\boldsymbol{\rho}}_{g,n})) (\mathbf{A}_{r,n} + \mathbf{A}'_{r,n}) (\mathbf{I}_n - \mathbf{R}_{g,n}^* (\widehat{\boldsymbol{\rho}}_{g,n})) \widehat{\mathbf{u}}_{g,n} \right]$.

5.4 Asymptotic Properties of Full Information Estimators

In this subsection, we derive results concerning the joint limiting distribution of the GS3SLS estimators $\widehat{\boldsymbol{\rho}}_n$ and $\widehat{\boldsymbol{\delta}}_n$ by applying again the generic limit theory developed in Theorem 4. In preparation, we first specialize the expressions for estimators of $\boldsymbol{\Psi}_n^{\delta\delta}$, $\boldsymbol{\Psi}_n^{\rho\rho}$ and $\boldsymbol{\Omega}_n^{\delta\delta}$, $\boldsymbol{\Omega}_n^{\delta\rho}$, $\boldsymbol{\Omega}_n^{\rho\rho}$ as implied by the specific structure of $\widehat{\boldsymbol{\rho}}_n$ and $\widehat{\boldsymbol{\delta}}_n$. More specifically, let

$$\widehat{\boldsymbol{\Omega}}_n^{\delta\delta} = \widehat{\boldsymbol{\Psi}}_n^{\delta\delta}, \quad \widehat{\boldsymbol{\Omega}}_n^{\delta\rho} = \widehat{\boldsymbol{\Psi}}_n^{\delta\rho} \text{diag}_{g=1}^G \left(\widehat{\mathbf{J}}_{g,n} \right), \quad \widehat{\boldsymbol{\Omega}}_n^{\rho\rho} = \text{diag}_{g=1}^G \left(\widehat{\mathbf{J}}'_{g,n} \right) \widehat{\boldsymbol{\Psi}}_n^{\rho\rho} \text{diag}_{g=1}^G \left(\widehat{\mathbf{J}}_{g,n} \right),$$

with

$$\begin{aligned} \widehat{\boldsymbol{\Psi}}_n^{\delta\rho} &= \widehat{\boldsymbol{\Psi}}_n^{\delta\delta} \text{diag}_{g=1}^G \left([\widehat{\boldsymbol{\alpha}}_{g,1,n}, \dots, \widehat{\boldsymbol{\alpha}}_{g,S,n}] \right), \\ \widehat{\mathbf{J}}_{g,n} &= \left(\widehat{\boldsymbol{\Psi}}_{gg,n}^{\rho\rho} \right)^{-1} \widehat{\mathbf{J}}_{g,n} \left[\widehat{\mathbf{J}}'_{g,n} \left(\widehat{\boldsymbol{\Psi}}_{gg,n}^{\rho\rho} \right)^{-1} \widehat{\mathbf{J}}_{g,n} \right]^{-1}, \end{aligned}$$

and with $\widehat{\mathbf{J}}_{g,n} = \mathbf{J}_{g,n}(\widehat{\boldsymbol{\rho}}_{g,n})$. The next theorem establishes the joint limiting distribution of $\widehat{\boldsymbol{\rho}}_n$ and $\widehat{\boldsymbol{\delta}}_n$.

Theorem 6 (*Joint Asymptotic Normality of $\widehat{\boldsymbol{\rho}}_n$ and $\widehat{\boldsymbol{\delta}}_n$*) *Suppose Assumptions 1-8 hold, and that the smallest eigenvalues of $\boldsymbol{\Psi}_n$ are bounded away from zero.²² Then,*

$$\begin{bmatrix} \widehat{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_n \\ \widehat{\boldsymbol{\rho}}_n - \boldsymbol{\rho}_n \end{bmatrix} \sim AN \left[n^{-1} \begin{pmatrix} \widehat{\boldsymbol{\Omega}}_n^{\delta\delta} & \widehat{\boldsymbol{\Omega}}_n^{\delta\rho} \\ \widehat{\boldsymbol{\Omega}}_n^{\delta\rho'} & \widehat{\boldsymbol{\Omega}}_n^{\rho\rho} \end{pmatrix} \right].$$

The estimators $\widehat{\boldsymbol{\rho}}_{g,n}$ based on GS3SLS residuals are efficient within their class.

In the above theorem the estimators for the asymptotic VC matrix of $\widehat{\boldsymbol{\rho}}_n$ and $\widehat{\boldsymbol{\delta}}_n$ employ $\widehat{\boldsymbol{\rho}}_{g,n}$ as an estimator for $\boldsymbol{\rho}_{g,n}$. The theorem continues to hold if $\widehat{\boldsymbol{\rho}}_{g,n}$ is replaced by $\widehat{\boldsymbol{\rho}}_{g,n}$ (or any other consistent estimator).

²²Explicit expressions for the sub-matrices $\boldsymbol{\Psi}_n$ are given in the proof of the theorem.

6 Limited and Full Information One-Step Estimators

In the following we discuss the one-step analogues to the two-step estimators considered in the previous section. For simplicity we assume the availability of a consistent estimator for Σ , say $\tilde{\Sigma}_n = (\tilde{\sigma}_{gh,n})$.

Towards defining our one-step limited information GMM estimator, consider the stacked moment vector

$$\mathbf{m}_{g,n}(\boldsymbol{\rho}_{g,n}, \boldsymbol{\delta}_{g,n}) = \begin{bmatrix} \mathbf{m}_{g,n}^L(\boldsymbol{\rho}_{g,n}, \boldsymbol{\delta}_{g,n}) \\ \mathbf{m}_{g,n}^Q(\boldsymbol{\rho}_{g,n}, \boldsymbol{\delta}_{g,n}) \end{bmatrix},$$

where $\mathbf{m}_{g,n}^L(\boldsymbol{\rho}_{g,n}, \boldsymbol{\delta}_{g,n}) = \mathbf{m}_{g,n}^\delta(\boldsymbol{\rho}_{g,n}, \boldsymbol{\delta}_{g,n})$ and $\mathbf{m}_{g,n}^Q(\boldsymbol{\rho}_{g,n}, \boldsymbol{\delta}_{g,n}) = \mathbf{m}_{g,n}^\rho(\boldsymbol{\rho}_{g,n}, \boldsymbol{\delta}_{g,n})$ are the vectors of linear and quadratic sample moments for the g -th equation as defined in (12) and (13).²³ Let

$$\boldsymbol{\Psi}_{gg,n}^{LL} = \sigma_{gg,n} [n^{-1} \mathbf{H}'_n \mathbf{H}_n] \text{ and } \boldsymbol{\Psi}_{gg,n}^{QQ} = \sigma_{gg,n}^2 \mathbf{K}_n^{QQ},$$

with $\mathbf{K}_n^{QQ} = (k_{rs,n}^{QQ})$ and

$$k_{rs,n}^{QQ} = (2n)^{-1} \text{tr} [(\mathbf{A}_{r,n} + \mathbf{A}'_{r,n})(\mathbf{A}_{s,n} + \mathbf{A}'_{s,n})].$$

Then, $E\mathbf{m}_{g,n}(\boldsymbol{\rho}_{g,n}, \boldsymbol{\delta}_{g,n}) = \mathbf{0}$ and

$$VC(n^{1/2} \mathbf{m}_{g,n}(\boldsymbol{\rho}_{g,n}, \boldsymbol{\delta}_{g,n})) = \begin{bmatrix} \boldsymbol{\Psi}_{gg,n}^{LL} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi}_{gg,n}^{QQ} \end{bmatrix}.$$

Let $\tilde{\boldsymbol{\Psi}}_{gg,n}^{LL}$ and $\tilde{\boldsymbol{\Psi}}_{gg,n}^{QQ}$ denote the corresponding estimators, where $\sigma_{gg,n}$ is replaced by some consistent estimator $\tilde{\sigma}_{gg,n}$. Then the one-step limited information GMM estimator is defined as

$$(\hat{\boldsymbol{\delta}}_{g,n}^o, \hat{\boldsymbol{\rho}}_{g,n}^o) = \arg \min_{\bar{\boldsymbol{\rho}}_{g,n}, \bar{\boldsymbol{\delta}}_{g,n}} \mathbf{m}_{g,n}(\bar{\boldsymbol{\rho}}_{g,n}, \bar{\boldsymbol{\delta}}_{g,n})' \begin{bmatrix} \tilde{\boldsymbol{\Psi}}_{gg,n}^{LL} & \mathbf{0} \\ \mathbf{0} & \tilde{\boldsymbol{\Psi}}_{gg,n}^{QQ} \end{bmatrix}^{-1} \mathbf{m}_{g,n}(\bar{\boldsymbol{\rho}}_{g,n}, \bar{\boldsymbol{\delta}}_{g,n}). \quad (36)$$

For ease of distinction from the two-step limited information estimators defined in the previous section we refer to this estimator as the Linear-Quadratic

²³The purpose of switching the superscripts from δ and ρ to L for ‘‘linear’’ and Q for ‘‘quadratic’’ is to emphasize that for one-step estimators both moment vectors are used for estimating both $\boldsymbol{\rho}_{g,n}$ and $\boldsymbol{\delta}_{g,n}$.

Generalizes Spatial 2SLS (LQ-GS2SLS) estimator. A simple adaptation of the methodology used to derive the limiting distribution of two-step estimators yields

$$\begin{bmatrix} (\widehat{\boldsymbol{\delta}}_{g,n}^o - \boldsymbol{\delta}_{g,n}) \\ (\widehat{\boldsymbol{\rho}}_{g,n}^o - \boldsymbol{\rho}_{g,n}) \end{bmatrix} \sim AN \left[0, n^{-1} \widehat{\boldsymbol{\Omega}}_{g,n}^o \right].$$

with $\widehat{\boldsymbol{\Omega}}_{g,n}^o = (\widehat{\mathbf{S}}_{g,n}^o)^{-1}$ and

$$\begin{aligned} \widehat{\mathbf{S}}_{g,n}^o &= \begin{bmatrix} \widehat{\mathbf{S}}_{g,n}^{o,LL} & \widehat{\mathbf{S}}_{g,n}^{o,LQ} \\ \widehat{\mathbf{S}}_{g,n}^{o,QL} & \widehat{\mathbf{S}}_{g,n}^{o,QQ} \end{bmatrix}, \\ \widehat{\mathbf{S}}_{g,n}^{o,LL} &= \widehat{\sigma}_{gg,n}^{o-1} [n^{-1} \widehat{\mathbf{Z}}_{*g,n}(\widehat{\boldsymbol{\rho}}_{g,n}^o)' \widehat{\mathbf{Z}}_{*g,n}(\widehat{\boldsymbol{\rho}}_{g,n}^o)] + \widehat{\sigma}_{gg,n}^{o-2} \widehat{\boldsymbol{\alpha}}_{g,n}^o (\mathbf{K}_n^{QQ})^{-1} \widehat{\boldsymbol{\alpha}}_{g,n}^{o'}, \\ \widehat{\mathbf{S}}_{g,n}^{o,LQ} &= \widehat{\mathbf{S}}_{g,n}^{o,QL'} = -\widehat{\sigma}_{gg,n}^{o-2} \widehat{\boldsymbol{\alpha}}_{g,n}^o (\mathbf{K}_n^{QQ})^{-1} \mathbf{J}_{g,n}(\widehat{\boldsymbol{\rho}}_{g,n}^o), \\ \widehat{\mathbf{S}}_{g,n}^{o,QQ} &= \widehat{\sigma}_{gg,n}^{o-2} \mathbf{J}_{g,n}'(\widehat{\boldsymbol{\rho}}_{g,n}^o) (\mathbf{K}_n^{QQ})^{-1} \mathbf{J}_{g,n}(\widehat{\boldsymbol{\rho}}_{g,n}^o), \end{aligned}$$

where $\widehat{\sigma}_{gg,n}^o = n^{-1} \widehat{\boldsymbol{\varepsilon}}_{g,n}^{o'} \widehat{\boldsymbol{\varepsilon}}_{g,n}^o$ with $\widehat{\boldsymbol{\varepsilon}}_{g,n}^o = \mathbf{y}_{*g,n}(\widehat{\boldsymbol{\rho}}_{g,n}^o) - \mathbf{Z}_{*g,n}(\widehat{\boldsymbol{\rho}}_{g,n}^o) \widehat{\boldsymbol{\delta}}_{g,n}^o$, and where $\widehat{\boldsymbol{\alpha}}_{g,n}^o = [\widehat{\boldsymbol{\alpha}}_{g,1,n}^o, \dots, \widehat{\boldsymbol{\alpha}}_{g,S,n}^o]$ with $\widehat{\boldsymbol{\alpha}}_{g,r,n}^o = -n^{-1} [\mathbf{Z}'_{*g,n}(\widehat{\boldsymbol{\rho}}_{g,n}^o) (\mathbf{A}_{r,n} + \mathbf{A}'_{r,n}) (\mathbf{I}_n - \mathbf{R}_{g,n}^*(\widehat{\boldsymbol{\rho}}_{g,n}^o))] \widehat{\mathbf{u}}_{g,n}^o$ and $\widehat{\mathbf{u}}_{g,n}^o = \mathbf{y}_{g,n} - \mathbf{Z}_{g,n} \widehat{\boldsymbol{\delta}}_{g,n}^o$.

The one-step estimator defined in (36) is efficient among GMM estimators based on the moment conditions $E \mathbf{m}_{g,n}(\boldsymbol{\rho}_{g,n}, \boldsymbol{\delta}_{g,n}) = \mathbf{0}$. However, a comparison of the above expressions for the asymptotic VC matrix of the one-step estimator with those for the two-step estimator given by Theorem 5 reveals that in the case where $\mathbf{Z}_{g,n}$ only contains exogenous variables, and thus $\boldsymbol{\alpha}_{g,n} = 0$, both estimators have the same limiting distribution.

Our setup contains as an important special case models without spatial lags in the disturbances (i.e., where $\boldsymbol{\rho}_{g,n} = 0$ is known). Obviously the LQ-GS2SLS estimator, which combines both linear and quadratic moment conditions, remains well defined in this case, and we obtain $(\widehat{\boldsymbol{\delta}}_{g,n}^o - \boldsymbol{\delta}_{g,n}) \sim AN \left[0, n^{-1} (\widehat{\mathbf{S}}_{g,n}^{o,LL})^{-1} \right]$ as a special case. The use of quadratic moment conditions may be especially beneficial in cases where identification by the linear moments is weak; see, e.g., Kuersteiner and Prucha (2020) for a recent discussion.

Towards defining our one-step full information GMM estimator consider the

stacked sample moment vector

$$\mathbf{m}_n(\boldsymbol{\rho}_n, \boldsymbol{\delta}_n) = \begin{bmatrix} \mathbf{m}_{1,n}^L(\boldsymbol{\rho}_{1,n}, \boldsymbol{\delta}_{1,n}) \\ \vdots \\ \mathbf{m}_{G,n}^L(\boldsymbol{\rho}_{G,n}, \boldsymbol{\delta}_{G,n}) \\ \mathbf{m}_{1,n}^Q(\boldsymbol{\rho}_{1,n}, \boldsymbol{\delta}_{1,n}) \\ \vdots \\ \mathbf{m}_{G,n}^Q(\boldsymbol{\rho}_{G,n}, \boldsymbol{\delta}_{G,n}) \end{bmatrix}.$$

Let

$$\begin{aligned} \boldsymbol{\Psi}_n^{LL} &= \boldsymbol{\Sigma}_n \otimes [n^{-1} \mathbf{H}'_n \mathbf{H}_n], \\ \boldsymbol{\Psi}_n^{QQ} &= \begin{bmatrix} \boldsymbol{\Psi}_{11,n}^{QQ} & \cdots & \boldsymbol{\Psi}_{1G,n}^{QQ} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Psi}_{G1,n}^{QQ} & \cdots & \boldsymbol{\Psi}_{GG,n}^{QQ} \end{bmatrix} = \boldsymbol{\Sigma}_{SQ,n} \otimes \mathbf{K}_n^{QQ}, \end{aligned}$$

with $\boldsymbol{\Psi}_{gh,n}^{QQ} = \sigma_{gh,n}^2 \mathbf{K}_n^{QQ}$ and $\boldsymbol{\Sigma}_{SQ,n} = (\sigma_{gh,n}^2)$. Then, $E\mathbf{m}(\boldsymbol{\rho}_n, \boldsymbol{\delta}_n) = \mathbf{0}$ and

$$VC(\mathbf{m}_n(\boldsymbol{\rho}_n, \boldsymbol{\delta}_n)) = n^{-1} \begin{bmatrix} \boldsymbol{\Psi}_n^{LL} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi}_n^{QQ} \end{bmatrix}.$$

Let $\widehat{\boldsymbol{\Psi}}_n^{LL}$ and $\widehat{\boldsymbol{\Psi}}_n^{QQ}$ denote the corresponding estimators where $\sigma_{gh,n}$ is replaced by some consistent estimator $\tilde{\sigma}_{gh,n}$. Then the one-step full information GMM estimator is defined as

$$(\widehat{\boldsymbol{\delta}}_n^o, \widehat{\boldsymbol{\rho}}_n^o) = \arg \min_{\bar{\boldsymbol{\rho}}_n, \bar{\boldsymbol{\delta}}_n} \mathbf{m}_n(\bar{\boldsymbol{\rho}}_n, \bar{\boldsymbol{\delta}}_n)' \begin{bmatrix} \widehat{\boldsymbol{\Psi}}_n^{LL} & \mathbf{0} \\ \mathbf{0} & \widehat{\boldsymbol{\Psi}}_n^{QQ} \end{bmatrix}^{-1} \mathbf{m}_n(\bar{\boldsymbol{\rho}}_n, \bar{\boldsymbol{\delta}}_n).$$

For ease of distinction from the two-step full information estimators defined in the previous section we refer to this estimator as the Linear-Quadratic Generalizes Spatial 3SLS (LQ-GS3SLS) estimator. Again, a simple adaptation of the methodology used to derive the limiting distribution of two-step estimators yields

$$\begin{bmatrix} \widehat{\boldsymbol{\delta}}_n^o - \boldsymbol{\delta}_n \\ \widehat{\boldsymbol{\rho}}_n^o - \boldsymbol{\rho}_n \end{bmatrix} \sim AN \left[n^{-1} \widehat{\boldsymbol{\Omega}}_n^o \right],$$

with $\widehat{\boldsymbol{\Omega}}_n^o = (\widehat{\mathbf{S}}_n^o)^{-1}$ and

$$\begin{aligned}\widehat{\mathbf{S}}_n^o &= \begin{bmatrix} \widehat{\mathbf{S}}_n^{LL} & \widehat{\mathbf{S}}_n^{LQ} \\ \widehat{\mathbf{S}}_n^{QL} & \widehat{\mathbf{S}}_n^{QQ} \end{bmatrix}, \\ \widehat{\mathbf{S}}_n^{o,LL} &= n^{-1} \widehat{\mathbf{Z}}_{*n}'(\widehat{\boldsymbol{\rho}}_n^o) [(\widehat{\boldsymbol{\Sigma}}_n^o)^{-1} \otimes \mathbf{I}_n] \widehat{\mathbf{Z}}_{*n}(\widehat{\boldsymbol{\rho}}_n^o) \\ &\quad + \text{diag}_{g=1}^G [\widehat{\boldsymbol{\alpha}}_{g,n}^o] [(\widehat{\boldsymbol{\Sigma}}_{SQ,n}^o)^{-1} \otimes (\mathbf{K}_n^{QQ})^{-1}] \text{diag}_{g=1}^G [\widehat{\boldsymbol{\alpha}}_{g,n}^{o'}], \\ \widehat{\mathbf{S}}_{g,n}^{o,LQ} &= \widehat{\mathbf{S}}_{g,n}^{o,LQ'} = -\text{diag}_{g=1}^G [\widehat{\boldsymbol{\alpha}}_{g,n}^o] [(\widehat{\boldsymbol{\Sigma}}_{SQ,n}^o)^{-1} \otimes (\mathbf{K}_n^{QQ})^{-1}] \text{diag}_{g=1}^G [\mathbf{J}_{g,n}(\widehat{\boldsymbol{\rho}}_n^o)], \\ \widehat{\mathbf{S}}_{g,n}^{o,QQ} &= \text{diag}_{g=1}^G [\mathbf{J}'_{g,n}(\widehat{\boldsymbol{\rho}}_n^o)] [(\widehat{\boldsymbol{\Sigma}}_{SQ,n}^o)^{-1} \otimes (\mathbf{K}_n^{QQ})^{-1}] \text{diag}_{g=1}^G [\mathbf{J}_{g,n}(\widehat{\boldsymbol{\rho}}_n^o)],\end{aligned}$$

where $\widehat{\boldsymbol{\Sigma}}_n^o = (\widehat{\sigma}_{gh,n}^o)$ and $\widehat{\boldsymbol{\Sigma}}_{SQ,n}^o = (\widehat{\sigma}_{gh,n}^{o2})$ with $\widehat{\sigma}_{gh,n}^o = n^{-1} \widehat{\boldsymbol{\varepsilon}}_{g,n}^{o'} \widehat{\boldsymbol{\varepsilon}}_{h,n}^o$ and $\widehat{\boldsymbol{\varepsilon}}_{g,n}^o = \mathbf{y}_{*g,n}(\widehat{\boldsymbol{\rho}}_{g,n}^o) - \mathbf{Z}_{*g,n}(\widehat{\boldsymbol{\rho}}_{g,n}^o) \widehat{\boldsymbol{\delta}}_{g,n}^o$, and where $\widehat{\boldsymbol{\alpha}}_{g,n}^o = [\widehat{\alpha}_{g,1,n}^o, \dots, \widehat{\alpha}_{g,S,n}^o]$ with

$$\widehat{\alpha}_{g,r,n}^o = -n^{-1} \left[\mathbf{Z}'_{*g,n}(\widehat{\boldsymbol{\rho}}_{g,n}^o) (\mathbf{A}_{r,n} + \mathbf{A}'_{r,n}) (\mathbf{I}_n - \mathbf{R}_{g,n}^*(\widehat{\boldsymbol{\rho}}_{g,n}^o)) \widehat{\mathbf{u}}_{g,n} \right]$$

and $\widehat{\mathbf{u}}_{g,n}^o = \mathbf{y}_{g,n} - \mathbf{Z}_{g,n} \widehat{\boldsymbol{\delta}}_n^o$.

A comparison of the above expression for the asymptotic VC matrix of the one-step full information estimator with that for the two-step estimator given by Theorem 6 reveals that in the case where the $\mathbf{Z}_{g,n}$ only contain exogenous variables, and thus $\alpha_{g,n} = 0$, both estimators have the same limiting distribution.

7 Concluding Remarks

This paper develops estimation methodologies for a cross-sectional simultaneous equation model in G variables, where simultaneity stems from interdependencies in the G variables as well as from network interdependencies. Taking guidance from the spatial literature network interdependencies are modeled in the form of weighted averages. For simplicity, and consistent with the spatial literature, we refer to those weighted averages as spatial lags. We allow for higher order spatial lags in the endogenous variables, exogenous variables and disturbances. As a consequence, the model provides for significant flexibility in modeling network effects.

The paper develops an estimation theory for both limited and full informa-

tion generalized method of moments estimators, which utilize both linear and quadratic moment conditions. We consider both two-step and one-step estimators. An important aim in specifying our estimators was that the estimators remain feasible even for very large data sets and general weights matrices.

To explore the small sample properties of our estimators we conducted a Monte Carlo study. The details of the design and the results of that study are given in Appendix G in the Online Supplementary Appendix. In general, the results are encouraging. We consider scenarios where the parameters are well identified by the linear moment conditions as well as scenarios where they are not. The study includes the maximum likelihood (ML) estimator for comparison. In the well identified case the biases of all considered estimators are fairly small. As expected, the ML estimator has the smallest root mean squared error (RMSE). However, in general, the ML estimator only dominates the GS3SLS estimator slightly, and the GS3SLS estimator dominates the GS2SLS estimator. The differences in RMSE are the most pronounced for the estimates of the autoregressive parameters in the disturbance process. For the scenarios where identification by linear moment conditions alone is weak, the GS2SLS and GS3SLS estimators can, as expected, be substantially biased. In these scenarios the LQ-GS2SLS and LQ-GS3SLS estimators can greatly outperform the GS2SLS and GS3SLS estimators. However, for the well identified scenarios the benefit of combining linear and quadratic moment conditions seems limited.

We expect the model will be helpful for empirical research in both macro and micro economic settings, as well as areas outside of economics. As illustrated in the paper, one potential application is for modeling social interactions in different activities; e.g., the level of different physical activities among groups of friends connected via an activity tracker such as Fitbit. Suggestions for future research include an extension of the methodology to panel data, as well as an extension that allows for measurement errors in the data.

A Appendix: Preliminary Results

In this appendix we collect some preliminary results. All proofs are relegated to the Online Supplementary Appendix.

A.1 Asymptotic Linearity of S2SLS, GS2SLS and GS3SLS

Assumption 10 postulates that the estimators of the regression parameters are asymptotically linear. In the following we show that the S2SLS, GS2SLS and GS3SLS estimators are indeed asymptotically linear.

Lemma A.1 : Let $\mathbf{A}_n = (a_{ij,n})$ be some $m_n \times m_n$ real matrix where the row sums of the absolute elements are bounded uniformly in n by some finite constant. Let $\boldsymbol{\mu}_n = (\mu_{1,n}, \dots, \mu_{m_n,n})$ and $\boldsymbol{\eta}_n = (\eta_{1,n}, \dots, \eta_{m_n,n})'$ be some $m_n \times 1$ random vectors with $\sup_n \max_{i=1}^{m_n} E |\mu_{i,n}|^p < \infty$ and $\sup_n \max_{i=1}^{m_n} E |\eta_{i,n}|^p < \infty$ for some $p > 1$, and let $\boldsymbol{\xi}_n = (\xi_{1,n}, \dots, \xi_{m_n,n})' = \boldsymbol{\mu}_n + \mathbf{A}_n \boldsymbol{\eta}_n$. Then $\sup_n \max_{i=1}^{m_n} E |\xi_{i,n}|^p < \infty$.

Lemma A.2 : Suppose Assumptions 1-4 hold. Let $\underline{\mathbf{Z}}_n = [\mathbf{Y}_n, \mathbf{X}_n, \overline{\mathbf{Y}}_n]$, then

$$E |\underline{z}_{ij,n}|^4 \leq C < \infty, \quad (\text{A.1})$$

where C does not depend on i, j and n .

Lemma A.3 : Suppose Assumptions 1-4 hold. Let $\underline{\mathbf{Z}}_n = [\mathbf{Y}_n, \mathbf{X}_n, \overline{\mathbf{Y}}_n]$ and let $\mathbf{A}_n = (a_{ij,n})$ be some $n \times n$ matrix, where the row and column sums of the absolute elements are bounded uniformly in n by some finite constant. Then $n^{-1} \mathbf{u}'_{h,n} \mathbf{A}_n \mathbf{u}_{g,n} = O_p(1)$, $n^{-1} \underline{\mathbf{Z}}_n \mathbf{A}_n \mathbf{u}_{g,n} = O_p(1)$ and $n^{-1} \underline{\mathbf{Z}}_n \mathbf{A}_n \underline{\mathbf{Z}}_n = O_p(1)$ and furthermore

$$n^{-1} \underline{\mathbf{Z}}_n \mathbf{A}_n \mathbf{u}_{g,n} - n^{-1} E \underline{\mathbf{Z}}_n \mathbf{A}_n \mathbf{u}_{g,n} = o_p(1).$$

Lemma A.4 : Suppose Assumptions 1-4, 5 and 6 hold. Consider the S2SLS estimator

$$\tilde{\boldsymbol{\delta}}_{g,n} = (\widehat{\mathbf{Z}}'_{g,n} \mathbf{Z}_{g,n})^{-1} \widehat{\mathbf{Z}}'_{g,n} \mathbf{y}_{g,n},$$

where $\widehat{\mathbf{Z}}_{g,n} = \mathbf{P}_{\mathbf{H}_n} \mathbf{Z}_{g,n}$ and $\mathbf{P}_{\mathbf{H}_n} = \mathbf{H}_n (\mathbf{H}'_n \mathbf{H}_n)^{-1} \mathbf{H}'_n$. Then,
(a) $n^{1/2}(\widetilde{\boldsymbol{\delta}}_{g,n} - \boldsymbol{\delta}_{g,n}) = n^{-1/2} \mathbf{T}'_{gg,n} \boldsymbol{\varepsilon}_{g,n} + o_p(1)$ with $\mathbf{T}_{gg,n} = \mathbf{F}_{gg,n} \mathbf{P}_{gg,n}$ and where

$$\begin{aligned} \mathbf{P}_{gg} &= \mathbf{Q}_{\mathbf{H}\mathbf{H}}^{-1} \mathbf{Q}_{\mathbf{H}\mathbf{Z},g} [\mathbf{Q}'_{\mathbf{H}\mathbf{Z},g} \mathbf{Q}_{\mathbf{H}\mathbf{H}}^{-1} \mathbf{Q}_{\mathbf{H}\mathbf{Z},g}]^{-1}, \\ \mathbf{F}_{gg,n} &= (\mathbf{I}_n - \mathbf{R}_{g,n}^{*'})^{-1} \mathbf{H}_n. \end{aligned}$$

(b) $n^{-1/2} \mathbf{T}'_{gg,n} \boldsymbol{\varepsilon}_{g,n} = O_p(1)$.
(c) \mathbf{P}_{gg} is a finite matrix and $\widetilde{\mathbf{P}}_{gg,n} - \mathbf{P}_{gg} = o_p(1)$ for

$$\begin{aligned} \widetilde{\mathbf{P}}_{gg,n} &= (n^{-1} \mathbf{H}'_n \mathbf{H}_n)^{-1} (n^{-1} \mathbf{H}'_n \mathbf{Z}_{g,n}) \times \\ &\quad [(n^{-1} \mathbf{Z}'_{g,n} \mathbf{H}_n) (n^{-1} \mathbf{H}'_n \mathbf{H}_n)^{-1} (n^{-1} \mathbf{H}'_n \mathbf{Z}_{g,n})]^{-1}. \end{aligned}$$

(d) $\lambda_{\min}(n^{-1} \mathbf{T}'_{gg,n} \mathbf{T}_{gg,n}) \geq c$ for some $c > 0$ for all large n .

Lemma A.5 : Suppose Assumptions 1-4, 5 and 6 hold. Consider the GS2SLS estimator

$$\widehat{\boldsymbol{\delta}}_{g,n} = [\widehat{\mathbf{Z}}_{*g,n}(\widehat{\boldsymbol{\rho}}_{g,n})' \mathbf{Z}_{*g,n}(\widehat{\boldsymbol{\rho}}_{g,n})]^{-1} \widehat{\mathbf{Z}}_{*g,n}(\widehat{\boldsymbol{\rho}}_{g,n})' \mathbf{y}_{*g,n}(\widehat{\boldsymbol{\rho}}_{g,n}),$$

where $\widehat{\mathbf{Z}}_{*g,n}(\widehat{\boldsymbol{\rho}}_{g,n}) = \mathbf{P}_{\mathbf{H}_n} \mathbf{Z}_{*g,n}(\widehat{\boldsymbol{\rho}}_{g,n})$, where $\widehat{\boldsymbol{\rho}}_{g,n}$ is any consistent estimator for $\boldsymbol{\rho}_{g,n}$. Then,

(a) $n^{1/2}(\widehat{\boldsymbol{\delta}}_{g,n} - \boldsymbol{\delta}_{g,n}) = n^{-1/2} \mathbf{T}^{*'}_{gg,n} \boldsymbol{\varepsilon}_{g,n} + o_p(1)$ with $\mathbf{T}^*_{gg,n} = \mathbf{F}^*_{gg,n} \mathbf{P}^*_{gg,n}$ and where

$$\begin{aligned} \mathbf{P}^*_{gg,n} &= \mathbf{Q}_{\mathbf{H}\mathbf{H}}^{-1} \mathbf{Q}_{\mathbf{H}\mathbf{Z},g^*}(\boldsymbol{\rho}_{g,n}) [\mathbf{Q}'_{\mathbf{H}\mathbf{Z},g^*}(\boldsymbol{\rho}_{g,n}) \mathbf{Q}_{\mathbf{H}\mathbf{H}}^{-1} \mathbf{Q}_{\mathbf{H}\mathbf{Z},g^*}(\boldsymbol{\rho}_{g,n})]^{-1} \\ \mathbf{F}^*_{gg,n} &= \mathbf{H}_n. \end{aligned}$$

(b) $n^{-1/2} \mathbf{T}^{*'}_{gg,n} \boldsymbol{\varepsilon}_{g,n} = O_p(1)$.
(c) $\mathbf{P}^*_{gg,n} = O_p(1)$ and $\widetilde{\mathbf{P}}^*_{gg,n} - \mathbf{P}^*_{gg,n} = o_p(1)$ for

$$\begin{aligned} \widetilde{\mathbf{P}}^*_{gg,n} &= (n^{-1} \mathbf{H}'_n \mathbf{H}_n)^{-1} (n^{-1} \mathbf{H}'_n \mathbf{Z}_{*g,n}(\widehat{\boldsymbol{\rho}}_{g,n})) \times \\ &\quad [(n^{-1} \mathbf{Z}'_{*g,n}(\widehat{\boldsymbol{\rho}}_{g,n}) \mathbf{H}_n) (n^{-1} \mathbf{H}'_n \mathbf{H}_n)^{-1} (n^{-1} \mathbf{H}'_n \mathbf{Z}_{*g,n}(\widehat{\boldsymbol{\rho}}_{g,n}))]^{-1}. \end{aligned}$$

(d) $\lambda_{\min}(n^{-1} \mathbf{T}^{*'}_{gg,n} \mathbf{T}^*_{gg,n}) \geq c$ for some $c > 0$ for large n .

Lemma A.6 : Suppose Assumptions 1-4, 5-6 hold. Consider the GS3SLS estimator

$$\widehat{\boldsymbol{\delta}}_n = \left[\widehat{\mathbf{Z}}'_{*n}(\widehat{\boldsymbol{\rho}}_n)(\widehat{\boldsymbol{\Sigma}}_n^{-1} \otimes \mathbf{I}_n) \mathbf{Z}_{*n}(\widehat{\boldsymbol{\rho}}_n) \right]^{-1} \widehat{\mathbf{Z}}'_{*n}(\widehat{\boldsymbol{\rho}}_n)(\widehat{\boldsymbol{\Sigma}}_n^{-1} \otimes \mathbf{I}_n) \mathbf{y}_{*n}(\widehat{\boldsymbol{\rho}}_n),$$

where $\widehat{\mathbf{Z}}_{*n}(\widehat{\boldsymbol{\rho}}_n) = \text{diag}_{g=1}^G \left[\widehat{\mathbf{Z}}'_{*g,n}(\widehat{\boldsymbol{\rho}}_{g,n}) \right]$ with $\widehat{\mathbf{Z}}_{*g,n}(\widehat{\boldsymbol{\rho}}_{g,n}) = \mathbf{P}_{\mathbf{H}_n} \mathbf{Z}_{*g,n}(\widehat{\boldsymbol{\rho}}_n)$, and $\widehat{\boldsymbol{\rho}}_n = [\widehat{\boldsymbol{\rho}}'_{1,n}, \dots, \widehat{\boldsymbol{\rho}}'_{G,n}]'$ is any consistent estimator for $\boldsymbol{\rho}_n$ and $\widehat{\boldsymbol{\Sigma}}_n$ is any consistent estimator for $\boldsymbol{\Sigma}$. Then

(a) $n^{1/2}[\widehat{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_n] = n^{-1/2} \mathbf{T}_n^{**'} \boldsymbol{\varepsilon}_n + o_p(1)$ with $\mathbf{T}_n^{**} = \mathbf{F}_n^{**} \mathbf{P}_n^{**}$, and where

$$\begin{aligned} \mathbf{P}_n^{**} &= \left[\boldsymbol{\Sigma}^{-1} \otimes \mathbf{Q}_{\mathbf{H}\mathbf{H}}^{-1} \right] \text{diag} \left[\mathbf{Q}_{\mathbf{H}\mathbf{Z},g^*}(\boldsymbol{\rho}_{g,n}) \right] \\ &\quad \times \left\{ \text{diag} \left[\mathbf{Q}'_{\mathbf{H}\mathbf{Z},g^*}(\boldsymbol{\rho}_{g,n}) \right] \left[\boldsymbol{\Sigma}^{-1} \otimes \mathbf{Q}_{\mathbf{H}\mathbf{H}}^{-1} \right] \text{diag} \left[\mathbf{Q}_{\mathbf{H}\mathbf{Z},g^*}(\boldsymbol{\rho}_{g,n}) \right] \right\}^{-1} \end{aligned}$$

and

$$\mathbf{F}_n^{**} = \mathbf{I}_G \otimes \mathbf{H}_n.$$

(b) $n^{-1/2} \mathbf{T}_n^{**'} \boldsymbol{\varepsilon}_n = O_p(1)$.

(c) $\mathbf{P}_n^{**} = O_p(1)$ and $\widetilde{\mathbf{P}}_n^{**} - \mathbf{P}_n^{**} = o_p(1)$ for

$$\begin{aligned} \widetilde{\mathbf{P}}_n^{**} &= \left[\widehat{\boldsymbol{\Sigma}}_n^{-1} \otimes (n^{-1} \mathbf{H}'_n \mathbf{H}_n)^{-1} \right] \text{diag} \left[n^{-1} \mathbf{H}'_n \mathbf{Z}_{*g,n}(\widehat{\boldsymbol{\rho}}_{g,n}) \right] \\ &\quad \times \left[n^{-1} \widehat{\mathbf{Z}}'_{*n}(\widehat{\boldsymbol{\rho}}_n)(\widehat{\boldsymbol{\Sigma}}_n^{-1} \otimes \mathbf{I}_n) \mathbf{Z}_{*n}(\widehat{\boldsymbol{\rho}}_n) \right]^{-1}. \end{aligned}$$

(d) $\lambda_{\min}(n^{-1} \mathbf{T}_n^{**'} \mathbf{T}_n^{**}) \geq c$ for some $c > 0$ for large n .

A.2 Auxiliary Results for Linear Quadratic Forms

In the following we establish some auxiliary results on the relationship between linear and quadratic forms based on some $n \times 1$ disturbance vector \mathbf{u}_n and corresponding forms based on a predictor $\widetilde{\mathbf{u}}_n$.

Assumption A.1 For $n \geq 1$ the $n \times 1$ disturbance vector \mathbf{u}_n is generated by

$$\begin{matrix} \mathfrak{R}_n \mathbf{u}_n = \boldsymbol{\varepsilon}_n, \\ n \times n n \times 1 & n \times 1 \end{matrix}$$

where \mathfrak{R}_n is a non-stochastic nonsingular $n \times n$ matrix, and the row and column sums of the absolute elements of \mathfrak{R}_n and \mathfrak{R}_n^{-1} are bounded uniformly by some finite constant, and the innovations $\boldsymbol{\varepsilon}_n = (\boldsymbol{\varepsilon}_{1,n}, \dots, \boldsymbol{\varepsilon}_{n,n})'$ have the following

properties: For each $n \geq 1$ the random variables $\boldsymbol{\epsilon}_{1,n}, \dots, \boldsymbol{\epsilon}_{n,n}$ are totally independent with $E\boldsymbol{\epsilon}_{i,n} = 0$, $E(\boldsymbol{\epsilon}_{i,n}^2) = \sigma^2 > 0$, and $\sup_{1 \leq i \leq n, n \geq 1} E|\boldsymbol{\epsilon}_{i,n}|^{4+v} < \infty$ for some $v > 0$.

Assumption A.2 : The predictor $\tilde{\mathbf{u}}_n$ for \mathbf{u}_n satisfies that

$$\tilde{\mathbf{u}}_n - \mathbf{u}_n = \mathfrak{D}_n \Delta_n,$$

where $\mathfrak{D}_n = (\mathfrak{d}_{ij,n})$ is an $n \times p_\Delta$ random matrix and Δ_n is a $p_\Delta \times 1$ random vector. Furthermore $\sup_n \sup_{1 \leq i \leq n, 1 \leq j \leq p_\Delta} E|\mathfrak{d}_{ij,n}|^{2+\delta} < \infty$ for some $\delta > 0$, and $n^{1/2} \|\Delta_n\| = O_p(1)$.

Assumption A.3 For any $n \times n$ real matrix \mathbf{A}_n^* , whose row and column sums are bounded uniformly in absolute value,

$$n^{-1} \mathfrak{D}'_n \mathbf{A}_n^* \mathbf{u}_n - n^{-1} E \mathfrak{D}'_n \mathbf{A}_n^* \mathbf{u}_n = o_p(1).$$

Remark: The above assumptions are formulated in a general fashion, so that the results on the properties of linear quadratic forms established below can also be utilized in a variety of contexts. For an interpretation of the results specific to this paper, suppose that \mathbf{u}_n corresponds to the disturbance term of the g -th equation of the model defined by (1)-(8). Then the quantities considered in Assumptions A.1-A.3 should be interpreted as $\mathbf{u}_n = \mathbf{u}_{g,n}$, $\mathfrak{D}_n = -\mathbf{Z}_{g,n}$, $\mathfrak{R}_n = \mathbf{I} - \sum_{r \in \mathbf{I}_{g,p}} \rho_{g,r,n} \mathbf{M}_{r,n}$, $\boldsymbol{\epsilon}_n = \boldsymbol{\epsilon}_{g,n}$, and $\Delta_n = \tilde{\boldsymbol{\delta}}_{g,n} - \boldsymbol{\delta}_{g,n}$, where $\tilde{\boldsymbol{\delta}}_{g,n}$ is some estimator for the parameter vector $\boldsymbol{\delta}_{g,n}$. Observe that under Assumptions 1-4, and given $\tilde{\boldsymbol{\delta}}_{g,n}$ a $n^{1/2}$ -consistent estimator, these quantities clearly satisfy Assumptions A.1-A.3 in light of Lemmata A.1 and A.2.

Lemma A.7 : Let \mathbf{A}_n^* be an $n \times n$ real matrix whose row and column sums are bounded uniformly in absolute value. Suppose Assumptions A.1 and A.2 hold, then:

(a) $n^{-1} E |\mathbf{u}'_n \mathbf{A}_n^* \mathbf{u}_n| = O(1)$, $\text{var}(n^{-1} \mathbf{u}'_n \mathbf{A}_n^* \mathbf{u}_n) = o(1)$ and

$$n^{-1} \tilde{\mathbf{u}}'_n \mathbf{A}_n^* \tilde{\mathbf{u}}_n - n^{-1} E \mathbf{u}'_n \mathbf{A}_n^* \mathbf{u}_n = o_p(1).$$

(b) $n^{-1} E |\mathfrak{D}'_n \mathbf{A}_n^* \mathbf{u}_n| = O(1)$, and

$$n^{-1} \mathfrak{D}'_n \mathbf{A}_n^* \tilde{\mathbf{u}}_n - n^{-1} \mathfrak{D}'_n \mathbf{A}_n^* \mathbf{u}_n = o_p(1).$$

(c) If furthermore Assumption A.3 holds, then

$$n^{-1/2}\tilde{\mathbf{u}}_n'\mathbf{A}_n^*\tilde{\mathbf{u}}_n = n^{-1/2}\mathbf{u}'_n\mathbf{A}_n^*\mathbf{u}_n + \boldsymbol{\alpha}_n^{*'}n^{1/2}\Delta_n + o_p(1)$$

with $\boldsymbol{\alpha}_n^* = n^{-1}E\mathcal{D}'_n(\mathbf{A}_n^* + \mathbf{A}_n^{*'})\mathbf{u}_n$. (Of course, in light of (a) and (b) we have $\boldsymbol{\alpha}_n^* = O(1)$ and $n^{-1}\mathcal{D}'_n(\mathbf{A}_n^* + \mathbf{A}_n^{*'})\tilde{\mathbf{u}}_n - \boldsymbol{\alpha}_n^* = o_p(1)$.)

We next state an additional assumption regarding Δ_n , which is satisfied by the various IV estimators considered in the paper; compare Lemmata A.4-A.6. This assumption then permits a specialization of the r.h.s. expression for $n^{-1/2}\tilde{\mathbf{u}}_n'\mathbf{A}_n^*\tilde{\mathbf{u}}_n$ given in Lemma A.7(c) for the case where $\mathbf{A}_n^* = \mathfrak{R}'_n\mathbf{A}_n\mathfrak{R}_n$, which will be helpful for deriving its limiting distribution. Note that given Assumption A.1 and $\mathbf{A}_n^* = \mathfrak{R}'_n\mathbf{A}_n\mathfrak{R}_n$ we have $\mathbf{u}'_n\mathbf{A}_n^*\mathbf{u}_n = \mathbf{u}'_n\mathfrak{R}'_n\mathbf{A}_n\mathfrak{R}_n\mathbf{u}_n = \boldsymbol{\epsilon}'_n\mathbf{A}_n\boldsymbol{\epsilon}_n$.

Assumption A.4 (a) The vector of innovations $\boldsymbol{\epsilon}_n = [\boldsymbol{\epsilon}'_{1,n}, \dots, \boldsymbol{\epsilon}'_{G,n}]'$ satisfies Assumption 4. (b) The estimator Δ_n is asymptotically linear in the sense that

$$n^{1/2}\Delta_n = n^{-1/2}\sum_{h=1}^G\mathbf{T}'_{h,n}\boldsymbol{\epsilon}_{h,n} + o_p(1),$$

with $\mathbf{T}_{h,n} = \mathbf{F}_{h,n}\mathbf{P}_{h,n}$ where the $\mathbf{F}_{h,n} = (f_{his,n})$ are $n \times p_F$ dimensional real nonstochastic matrices with $\sup_n n^{-1}\sum_{i=1}^n |f_{his,n}|^\eta < \infty$ with $\eta > 2$ for $h = 1, \dots, G, s = 1, \dots, p_F$, and $\mathbf{P}_{h,n} = (p_{hkl,n})$ are $p_F \times p_\Delta$ dimensional real nonstochastic matrices whose elements are uniformly bounded in absolute value.

We now have the following specialization of Lemma A.7(c).

Lemma A.8 : Suppose Assumptions A.1-A.3 hold where Δ_n is asymptotically linear satisfying Assumption A.4, and suppose $\boldsymbol{\epsilon}_n = \boldsymbol{\epsilon}_{g,n}$. Furthermore, let \mathbf{A}_n be an $n \times n$ real matrix whose row and column sums are bounded uniformly in absolute value.

(a) Then

$$n^{-1/2}\tilde{\mathbf{u}}_n'\mathfrak{R}'_n\mathbf{A}_n\mathfrak{R}_n\tilde{\mathbf{u}}_n = n^{-1/2}\boldsymbol{\epsilon}'_{g,n}\mathbf{A}_n\boldsymbol{\epsilon}_{g,n} + n^{-1/2}\sum_{h=1}^G\mathbf{a}'_{h,n}\boldsymbol{\epsilon}_{h,n} + o_p(1),$$

where $\mathbf{a}_{h,n} = (a_{h1,n}, \dots, a_{hn,n})' = \mathbf{T}_{h,n}\boldsymbol{\alpha}_n$ with $\boldsymbol{\alpha}_n = n^{-1}E\mathcal{D}'_n\mathfrak{R}'_n(\mathbf{A}_n + \mathbf{A}'_n)\mathfrak{R}_n\mathbf{u}_n$. Furthermore, $\sup_n n^{-1}\sum_{i=1}^n |a_{hi,n}|^\eta < \infty$ for $\eta > 2$.

(b) If furthermore the diagonal elements of \mathbf{A}_n are zero, then

$$\begin{aligned}
& E \left[n^{-1/2} \boldsymbol{\varepsilon}'_{g,n} \mathbf{A}_n \boldsymbol{\varepsilon}_{g,n} + n^{-1/2} \sum_{h=1}^G \mathbf{a}'_{h,n} \boldsymbol{\varepsilon}_{h,n} \right] = 0, \\
& \text{var} \left[n^{-1/2} \boldsymbol{\varepsilon}'_{g,n} \mathbf{A}_n \boldsymbol{\varepsilon}_{g,n} + n^{-1/2} \sum_{h=1}^G \mathbf{a}'_{h,n} \boldsymbol{\varepsilon}_{h,n} \right] \\
& = n^{-1} \sigma_{gg}^2 \text{tr} [\mathbf{A}_n (\mathbf{A}_n + \mathbf{A}'_n)] + n^{-1} \sum_{h=1}^G \sum_{l=1}^G \sigma_{hl} \mathbf{a}'_{h,n} \mathbf{a}_{l,n}.
\end{aligned}$$

B Appendix: Proofs for Section 4

Proof of Theorem 1: The existence and measurability of $\tilde{\boldsymbol{\rho}}_{g,n}$ is assured by, e.g., Lemma 3.4 in Pötscher and Prucha (1997). The objective function of the weighted nonlinear least squares estimator and its corresponding non-stochastic counterpart are given by, respectively,

$$\begin{aligned}
R_n(\omega, \bar{\boldsymbol{\rho}}_g) &= \mathbf{m}_{g,n}^\rho(\bar{\boldsymbol{\rho}}_g, \tilde{\boldsymbol{\delta}}_{g,n})' \tilde{\boldsymbol{\Upsilon}}_{g,n} \mathbf{m}_{g,n}^\rho(\bar{\boldsymbol{\rho}}_g, \tilde{\boldsymbol{\delta}}_{g,n}) \\
&= \left[\tilde{\boldsymbol{\Gamma}}_{g,n} \mathbf{r}_{g,n}(\bar{\boldsymbol{\rho}}_g) - \tilde{\boldsymbol{\gamma}}_{g,n} \right]' \tilde{\boldsymbol{\Upsilon}}_{g,n} \left[\tilde{\boldsymbol{\Gamma}}_{g,n} \mathbf{r}_{g,n}(\bar{\boldsymbol{\rho}}_g) - \tilde{\boldsymbol{\gamma}}_{g,n} \right] \\
\bar{R}_n(\bar{\boldsymbol{\rho}}_g) &= \left[\boldsymbol{\Gamma}_{g,n} \mathbf{r}_{g,n}(\bar{\boldsymbol{\rho}}_g) - \boldsymbol{\gamma}_n \right]' \boldsymbol{\Upsilon}_n \left[\boldsymbol{\Gamma}_{g,n} \mathbf{r}_{g,n}(\bar{\boldsymbol{\rho}}_g) - \boldsymbol{\gamma}_n \right],
\end{aligned}$$

where the quantities appearing in the above equation are defined before the theorem in the text. To prove the consistency of $\tilde{\boldsymbol{\rho}}_{g,n}$ we show that the conditions of, e.g., Lemma 3.1 in Pötscher and Prucha (1997) are satisfied for the problem at hand. We first show that $\boldsymbol{\rho}_{g,n}$ is an identifiably unique sequence of minimizers of \bar{R}_n . Observe that $\bar{R}_n(\bar{\boldsymbol{\rho}}_g) \geq 0$ and that $\bar{R}_n(\boldsymbol{\rho}_{g,n}) = 0$, since $\boldsymbol{\gamma}_n = \boldsymbol{\Gamma}_{g,n} \mathbf{r}_{g,n}(\boldsymbol{\rho}_{g,n})$ by (14). Utilizing Assumptions 8 and 9 we get

$$\begin{aligned}
& \bar{R}_n(\bar{\rho}_g) - \bar{R}_n(\rho_{g,n}) = \bar{R}_n(\bar{\rho}_g) \\
&= [\mathbf{\Gamma}_{g,n} \mathbf{r}_{g,n}(\bar{\rho}_g) - \mathbf{\Gamma}_{g,n} \mathbf{r}_{g,n}(\rho_{g,n})]' \mathbf{\Upsilon}_n [\mathbf{\Gamma}_{g,n} \mathbf{r}_{g,n}(\bar{\rho}_g) - \mathbf{\Gamma}_{g,n} \mathbf{r}_{g,n}(\rho_{g,n})] \\
&= [\mathbf{r}_{g,n}(\bar{\rho}_g) - \mathbf{r}_{g,n}(\rho_{g,n})]' \mathbf{\Gamma}'_{g,n} \mathbf{\Upsilon}_n \mathbf{\Gamma}_{g,n} [\mathbf{r}_{g,n}(\bar{\rho}_g) - \mathbf{r}_{g,n}(\rho_{g,n})] \\
&\geq \lambda_{\min}(\mathbf{\Upsilon}_n) \lambda_{\min}(\mathbf{\Gamma}'_{g,n} \mathbf{\Gamma}_{g,n}) [\mathbf{r}_{g,n}(\bar{\rho}_g) - \mathbf{r}_{g,n}(\rho_{g,n})]' [\mathbf{r}_{g,n}(\bar{\rho}_g) - \mathbf{r}_{g,n}(\rho_{g,n})] \\
&\geq \lambda_* \sum_{s=1}^q [\bar{\rho}_{g,s} - \rho_{g,s,n}]^2
\end{aligned}$$

for some $\lambda_* > 0$. Hence, for every $\varepsilon > 0$ and n we have

$$\begin{aligned}
& \inf_{\bar{\rho}_g: \sum_{r \in \mathbf{I}_{g,\rho}} |\bar{\rho}_{g,r}| \in [-a_\rho, a_\rho] \text{ and } \|\bar{\rho}_g - \rho_{g,n}\| \geq \varepsilon} [\bar{R}_n(\bar{\rho}_g) - \bar{R}_n(\rho_{g,n})] \\
&\geq \inf_{\bar{\rho}_g: \sum_{r \in \mathbf{I}_{g,\rho}} |\bar{\rho}_{g,r}| \in [-a_\rho, a_\rho] \text{ and } \|\bar{\rho}_g - \rho_{g,n}\| \geq \varepsilon} \lambda_* \sum_{r \in \mathbf{I}_{g,\rho}} [\bar{\rho}_{g,r} - \rho_{g,r,n}]^2 = \lambda_* \varepsilon^2 > 0,
\end{aligned}$$

which proves that $\rho_{g,n}$ is identifiably unique. Next let $\mathbf{\Phi}_n = [\mathbf{\Gamma}_n, -\gamma_n]$ and $\tilde{\mathbf{\Phi}}_n = [\tilde{\mathbf{\Gamma}}_n, -\tilde{\gamma}_n]$, then

$$\begin{aligned}
& |R_n(\omega, \bar{\rho}_g) - \bar{R}_n(\bar{\rho}_g)| \\
&= \left| [\mathbf{r}_{g,n}(\bar{\rho}_g)', 1] \left[\tilde{\mathbf{\Phi}}'_n \tilde{\mathbf{\Upsilon}}_n \tilde{\mathbf{\Phi}}_n - \mathbf{\Phi}'_n \mathbf{\Upsilon}_n \mathbf{\Phi}_n \right] [\mathbf{r}_{g,n}(\bar{\rho}_g)', 1]' \right| \\
&\leq \left\| \tilde{\mathbf{\Phi}}'_n \tilde{\mathbf{\Upsilon}}_n \tilde{\mathbf{\Phi}}_n - \mathbf{\Phi}'_n \mathbf{\Upsilon}_n \mathbf{\Phi}_n \right\| \left\| [\mathbf{r}_{g,n}(\bar{\rho}_g)', 1] \right\|^2 \\
&\leq \left\| \tilde{\mathbf{\Phi}}'_n \tilde{\mathbf{\Upsilon}}_n \tilde{\mathbf{\Phi}}_n - \mathbf{\Phi}'_n \mathbf{\Upsilon}_n \mathbf{\Phi}_n \right\| \{1 + [2q + q(q-1)/2] a_\rho^4\}.
\end{aligned}$$

Next observe that the elements of $\mathbf{\Phi}_n$ and $\tilde{\mathbf{\Phi}}_n$ are all of the form $n^{-1} E \mathbf{u}'_n A_n \mathbf{u}_n$ and $n^{-1} \tilde{\mathbf{u}}'_n A_n \tilde{\mathbf{u}}_n$, where the row and column sums of A_n are bounded uniformly in absolute value. Recall that $\mathbf{u}_{g,n} = \left[\sum_{r \in \mathbf{I}_{g,\rho}} \rho_{g,r,n} \mathbf{M}_{r,n} \right] \mathbf{u}_{g,n} + \boldsymbol{\varepsilon}_{g,n}$ and observe that $\tilde{\mathbf{u}}_{g,n} - \mathbf{u}_{g,n} = -\mathbf{Z}_{g,n} (\tilde{\boldsymbol{\delta}}_{g,n} - \boldsymbol{\delta}_{g,n})$. By Assumption 1 and 2 the row and column sums of the absolute elements of $\sum_{r \in \mathbf{I}_{g,\rho}} \rho_{g,r,n} \mathbf{M}_{r,n}$ and $\left[\sum_{r \in \mathbf{I}_{g,\rho}} \rho_{g,r,n} \mathbf{M}_{r,n} \right]^{-1}$ are bounded in absolute value by some finite constant. By Assumptions 4 the elements of $\boldsymbol{\varepsilon}_{g,n}$ are totally independent with zero mean, positive variance and finite $4 + v$ absolute moments for some $v > 0$. Furthermore, observe that by Lemma A.2 the fourth absolute moments of the elements of $\mathbf{Z}_{g,n}$ are uniformly bounded by a finite constant and $n^{1/2} (\tilde{\boldsymbol{\delta}}_{g,n} - \boldsymbol{\delta}_{g,n}) = O_p(1)$. It now follows immediately from Lemma A.7(a) that $\tilde{\mathbf{\Phi}}_n - \mathbf{\Phi}_n \xrightarrow{P} 0$, that the elements of $\mathbf{\Phi}_n$ are $O(1)$ and, consequently, that the elements of $\tilde{\mathbf{\Phi}}_n$ are $O_p(1)$.

The elements of $\tilde{\Upsilon}_n$ and Υ_n have the analogous properties in light of condition (b) in the theorem. Given this it follows from the above inequality that $R_n(\omega, \bar{\rho}_g) - \bar{R}_n(\bar{\rho}_g)$ converges to zero uniformly over the optimization space, i.e.,

$$\begin{aligned} & \sup_{\bar{\rho}_g: \sum_{r \in \mathcal{I}_{g,\rho}} |\bar{\rho}_{g,r}| \in [-a_\rho, a_\rho]} |R_n(\omega, \bar{\rho}_g) - \bar{R}_n(\bar{\rho}_g)| \\ & \leq \left\| \tilde{\Phi}'_n \tilde{\Upsilon}_n \tilde{\Phi}_n - \Phi'_n \Upsilon_n \Phi_n \right\| \{1 + [2q + q(q-1)/2] a_\rho^4\} \xrightarrow{P} 0 \end{aligned}$$

as $n \rightarrow \infty$. The consistency of $\tilde{\rho}_{g,n}$ now follows directly from Lemma 3.1 in Pötscher and Prucha (1997). \blacksquare

Proof of Theorem 2: We have shown in Theorem 1 that the GMM estimator $\tilde{\rho}_{g,n}$ defined in (16) is consistent. Apart on a set, whose probability tends to zero, the estimator satisfies the following first-order condition

$$\mathbf{m}_{g,n}^\rho(\tilde{\rho}_{g,n}, \tilde{\delta}_{g,n})' \tilde{\Upsilon}_{g,n} \frac{\partial \mathbf{m}_{g,n}^\rho(\tilde{\rho}_{g,n}, \tilde{\delta}_{g,n})}{\partial \rho_g} = 0.$$

Substituting the mean value theorem expression

$$\mathbf{m}_{g,n}^\rho(\tilde{\rho}_{g,n}, \tilde{\delta}_{g,n}) = \mathbf{m}_{g,n}^\rho(\rho_{g,n}, \tilde{\delta}_{g,n}) + \frac{\partial \mathbf{m}_{g,n}^\rho(\check{\rho}_{g,n}, \tilde{\delta}_{g,n})}{\partial \rho_g} (\tilde{\rho}_{g,n} - \rho_{g,n}),$$

where $\check{\rho}_{g,n}$ is some between value, into the first-order condition yields

$$\begin{aligned} & \frac{\partial \mathbf{m}_{g,n}^\rho(\tilde{\rho}_{g,n}, \tilde{\delta}_{g,n})}{\partial \rho'_g} \tilde{\Upsilon}_{g,n} \frac{\partial \mathbf{m}_{g,n}^\rho(\check{\rho}_{g,n}, \tilde{\delta}_{g,n})}{\partial \rho_g} n^{1/2} (\tilde{\rho}_{g,n} - \rho_{g,n}) \quad (\text{B.1}) \\ & = - \frac{\partial \mathbf{m}_{g,n}^\rho(\tilde{\rho}_{g,n}, \tilde{\delta}_{g,n})}{\partial \rho'_g} \tilde{\Upsilon}_{g,n} \mathbf{m}_{g,n}^\rho(\rho_{g,n}, \tilde{\delta}_{g,n}). \end{aligned}$$

Observe that

$$\frac{\partial \mathbf{m}_{g,n}^\rho(\rho_g, \tilde{\delta}_{g,n})}{\partial \rho_g} = -\tilde{\Gamma}_{g,n} \frac{\partial \mathbf{r}_{g,n}(\rho_g)}{\partial \rho_g} \quad (\text{B.2})$$

and consider

$$\begin{aligned}
\tilde{\Xi}_{g,n} &= \frac{\partial \mathbf{m}_{g,n}^\rho(\tilde{\rho}_{g,n}, \tilde{\delta}_{g,n})}{\partial \rho'_g} \tilde{\Upsilon}_{g,n} \frac{\partial \mathbf{m}_{g,n}^\rho(\check{\rho}_{g,n}, \check{\delta}_{g,n})}{\partial \rho_g} \\
&= \frac{\partial \mathbf{r}_{g,n}(\tilde{\rho}_{g,n})}{\partial \rho'_g} \tilde{\Gamma}'_{g,n} \tilde{\Upsilon}_{g,n} \tilde{\Gamma}_{g,n} \frac{\partial \mathbf{r}_{g,n}(\check{\rho}_{g,n})}{\partial \rho_g}, \\
\Xi_{g,n} &= \frac{\partial \mathbf{r}_{g,n}(\rho_{g,n})}{\partial \rho'_g} \Gamma'_{g,n} \Upsilon_{g,n} \Gamma_{g,n} \frac{\partial \mathbf{r}_{g,n}(\rho_{g,n})}{\partial \rho_g}.
\end{aligned} \tag{B.3}$$

In proving Theorem 1 we have demonstrated that $\tilde{\Gamma}_{g,n} - \Gamma_{g,n} \xrightarrow{p} \mathbf{0}$ and that the elements of $\tilde{\Gamma}_{g,n}$ and $\Gamma_{g,n}$ are $O_p(1)$ and $O(1)$, respectively. By Assumption 9 we have $\tilde{\Upsilon}_{g,n} - \Upsilon_{g,n} = o_p(1)$ and also that the elements of $\tilde{\Upsilon}_{g,n}$ and $\Upsilon_{g,n}$ are $O_p(1)$ and $O(1)$, respectively. Since $\tilde{\rho}_{g,n}$ and $\check{\rho}_{g,n}$ are consistent for $\rho_{g,n}$, and the elements of $\rho_{g,n}$ are bounded in absolute value, clearly

$$\tilde{\Xi}_{g,n} - \Xi_{g,n} \xrightarrow{p} \mathbf{0} \tag{B.4}$$

as $n \rightarrow \infty$, and furthermore $\tilde{\Xi}_{g,n} = O_p(1)$ and $\Xi_{g,n} = O(1)$. In particular $\lambda_{\max}(\Xi_{g,n}) \leq \lambda_{\Xi}^{**}$ where λ_{Ξ}^{**} is some finite constant. Observe that in light of Assumptions 8 and 9 we have

$$\lambda_{\min}(\Xi_{g,n}) \geq \lambda_{\min}(\Upsilon_{g,n}) \lambda_{\min}(\Gamma'_{g,n} \Gamma_{g,n}) \lambda_{\min} \left\{ \frac{\partial \mathbf{r}_{g,n}(\rho_{g,n})}{\partial \rho'_g} \frac{\partial \mathbf{r}_{g,n}(\rho_{g,n})}{\partial \rho_g} \right\} \geq \lambda_{\Xi}^*$$

for some $\lambda_{\Xi}^* > 0$, observing that

$$\lambda_{\min} \left\{ \frac{\partial \mathbf{r}_{g,n}(\rho_{g,n})}{\partial \rho'_g} \frac{\partial \mathbf{r}_{g,n}(\rho_{g,n})}{\partial \rho_g} \right\} = \lambda_{\min} \{ I_{q_g} + \text{semi-positive matrix} \} \geq 1.$$

Hence $0 < \lambda_{\max}\{\Xi_{g,n}^{-1}\} = 1/\lambda_{\min}(\Xi_{g,n}) \leq 1/\lambda_{\Xi}^* < \infty$, and thus we also have $\Xi_{g,n}^{-1} = O(1)$. Let $\tilde{\Xi}_{g,n}^+$ denote the generalized inverse of $\tilde{\Xi}_{g,n}$. It then follows as a special case of Lemma F1 in Pötscher and Prucha (1997) that $\tilde{\Xi}_{g,n}$ is nonsingular eventually with probability tending to one, that $\tilde{\Xi}_{g,n}^+ = O_p(1)$, and that

$$\tilde{\Xi}_{g,n}^+ - \Xi_{g,n}^{-1} \xrightarrow{p} \mathbf{0} \tag{B.5}$$

as $n \rightarrow \infty$.

Premultiplying (B.1) by $\tilde{\Xi}_{g,n}^+$ and rearranging terms yields

$$\begin{aligned} n^{1/2}(\tilde{\rho}_{g,n} - \rho_{g,n}) &= \left[I - \tilde{\Xi}_{g,n}^+ \tilde{\Xi}_{g,n} \right] n^{1/2}(\tilde{\rho}_{g,n} - \rho_{g,n}) \\ &\quad - \tilde{\Xi}_{g,n}^+ \frac{\partial \mathbf{m}_{g,n}^\rho(\tilde{\rho}_{g,n}, \tilde{\delta}_{g,n})}{\partial \rho'_g} \tilde{\Upsilon}_{g,n} n^{1/2} \mathbf{m}_{g,n}^\rho(\rho_{g,n}, \tilde{\delta}_{g,n}). \end{aligned}$$

In light of the above discussion the first term on the r.h.s. is zero on ω -sets of probability tending to one. This yields

$$n^{1/2}(\tilde{\rho}_{g,n} - \rho_{g,n}) = -\tilde{\Xi}_{g,n}^+ \frac{\partial \mathbf{m}_{g,n}^\rho(\tilde{\rho}_{g,n}, \tilde{\delta}_{g,n})}{\partial \rho'_g} \tilde{\Upsilon}_{g,n} n^{1/2} \mathbf{m}_{g,n}^\rho(\rho_{g,n}, \tilde{\delta}_{g,n}) + o_p(1). \quad (\text{B.6})$$

Observe that

$$\tilde{\Xi}_{g,n}^+ \frac{\partial \mathbf{m}_{g,n}^\rho(\tilde{\rho}_{g,n}, \tilde{\delta}_{g,n})}{\partial \rho'_g} \tilde{\Upsilon}_{g,n} - \Xi_{g,n}^{-1} \frac{\partial \mathbf{r}_{g,n}(\rho_{n.g})}{\partial \rho'_g} \Gamma'_{g,n} \Upsilon_{g,n} = o_p(1). \quad (\text{B.7})$$

In light of (15) the elements of $n^{1/2} \mathbf{m}_{g,n}^\rho(\rho_{g,n}, \tilde{\delta}_{g,n})$ are of the form ($s = 1, \dots, S$)

$$n^{-1/2} \tilde{\mathbf{u}}'_{g,n} [\mathbf{I}_n - \mathbf{R}_{g,n}^{*'}(\rho_{g,n})] \mathbf{A}_{s,n} [\mathbf{I}_n - \mathbf{R}_{g,n}^*(\rho_{g,n})] \tilde{\mathbf{u}}_{g,n}$$

with

$$\begin{aligned} [\mathbf{I}_n - \mathbf{R}_{g,n}^*(\rho_{g,n})] \mathbf{u}_{g,n} &= \boldsymbol{\varepsilon}_{g,n}, \\ \tilde{\mathbf{u}}_{g,n} - \mathbf{u}_{g,n} &= -\mathbf{Z}_{g,n}(\tilde{\delta}_{g,n} - \delta_{g,n}), \\ n^{1/2}(\tilde{\delta}_{g,n} - \delta_{g,n}) &= n^{-1/2} \sum_{h=1}^G \mathbf{T}'_{gh,n} \boldsymbol{\varepsilon}_{h,n} + o_p(1). \end{aligned}$$

Recall that if we define $\tilde{\mathbf{u}}_n = \tilde{\mathbf{u}}_{g,n}$, $\mathbf{u}_n = \mathbf{u}_{g,n}$, $\boldsymbol{\varepsilon}_n = \boldsymbol{\varepsilon}_{g,n}$, $\mathfrak{R}_n = \mathbf{I}_n - \mathbf{R}_{g,n}^*(\rho_{g,n})$, $\mathfrak{D}_n = -\mathbf{Z}_{g,n}$, $\boldsymbol{\Delta}_n = \tilde{\delta}_{g,n} - \delta_{g,n}$ these quantities satisfy Assumptions A.1-A.3 in light of Lemmata A.1-A.3. Hence it follows from Lemma A.8 that

$$\begin{aligned} &n^{-1/2} \tilde{\mathbf{u}}'_{g,n} [\mathbf{I}_n - \mathbf{R}_{g,n}^{*'}(\rho_{g,n})] \mathbf{A}_{s,n} [\mathbf{I}_n - \mathbf{R}_{g,n}^*(\rho_{g,n})] \tilde{\mathbf{u}}_{g,n} \\ &= n^{-1/2} \frac{1}{2} \boldsymbol{\varepsilon}'_{g,n} (\mathbf{A}_{s,n} + \mathbf{A}'_{s,n}) \boldsymbol{\varepsilon}_{g,n} + n^{-1/2} \alpha'_{g,s,n} \sum_{h=1}^G \mathbf{T}'_{gh,n} \boldsymbol{\varepsilon}_{h,n} + o_p(1), \end{aligned}$$

where $\alpha_{g,s,n} = -n^{-1} E \mathbf{Z}'_{g,n} [\mathbf{I}_n - \mathbf{R}_{g,n}^{*'}(\rho_{g,n})] (\mathbf{A}_{s,n} + \mathbf{A}'_{s,n}) \boldsymbol{\varepsilon}_{g,n}$. Furthermore, the lemma implies that for some $\eta > 2$ the sample moments of the absolute

elements $\mathbf{T}_{h,n}\boldsymbol{\alpha}_{g,s,n}$ are uniformly bounded. We now have

$$\begin{aligned} & n^{1/2}\mathbf{m}_{g,n}^\rho(\boldsymbol{\rho}_{g,n},\tilde{\delta}_{g,n}) \\ = & n^{-1/2}\begin{bmatrix} \frac{1}{2}\boldsymbol{\varepsilon}'_{g,n}(\mathbf{A}_{1,n}+\mathbf{A}'_{1,n})\boldsymbol{\varepsilon}_{g,n}+\boldsymbol{\alpha}'_{g,1,n}\sum_{h=1}^G\mathbf{T}'_{gh,n}\boldsymbol{\varepsilon}_{h,n} \\ \vdots \\ \frac{1}{2}\boldsymbol{\varepsilon}'_{g,n}(\mathbf{A}_{S,n}+\mathbf{A}'_{S,n})\boldsymbol{\varepsilon}_{g,n}+\boldsymbol{\alpha}'_{g,S,n}\sum_{h=1}^G\mathbf{T}'_{gh,n}\boldsymbol{\varepsilon}_{h,n} \end{bmatrix}+o_p(1). \end{aligned} \quad (\text{B.8})$$

Let $\boldsymbol{\Psi}_{gg,n}^{\rho\rho}=(\psi_{rs,gg,n}^{\rho\rho})$ denote the VC matrix of the vector of linear quadratic forms on the r.h.s. of (B.8), then observing that the diagonal elements of $\mathbf{A}_{s,n}$ are zero and that the VC matrix of $\boldsymbol{\varepsilon}_n$ is $\boldsymbol{\Sigma}\otimes\mathbf{I}_n$, it follows from Lemma A.1 in Kelejian and Prucha (2010) that

$$\begin{aligned} \psi_{rs,gg,n}^{\rho\rho} &= \frac{1}{2n}\sigma_{gg}^2\text{tr}\left[(\mathbf{A}_{r,n}+\mathbf{A}'_{r,n})(\mathbf{A}_{s,n}+\mathbf{A}'_{s,n})\right] \\ &+ \frac{1}{n}\boldsymbol{\alpha}'_{g,r,n}\left[\sum_{h=1}^G\sum_{l=1}^G\sigma_{hl}\mathbf{T}'_{gh,n}\mathbf{T}_{gl,n}\right]\boldsymbol{\alpha}_{g,s,n}. \end{aligned} \quad (\text{B.9})$$

We note that $\lambda_{\min}(\boldsymbol{\Psi}_{gg,n}^{\rho\rho})\geq\text{const}>0$ by assumption. Since the matrices $\mathbf{A}_{s,n}$, the vectors of the linear forms, and – in light of Assumption 4 – the innovations $\boldsymbol{\varepsilon}_n$ satisfy all of the remaining assumptions of the central limit theorem for vectors of linear quadratic forms given in Kelejian and Prucha (2010) it now follows that

$$\begin{aligned} \boldsymbol{\xi}_{g,n} &= -(\boldsymbol{\Psi}_{gg,n}^{\rho\rho})^{-1/2}n^{-1/2}\begin{bmatrix} \frac{1}{2}\boldsymbol{\varepsilon}'_{g,n}(\mathbf{A}_{1,n}+\mathbf{A}'_{1,n})\boldsymbol{\varepsilon}_{g,n}+\boldsymbol{\alpha}'_{g,1,n}\sum_{h=1}^G\mathbf{T}'_{gh,n}\boldsymbol{\varepsilon}_{h,n} \\ \vdots \\ \frac{1}{2}\boldsymbol{\varepsilon}'_{g,n}(\mathbf{A}_{S,n}+\mathbf{A}'_{S,n})\boldsymbol{\varepsilon}_{g,n}+\boldsymbol{\alpha}'_{g,S,n}\sum_{h=1}^G\mathbf{T}'_{gh,n}\boldsymbol{\varepsilon}_{h,n} \end{bmatrix} \\ &\xrightarrow{d}N(\mathbf{0},\mathbf{I}_S). \end{aligned} \quad (\text{B.10})$$

Next observe that $\boldsymbol{\Psi}_{gg,n}^{\rho\rho}=O(1)$, and hence $(\boldsymbol{\Psi}_{gg,n}^{\rho\rho})^{1/2}=O(1)$, since the row and column sums of the absolute elements of $\mathbf{A}_{r,n}$ are uniformly bounded by assumption, and since in light of the above assumptions the terms $n^{-1}\boldsymbol{\alpha}'_{g,r,n}\mathbf{T}'_{gh,n}\mathbf{T}_{gl,n}\boldsymbol{\alpha}'_{g,s,n}$ are uniformly bounded in absolute value. It now follows from (B.6), (B.7) and (B.10) that

$$n^{1/2}(\tilde{\boldsymbol{\rho}}_{g,n}-\boldsymbol{\rho}_{g,n})=[\mathbf{J}'_{g,n}\boldsymbol{\Upsilon}_{g,n}\mathbf{J}_{g,n}]^{-1}\mathbf{J}'_{g,n}\boldsymbol{\Upsilon}_{g,n}(\boldsymbol{\Psi}_{gg,n}^{\rho\rho})^{1/2}\boldsymbol{\xi}_{g,n}+o_p(1), \quad (\text{B.11})$$

observing that $\boldsymbol{\Xi}_{g,n}=\mathbf{J}'_{g,n}\boldsymbol{\Upsilon}_{g,n}\mathbf{J}_{g,n}$. This establishes (19). Since all of the nonstochastic terms on the r.h.s. of (B.11) are $O(1)$, it follows that $n^{1/2}(\tilde{\boldsymbol{\rho}}_{g,n}-$

$\boldsymbol{\rho}_{g,n} = O_p(1)$. Next recall that $0 < \lambda_{\Xi}^* \leq \lambda_{\min}(\boldsymbol{\Xi}_{g,n}) \leq \lambda_{\max}(\boldsymbol{\Xi}_{g,n}) \leq \lambda_{\Xi}^{**} < \infty$ and observe that $\lambda_{\min}(\boldsymbol{\Xi}_{g,n}^{-1}) = \mathbf{1}/\lambda_{\max}(\boldsymbol{\Xi}_{g,n})$. Hence

$$\begin{aligned} & \lambda_{\min} \{ \boldsymbol{\Xi}_{g,n}^{-1} \mathbf{J}'_{g,n} \boldsymbol{\Upsilon}_{g,n} \boldsymbol{\Psi}_{gg,n}^{\rho\rho} \boldsymbol{\Upsilon}_{g,n} \mathbf{J}_{g,n} \boldsymbol{\Xi}_{g,n}^{-1} \} \\ & \geq \lambda_{\min}(\boldsymbol{\Psi}_{gg,n}^{\rho\rho}) \lambda_{\min}(\boldsymbol{\Upsilon}_{g,n}) \lambda_{\min}(\boldsymbol{\Xi}_{g,n}^{-1} \mathbf{J}'_{g,n} \boldsymbol{\Upsilon}_{g,n} \mathbf{J}_{g,n} \boldsymbol{\Xi}_{g,n}^{-1}) \\ & \geq \lambda_{\min}(\boldsymbol{\Psi}_{gg,n}^{\rho\rho}) \lambda_{\min}(\boldsymbol{\Upsilon}_{g,n}) / \lambda_{\Xi}^{**} \geq \text{const} > 0. \end{aligned}$$

This establishes the last claim of the theorem. \blacksquare

Proof of Lemma 1: Observe that $\tilde{\mathbf{u}}_{g,n} = \mathbf{u}_{g,n} - \mathbf{Z}_{g,n} \boldsymbol{\Delta}_{g,n}$ with $\boldsymbol{\Delta}_{g,n} = \tilde{\boldsymbol{\delta}}_{g,n} - \boldsymbol{\delta}_{g,n}$, and thus

$$\begin{aligned} \tilde{\boldsymbol{\varepsilon}}_{g,n} &= [\mathbf{I}_n - \mathbf{R}_{g,n}^*(\tilde{\boldsymbol{\rho}}_{g,n})] \tilde{\mathbf{u}}_{g,n} \\ &= \boldsymbol{\varepsilon}_{g,n} - \mathbf{R}_{g,n}^*(\tilde{\boldsymbol{\rho}}_{g,n} - \boldsymbol{\rho}_{g,n}) \mathbf{u}_{g,n} - [\mathbf{I}_n - \mathbf{R}_{g,n}^*(\tilde{\boldsymbol{\rho}}_{g,n})] \mathbf{Z}_{g,n} \boldsymbol{\Delta}_{g,n}. \end{aligned}$$

Consequently

$$\begin{aligned} \tilde{\sigma}_{gh,n} &= n^{-1} \tilde{\boldsymbol{\varepsilon}}'_{g,n} \tilde{\boldsymbol{\varepsilon}}_{h,n} = n^{-1} \boldsymbol{\varepsilon}'_{g,n} \boldsymbol{\varepsilon}_{h,n} \\ &+ n^{-1} \mathbf{u}'_{g,n} [\mathbf{R}_{g,n}^*(\tilde{\boldsymbol{\rho}}_{g,n} - \boldsymbol{\rho}_{g,n})]' \mathbf{R}_{h,n}^*(\tilde{\boldsymbol{\rho}}_{h,n} - \boldsymbol{\rho}_{h,n}) \mathbf{u}_{h,n} \\ &+ \boldsymbol{\Delta}'_{g,n} n^{-1} \mathbf{Z}'_{g,n} [\mathbf{I}_n - \mathbf{R}_{g,n}^*(\tilde{\boldsymbol{\rho}}_{g,n})]' [\mathbf{I}_n - \mathbf{R}_{h,n}^*(\tilde{\boldsymbol{\rho}}_{h,n})] \mathbf{Z}_{h,n} \boldsymbol{\Delta}_{h,n} \\ &- n^{-1} \boldsymbol{\varepsilon}'_{g,n} \mathbf{R}_{h,n}^*(\tilde{\boldsymbol{\rho}}_{h,n} - \boldsymbol{\rho}_{h,n}) \mathbf{u}_{h,n} - n^{-1} \boldsymbol{\varepsilon}'_{h,n} \mathbf{R}_{g,n}^*(\tilde{\boldsymbol{\rho}}_{g,n} - \boldsymbol{\rho}_{g,n}) \mathbf{u}_{g,n} \\ &- n^{-1} \boldsymbol{\varepsilon}'_{g,n} [\mathbf{I}_n - \mathbf{R}_{h,n}^*(\tilde{\boldsymbol{\rho}}_{h,n})] \mathbf{Z}_{h,n} \boldsymbol{\Delta}_{h,n} - n^{-1} \boldsymbol{\varepsilon}'_{h,n} [\mathbf{I}_n - \mathbf{R}_{g,n}^*(\tilde{\boldsymbol{\rho}}_{g,n})] \mathbf{Z}_{g,n} \boldsymbol{\Delta}_{g,n} \\ &+ n^{-1} \mathbf{u}'_{g,n} [\mathbf{R}_{g,n}^*(\tilde{\boldsymbol{\rho}}_{g,n} - \boldsymbol{\rho}_{g,n})]' [\mathbf{I}_n - \mathbf{R}_{h,n}^*(\tilde{\boldsymbol{\rho}}_{h,n})] \mathbf{Z}_{h,n} \boldsymbol{\Delta}_{h,n} \\ &+ n^{-1} \mathbf{u}'_{h,n} [\mathbf{R}_{h,n}^*(\tilde{\boldsymbol{\rho}}_{h,n} - \boldsymbol{\rho}_{h,n})]' [\mathbf{I}_n - \mathbf{R}_{g,n}^*(\tilde{\boldsymbol{\rho}}_{g,n})] \mathbf{Z}_{g,n} \boldsymbol{\Delta}_{g,n}. \end{aligned}$$

By Assumption 4 we have $\sigma_{gh} = \sum_{l=1}^G \sigma_{*lg} \sigma_{*lh}$ and $\boldsymbol{\varepsilon}_g = \sum_{l=1}^G \sigma_{*lg} \mathbf{v}_l$. Since the elements of \mathbf{v} are i.i.d. (0,1), it follows that $n^{-1} \mathbf{v}'_l \mathbf{v}_l = 1 + o_p(1)$ and $n^{-1} \mathbf{v}'_l \mathbf{v}_k = o_p(1)$ for $l \neq k$. Hence

$$n^{-1} \boldsymbol{\varepsilon}'_{g,n} \boldsymbol{\varepsilon}_{h,n} = \sum_{l=1}^G \sum_{k=1}^G \sigma_{*lg} \sigma_{*kh} n^{-1} \mathbf{v}'_l \mathbf{v}_k = \sum_{l=1}^G \sigma_{*lg} \sigma_{*lh} + o_p(1) = \sigma_{gh} + o_p(1).$$

Next observe that, taking into account that $\boldsymbol{\varepsilon}_{h,n} = [\mathbf{I}_n - \mathbf{R}_{g,n}^*(\boldsymbol{\rho}_{g,n})] \mathbf{u}_{g,n}$, all the other terms consist of expressions of the form $o_p(1) n^{-1} \mathbf{u}'_{g,n} \mathbf{A}_n \mathbf{u}_{h,n}$, $o_p(1) n^{-1} \mathbf{u}'_{g,n} \mathbf{A}_n \mathbf{Z}_{h,n}$ and $o_p(1) n^{-1} \mathbf{Z}'_{g,n} \mathbf{A}_n \mathbf{Z}_{h,n}$, where \mathbf{A}_n is a matrix whose row and column sums of the absolute elements is uniformly bounded. In light of Lemma A.3 all those expressions are seen to be $o_p(1)$, and thus $\tilde{\sigma}_{gh,n} - \sigma_{gh} =$

$o_p(1)$ as claimed. \blacksquare

Proof of Theorem 3: We first demonstrate that $\tilde{\Psi}_{gg,n}^{\rho\rho} - \Psi_{gg,n}^{\rho\rho} = o_p(1)$, where the elements of $\tilde{\Psi}_{gg,n}^{\rho\rho}$ and $\Psi_{gg,n}^{\rho\rho}$ are defined in (18) and (24). By Lemma 1 we have $\tilde{\sigma}_{gh,n} - \sigma_{gh} = o_p(1)$. Furthermore, by assumption $n^{-1}\tilde{\mathbf{T}}_{gh,n}'\tilde{\mathbf{T}}_{gl,n} - n^{-1}\mathbf{T}_{gh,n}'\mathbf{T}_{gl,n} = o_p(1)$, where $n^{-1}\mathbf{T}_{gh,n}'\mathbf{T}_{gl,n} = O(1)$ in light of Assumption 10. Next, observe that under the maintained assumption the row and column sums of the absolute elements of the matrices $\mathbf{A}_{r,n}$ and $\mathbf{A}_{s,n}$ are uniformly bounded, and thus clearly are those of the matrices $\mathbf{A}_n = (a_{ij,n})$ with $a_{ij,n} = (a_{ij,r,n} + a_{ji,r,n})(a_{ij,s,n} + a_{ji,s,n})$. It then follows directly from Lemma A.7 that $\tilde{\alpha}_{g,r,n} - \alpha_{g,r,n} = o_p(1)$, where $\alpha_{g,r,n} = O(1)$ and $\tilde{\alpha}_{g,r,n} = O_p(1)$. Hence, clearly $\tilde{\psi}_{rs,gg,n}^{\rho\rho} - \psi_{rs,gg,n}^{\rho\rho} = o_p(1)$, $\psi_{rs,gg,n}^{\rho\rho} = O(1)$ and $\tilde{\psi}_{rs,gg,n}^{\rho\rho} = O_p(1)$, as well as $\tilde{\Psi}_{gg,n}^{\rho\rho} - \Psi_{gg,n}^{\rho\rho} = o_p(1)$, $\Psi_{gg,n}^{\rho\rho} = O(1)$ and $\tilde{\Psi}_{gg,n}^{\rho\rho} = O_p(1)$. Observing that $\lambda_{\min}(\Psi_{gg,n}^{\rho\rho}) \geq c_{\Psi}^* > 0$ it follows further that $(\tilde{\Psi}_{gg,n}^{\rho\rho})^{-1} - (\Psi_{gg,n}^{\rho\rho})^{-1} = o_p(1)$, $(\Psi_{gg,n}^{\rho\rho})^{-1} = O(1)$ and $(\tilde{\Psi}_{gg,n}^{\rho\rho})^{-1} = O_p(1)$.

By Assumption 9, $\tilde{\Upsilon}_{g,n} - \Upsilon_{g,n} = o_p(1)$, $\Upsilon_{g,n} = O(1)$ and thus $\tilde{\Upsilon}_{g,n} = O_p(1)$. In proving Theorem 2 we have verified that $\tilde{\mathbf{J}}_{g,n} - \mathbf{J}_{g,n} \xrightarrow{p} 0$, $\mathbf{J}_{g,n} = O(1)$ and $\tilde{\mathbf{J}}_{g,n} = O_p(1)$, and furthermore that $(\tilde{\mathbf{J}}_{g,n}'\tilde{\Upsilon}_{g,n}\tilde{\mathbf{J}}_{g,n})^+ - (\mathbf{J}_{g,n}'\Upsilon_{g,n}\mathbf{J}_{g,n})^{-1} = o_p(1)$, $(\tilde{\mathbf{J}}_{g,n}'\tilde{\Upsilon}_{g,n}\tilde{\mathbf{J}}_{g,n})^+ = O_p(1)$ and $(\mathbf{J}_{g,n}'\Upsilon_{g,n}\mathbf{J}_{g,n})^{-1} = O(1)$. The claim that $\tilde{\Omega}_{gg,n}^{\rho\rho} - \Omega_{gg,n}^{\rho\rho} = o_p(1)$ and $\Omega_{gg,n}^{\rho\rho} = O(1)$ is now obvious. Observing that $\lambda_{\min}[\Omega_{gg,n}^{\rho\rho}(\Upsilon_{g,n})] \geq \text{const} > 0$ by Theorem 2 it follows further that $(\tilde{\Omega}_{gg,n}^{\rho\rho})^{-1} - (\Omega_{gg,n}^{\rho\rho})^{-1} = o_p(1)$, $(\Omega_{gg,n}^{\rho\rho})^{-1} = O(1)$ and $(\tilde{\Omega}_{gg,n}^{\rho\rho})^{-1} = O_p(1)$. \blacksquare

We shall utilize the following lemma.

Lemma B.1 : *Suppose the $nG \times 1$ vector of innovations ε_n is generated as postulated in Assumption 4. Let $\mathbf{A} = \text{diag}_{g=1}^G(\mathbf{A}_{gg})$, and $\mathbf{B} = \text{diag}_{g=1}^G(\mathbf{B}_{gg})$ be symmetric $nG \times nG$ matrices with zero diagonal elements, and let $\mathbf{a} = [\mathbf{a}'_1, \dots, \mathbf{a}'_G]'$ and $\mathbf{b} = [\mathbf{b}'_1, \dots, \mathbf{b}'_G]'$ be $nG \times 1$ nonstochastic vectors. Then*

$$\begin{aligned} E(\varepsilon' \mathbf{A} \varepsilon + \mathbf{a}' \varepsilon) &= \mathbf{0} \\ \text{cov}(\varepsilon' \mathbf{A} \varepsilon + \mathbf{a}' \varepsilon, \varepsilon' \mathbf{B} \varepsilon + \mathbf{b}' \varepsilon) &= 2 \text{tr} [\mathbf{A}(\boldsymbol{\Sigma} \otimes \mathbf{I})\mathbf{B}(\boldsymbol{\Sigma} \otimes \mathbf{I})] + \mathbf{a}'(\boldsymbol{\Sigma} \otimes \mathbf{I})\mathbf{b} \\ &= 2 \sum_{h=1}^G \sum_{l=1}^G \sigma_{hl}^2 \text{tr} [\mathbf{A}_{hh}\mathbf{B}_{ll}] + \sum_{h=1}^G \sum_{l=1}^G \sigma_{hl} \mathbf{a}'_h \mathbf{b}_l. \end{aligned}$$

Proof. The result follows immediately from Lemma A1 in Kelejian and Prucha (2010), observing that the diagonal elements of $\mathbf{A}_* = (\boldsymbol{\Sigma}_* \otimes \mathbf{I})\mathbf{A}(\boldsymbol{\Sigma}'_* \otimes \mathbf{I})$ and $\mathbf{B}_* = (\boldsymbol{\Sigma}_* \otimes \mathbf{I})\mathbf{B}(\boldsymbol{\Sigma}'_* \otimes \mathbf{I})$ are zero. \blacksquare

Proof of Theorem 4: Observe that $\boldsymbol{\eta}_n = n^{-1/2}\mathbf{T}'_n\boldsymbol{\varepsilon}_n$ and thus clearly $\boldsymbol{\Psi}_n^{\delta\delta} = E\boldsymbol{\eta}_n\boldsymbol{\eta}'_n = n^{-1}\mathbf{T}'_nE\boldsymbol{\varepsilon}_n\boldsymbol{\varepsilon}'_n\mathbf{T}_n = n^{-1}\mathbf{T}'_n(\boldsymbol{\Sigma} \otimes \mathbf{I}_n)\mathbf{T}_n$ as claimed.

Next observe that $\boldsymbol{\xi}_{gs,n} = n^{-1/2}[\boldsymbol{\varepsilon}'_n\mathbf{B}\boldsymbol{\varepsilon}_n + \mathbf{b}'\boldsymbol{\varepsilon}_n]$ with $\mathbf{B} = \text{diag}(\mathbf{0}, \dots, \mathbf{B}_{gg}, \dots, \mathbf{0})$, where $\mathbf{B}_{gg} = (\mathbf{A}_{s,n} + \mathbf{A}'_{s,n})/2$ and $\mathbf{b} = \mathbf{T}_{g,n}\boldsymbol{\alpha}_{g,s,n}$ with $\mathbf{T}_{g,n} = [\mathbf{T}'_{g1,n}, \dots, \mathbf{T}'_{gG,n}]'$. It now follows from Lemma B.1 that

$$\begin{aligned} \text{cov}(\boldsymbol{\eta}_n, \boldsymbol{\xi}_{gs,n}) &= \text{cov}(n^{-1/2}\mathbf{T}'_n\boldsymbol{\varepsilon}_n, n^{-1/2}[\boldsymbol{\varepsilon}'_n\mathbf{B}\boldsymbol{\varepsilon}_n + \mathbf{b}'\boldsymbol{\varepsilon}_n]) = \text{cov}(n^{-1/2}\boldsymbol{\eta}_n, n^{-1/2}\mathbf{b}'\boldsymbol{\varepsilon}_n) \\ &= n^{-1}\mathbf{T}'_n(\boldsymbol{\Sigma} \otimes \mathbf{I}_n)\mathbf{T}_{g,n}\boldsymbol{\alpha}_{g,s,n} \end{aligned}$$

and thus

$$\text{cov}(\boldsymbol{\eta}_n, \boldsymbol{\xi}_{g,n}) = n^{-1}\mathbf{T}'_n(\boldsymbol{\Sigma} \otimes \mathbf{I}_n)\mathbf{T}_{g,n}[\boldsymbol{\alpha}_{g,1,n}, \dots, \boldsymbol{\alpha}_{g,S,n}]$$

and

$$\boldsymbol{\Psi}_n^{\delta\rho} = \text{cov}(\boldsymbol{\eta}_n, \boldsymbol{\xi}_n) = n^{-1}\mathbf{T}'_n(\boldsymbol{\Sigma} \otimes \mathbf{I}_n)\mathbf{T}_n \text{diag}_{g=1}^G [\boldsymbol{\alpha}_{g,1,n}, \dots, \boldsymbol{\alpha}_{g,S,n}]$$

as claimed.

Define $\boldsymbol{\xi}_{gr,n} = n^{-1/2}[\boldsymbol{\varepsilon}'_n\mathbf{A}\boldsymbol{\varepsilon}_n + \mathbf{a}'\boldsymbol{\varepsilon}_n]$ with $\mathbf{A} = \text{diag}(\mathbf{0}, \dots, \mathbf{A}_{gg}, \dots, \mathbf{0})$ where $\mathbf{A}_{gg} = (\mathbf{A}_{r,n} + \mathbf{A}'_{r,n})/2$ and $\mathbf{a} = \mathbf{T}_{g,n}\boldsymbol{\alpha}_{g,r,n}$. Then applying Lemma B.1 we see that

$$\begin{aligned} \text{cov}(\boldsymbol{\xi}_{gr,n}, \boldsymbol{\xi}_{hs,n}) &= \sigma_{gh,n}^2(2n)^{-1} \text{tr} [(\mathbf{A}_{r,n} + \mathbf{A}'_{r,n})(\mathbf{A}_{s,n} + \mathbf{A}'_{s,n})] \\ &\quad + \boldsymbol{\alpha}'_{g,r,n} \left[\sum_{u=1}^G \sum_{v=1}^G \sigma_{uv,n} \mathbf{T}'_{gu,n} \mathbf{T}_{hv,n} \right] \boldsymbol{\alpha}_{h,s,n}, \end{aligned}$$

which verifies the expressions for the elements of $\boldsymbol{\Psi}_n = \text{cov}(\boldsymbol{\xi}_n, \boldsymbol{\xi}_n)$.

In light of Assumption 10 and Theorem 2 we have

$$n^{1/2} \begin{bmatrix} \tilde{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_n \\ \tilde{\boldsymbol{\rho}}_n - \boldsymbol{\rho}_n \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \text{diag}[(\mathbf{J}'_{g,n} \boldsymbol{\Upsilon}_{g,n} \mathbf{J}_{g,n})^{-1} \mathbf{J}'_{g,n} \boldsymbol{\Upsilon}_{g,n}] \end{bmatrix} \begin{bmatrix} \boldsymbol{\eta}_n \\ \boldsymbol{\xi}_n \end{bmatrix} + o_p(1).$$

The vector of linear-quadratic forms $[\boldsymbol{\eta}'_n, \boldsymbol{\xi}'_n]'$ is readily seen to satisfy the assumptions of Theorem A.1 in Kelejian and Prucha (2010). Hence it follows from

that central limit theorem that

$$\Psi_n^{-1/2} \begin{bmatrix} \boldsymbol{\eta}_n \\ \boldsymbol{\xi}_n \end{bmatrix} \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_d),$$

which proves the first part of the theorem.

Note that $\lambda_{\min}(\Psi_{gg,n}^{\rho\rho}) \geq \lambda_{\min}(\Psi_n^{\rho\rho}) \geq \lambda_{\min}(\Psi_n) \geq \text{const} > 0$. In proving Theorem 3 we have shown that $\tilde{\Psi}_{gg,n}^{\rho\rho} - \Psi_{gg,n}^{\rho\rho} = o_p(1)$, $\Psi_{gg,n}^{\rho\rho} = O(1)$ and $\tilde{\Psi}_{gg,n}^{\rho\rho} = O_p(1)$. The proof that $\tilde{\Psi}_{gh,n}^{\rho\rho} - \Psi_{gh,n}^{\rho\rho} = o_p(1)$, $\Psi_{gh,n}^{\rho\rho} = O(1)$ and $\tilde{\Psi}_{gh,n}^{\rho\rho} = O_p(1)$ is analogous. Thus $\tilde{\Psi}_n^{\rho\rho} - \Psi_n^{\rho\rho} = o_p(1)$, $\Psi_n^{\rho\rho} = O(1)$ and $\tilde{\Psi}_n^{\rho\rho} = O_p(1)$. Observing that $\lambda_{\min}(\Psi_n^{\rho\rho}) \geq \text{const} > 0$ it follows further that $(\tilde{\Psi}_n^{\rho\rho})^{-1} - (\Psi_n^{\rho\rho})^{-1} = o_p(1)$, $(\Psi_n^{\rho\rho})^{-1} = O(1)$ and $(\tilde{\Psi}_n^{\rho\rho})^{-1} = O_p(1)$.

Next recall that in proving Theorem 3 we demonstrated that $\tilde{\boldsymbol{\alpha}}_{g,r,n} - \boldsymbol{\alpha}_{g,r,n} = o_p(1)$, $\boldsymbol{\alpha}_{g,r,n} = O(1)$ and $\tilde{\boldsymbol{\alpha}}_{g,r,n} = O_p(1)$, and furthermore that $\tilde{\sigma}_{gh,n} - \sigma_{gh} = o_p(1)$. Since $n^{-1}\tilde{\mathbf{T}}'_{gh,n}\tilde{\mathbf{T}}_{kl,n} - n^{-1}\mathbf{T}'_{gh,n}\mathbf{T}_{kl,n} = o_p(1)$ it follows that $\tilde{\Psi}_n^{\delta\delta} - \Psi_n^{\delta\delta} = o_p(1)$ and $\tilde{\Psi}_n^{\delta\rho} - \Psi_n^{\delta\rho} = o_p(1)$. Also observe that in light of Assumption 10 we have $\Psi_n^{\delta\delta} = O(1)$ and $\Psi_n^{\delta\rho} = O(1)$. This and the above results imply that $\tilde{\Psi}_n - \Psi_n = o_p(1)$, $\Psi_n = O(1)$ and $\tilde{\Psi}_n = O_p(1)$. Recalling that $\lambda_{\min}(\Psi_n) \geq \text{const} > 0$, it follows further that $\tilde{\Psi}_n^{-1} - \Psi_n^{-1} = o_p(1)$, $\Psi_n^{-1} = O(1)$ and $\tilde{\Psi}_n^{-1} = O_p(1)$.

By Assumption 9, $\tilde{\boldsymbol{\Upsilon}}_{g,n} - \boldsymbol{\Upsilon}_{g,n} = o_p(1)$, $\boldsymbol{\Upsilon}_{g,n} = O(1)$ and thus $\tilde{\boldsymbol{\Upsilon}}_{g,n} = O_p(1)$. Also recall from the proof of Theorem 3 that $\tilde{\mathbf{J}}_{g,n} - \mathbf{J}_{g,n} = o_p(1)$, $\mathbf{J}_{g,n} = O(1)$ and $\tilde{\mathbf{J}}_{g,n} = O_p(1)$, and furthermore that $(\tilde{\mathbf{J}}'_{g,n}\tilde{\boldsymbol{\Upsilon}}_{g,n}\tilde{\mathbf{J}}_{g,n})^+ - (\mathbf{J}'_{g,n}\boldsymbol{\Upsilon}_{g,n}\mathbf{J}_{g,n})^{-1} = o_p(1)$, $(\tilde{\mathbf{J}}'_{g,n}\tilde{\boldsymbol{\Upsilon}}_{g,n}\tilde{\mathbf{J}}_{g,n})^+ = O_p(1)$ and $(\mathbf{J}'_{g,n}\boldsymbol{\Upsilon}_{g,n}\mathbf{J}_{g,n})^{-1} = O(1)$. The claim that $\tilde{\boldsymbol{\Omega}}_n - \boldsymbol{\Omega}_n = o_p(1)$ and $\boldsymbol{\Omega}_n = O(1)$ is now obvious. Observing that

$$\lambda_{\min}(\boldsymbol{\Omega}_n) \geq \lambda_{\min}(\Psi_n)\lambda_{\min}\{\text{diag}_g[(\mathbf{J}'_{g,n}\boldsymbol{\Upsilon}_{g,n}\mathbf{J}_{g,n})^{-1}]\} \geq \text{const} > 0$$

utilizing that $\lambda_{\min}[\mathbf{J}'_{g,n}\boldsymbol{\Upsilon}_{g,n}\mathbf{J}_{g,n}]^{-1} \geq \text{const} > 0$ as demonstrated in the proof of Theorem 2, it follows further that $\tilde{\boldsymbol{\Omega}}_n^{-1} - \boldsymbol{\Omega}_n^{-1} = o_p(1)$, $\boldsymbol{\Omega}_n^{-1} = O(1)$ and $\tilde{\boldsymbol{\Omega}}_n^{-1} = O_p(1)$. \blacksquare

C Appendix: Proofs for Section 5

Lemma C.1 : *Suppose the assumptions of Theorem 5 hold. Let $\tilde{\boldsymbol{\rho}}_{g,n}$ be the initial GMM estimators defined in (30). Then $\tilde{\boldsymbol{\rho}}_{g,n} - \boldsymbol{\rho}_{g,n} = o_p(1)$.*

Proof. To prove the claim we verify the assumptions of Theorem 1. Assumptions 1-8 are maintained. Assumption 9 holds trivially with $\tilde{\Upsilon}_{g,n} = \Upsilon_{g,n} = \mathbf{I}$. Observing furthermore that by Lemma A.4 we have $n^{1/2}(\tilde{\delta}_{g,n} - \delta_{g,n}) = O_p(1)$ completes the proof. \blacksquare

Proof of Theorem 5: The proof of Theorem 5 is based on the generic limit theory developed in Theorem 4. In light of that theorem it proves convenient to first derive the limiting distribution of $\hat{\delta}_n = (\hat{\delta}'_{1,n}, \dots, \hat{\delta}'_{G,n})'$ and $\hat{\rho}_n = (\hat{\rho}'_{1,n}, \dots, \hat{\rho}'_{G,n})'$. The limiting distribution of $\hat{\delta}_{g,n}$ and $\hat{\rho}_{g,n}$ is then obtained as a trivial specialization. For clarity we divide the remainder of the proof into several parts.

Part 1: (Verification of Assumption 10 for $\hat{\delta}_{g,n}$) In light of Lemma A.5 we have $n^{1/2}[\hat{\delta}_{g,n} - \delta_{g,n}] = n^{-1/2}\mathbf{T}'_{gg,n}\boldsymbol{\varepsilon}_{g,n} + o_p(1)$ with

$$\mathbf{T}_{gg,n} = \mathbf{H}_n \mathbf{P}_{gg,n} \text{ with } \mathbf{P}_{gg,n} = \mathbf{Q}_{\mathbf{H}\mathbf{H}}^{-1} \mathbf{Q}_{\mathbf{H}\mathbf{Z},g^*}(\rho_{g,n}) [\mathbf{Q}'_{\mathbf{H}\mathbf{Z},g^*}(\rho_{g,n}) \mathbf{Q}_{\mathbf{H}\mathbf{H}}^{-1} \mathbf{Q}_{\mathbf{H}\mathbf{Z},g^*}(\rho_{g,n})]^{-1},$$

and thus

$$\begin{aligned} n^{-1}\mathbf{T}'_{gg,n} \mathbf{T}_{hh,n} &= [\mathbf{Q}'_{\mathbf{H}\mathbf{Z},g^*}(\rho_{g,n}) \mathbf{Q}_{\mathbf{H}\mathbf{H}}^{-1} \mathbf{Q}_{\mathbf{H}\mathbf{Z},g^*}(\rho_{g,n})]^{-1} \mathbf{Q}'_{\mathbf{H}\mathbf{Z},g^*}(\rho_{g,n}) \mathbf{Q}_{\mathbf{H}\mathbf{H}}^{-1} \\ &\quad \times (n^{-1}\mathbf{H}'_n \mathbf{H}_n) \mathbf{Q}_{\mathbf{H}\mathbf{H}}^{-1} \mathbf{Q}_{\mathbf{H}\mathbf{Z},h^*}(\rho_{h,n}) [\mathbf{Q}'_{\mathbf{H}\mathbf{Z},h^*}(\rho_{h,n}) \mathbf{Q}_{\mathbf{H}\mathbf{H}}^{-1} \mathbf{Q}_{\mathbf{H}\mathbf{Z},h^*}(\rho_{h,n})]^{-1}. \end{aligned}$$

The remaining conditions of Assumption 10 are also seen to hold in light of Lemma A.5.

Part 2: (Specialized Expressions for Ψ_n and the Corresponding Estimator) Next observe that the components of Ψ_n defined in (26) simplified in obvious ways in that $\mathbf{T}_n = \text{diag}_g(\mathbf{T}_{gg,n})$, and thus all terms involving a $\mathbf{T}_{gh,n}$ with $g \neq h$ are zero. In particular, the (g, h) -th blocks of $\Psi_n^{\delta\delta}$ and $\Psi_n^{\delta\rho}$ are given by

$$\begin{aligned} \Psi_{gh,n}^{\delta\delta} &= \sigma_{gh} n^{-1} \mathbf{T}'_{gg,n} \mathbf{T}_{hh,n}, \\ \Psi_{gh,n}^{\delta\rho} &= \sigma_{gh} n^{-1} \mathbf{T}'_{gg,n} \mathbf{T}_{hh,n} [\boldsymbol{\alpha}_{h,1,n}, \dots, \boldsymbol{\alpha}_{h,S,n}], \end{aligned}$$

and the elements of $\Psi_{gh,n}^{\rho\rho}$, i.e., the elements of the (g, h) -th block of $\Psi_n^{\rho\rho}$ are given by

$$\begin{aligned} \psi_{rs,gh,n}^{\rho\rho} &= \sigma_{gh,n}^2 (2n)^{-1} \text{tr} [(\mathbf{A}_{r,n} + \mathbf{A}'_{r,n}) (\mathbf{A}_{s,n} + \mathbf{A}'_{s,n})] \\ &\quad + \sigma_{gh,n} n^{-1} \boldsymbol{\alpha}'_{g,r,n} \mathbf{T}'_{gg,n} \mathbf{T}_{hh,n} \boldsymbol{\alpha}_{h,s,n}. \end{aligned}$$

Now consider the estimator

$$\begin{aligned}\widehat{\mathbf{T}}_{gg,n} &= \mathbf{H}_n(n^{-1}\mathbf{H}'_n\mathbf{H}_n)^{-1}(n^{-1}\mathbf{H}'_n\mathbf{Z}_{*g,n}(\tilde{\boldsymbol{\rho}}_{g,n})) \times \\ &\quad \left[(n^{-1}\mathbf{Z}'_{*g,n}(\tilde{\boldsymbol{\rho}}_{g,n})\mathbf{H}_n)(n^{-1}\mathbf{H}'_n\mathbf{H}_n)^{-1}(n^{-1}\mathbf{H}'_n\mathbf{Z}_{*g,n}(\tilde{\boldsymbol{\rho}}_{g,n})) \right]^{-1},\end{aligned}$$

where $\tilde{\boldsymbol{\rho}}_{g,n}$ denotes the first-stage estimator for $\boldsymbol{\rho}_{g,n}$, let

$$\widehat{\boldsymbol{\alpha}}_{g,r,n} = -n^{-1} \left[\mathbf{Z}'_{g,n}(\mathbf{I}_n - \mathbf{R}_{g,n}^*(\tilde{\boldsymbol{\rho}}_{g,n})) (\mathbf{A}_{r,n} + \mathbf{A}'_{r,n}) (\mathbf{I}_n - \mathbf{R}_{g,n}^*(\tilde{\boldsymbol{\rho}}_{g,n})) \widehat{\mathbf{u}}_{g,n} \right]$$

with $\widehat{\mathbf{u}}_{g,n} = \mathbf{y}_{g,n} - \mathbf{Z}_{g,n}\widehat{\boldsymbol{\delta}}_{g,n}$, and let $\widehat{\sigma}_{gh,n} = n^{-1}\widehat{\boldsymbol{\varepsilon}}'_{g,n}\widehat{\boldsymbol{\varepsilon}}_{h,n}$ with $\widehat{\boldsymbol{\varepsilon}}_{g,n} = \mathbf{y}_{*g,n}(\tilde{\boldsymbol{\rho}}_{g,n}) - \mathbf{Z}_{*g,n}(\tilde{\boldsymbol{\rho}}_{g,n})\widehat{\boldsymbol{\delta}}_{g,n}$. Then the components of the estimator for $\boldsymbol{\Psi}_n$ defined in (27) simplify to

$$\begin{aligned}\widehat{\boldsymbol{\Psi}}_{gh,n}^{\delta\delta} &= \widehat{\sigma}_{gh}n^{-1}\widehat{\mathbf{T}}'_{gg,n}\widehat{\mathbf{T}}_{hh,n}, \\ \widehat{\boldsymbol{\Psi}}_{gh,n}^{\delta\rho} &= \widehat{\sigma}_{gh}n^{-1}\widehat{\mathbf{T}}'_{gg,n}\widehat{\mathbf{T}}_{hh,n} [\widehat{\boldsymbol{\alpha}}_{h,1,n}, \dots, \widehat{\boldsymbol{\alpha}}_{h,S,n}],\end{aligned}$$

and the elements of the estimator $\widehat{\boldsymbol{\Psi}}_{gh,n}^{\rho\rho}$ of $\boldsymbol{\Psi}_{gh,n}^{\rho\rho}$ are given by

$$\begin{aligned}\widehat{\psi}_{rs,gh,n}^{\rho\rho} &= \widehat{\sigma}_{gh,n}^2(2n)^{-1}tr \left[(\mathbf{A}_{r,n} + \mathbf{A}'_{r,n}) (\mathbf{A}_{s,n} + \mathbf{A}'_{s,n}) \right] \\ &\quad + \widehat{\sigma}_{gh,n}n^{-1}\widehat{\boldsymbol{\alpha}}'_{g,r,n}\widehat{\mathbf{T}}'_{gg,n}\widehat{\mathbf{T}}_{hh,n}\widehat{\boldsymbol{\alpha}}_{h,s,n}.\end{aligned}$$

Note that

$$\begin{aligned}n^{-1}\widehat{\mathbf{T}}'_{gg,n}\widehat{\mathbf{T}}_{hh,n} &= \left[n^{-1}\widehat{\mathbf{Z}}'_{*g,n}(\tilde{\boldsymbol{\rho}}_{g,n})\widehat{\mathbf{Z}}_{*g,n}(\tilde{\boldsymbol{\rho}}_{g,n}) \right]^{-1} n^{-1}\widehat{\mathbf{Z}}'_{*g,n}(\tilde{\boldsymbol{\rho}}_{g,n})\widehat{\mathbf{Z}}_{*h,n}(\tilde{\boldsymbol{\rho}}_{h,n}) \\ &\quad \times \left[n^{-1}\widehat{\mathbf{Z}}'_{*h,n}(\tilde{\boldsymbol{\rho}}_{h,n})\widehat{\mathbf{Z}}_{*h,n}(\tilde{\boldsymbol{\rho}}_{h,n}) \right]^{-1},\end{aligned}$$

and

$$n^{-1}\widehat{\mathbf{T}}'_{gg,n}\widehat{\mathbf{T}}_{gg,n} = \left[n^{-1}\widehat{\mathbf{Z}}'_{*g,n}(\tilde{\boldsymbol{\rho}}_{g,n})\widehat{\mathbf{Z}}_{*g,n}(\tilde{\boldsymbol{\rho}}_{g,n}) \right]^{-1} = \left[n^{-1}\widehat{\mathbf{Z}}'_{*g,n}(\tilde{\boldsymbol{\rho}}_{g,n})\mathbf{Z}_{*g,n}(\tilde{\boldsymbol{\rho}}_{g,n}) \right]^{-1}.$$

Also note that in light of Lemma A.5 we have $n^{-1}\widehat{\mathbf{T}}'_{gg,n}\widehat{\mathbf{T}}_{hh,n} - n^{-1}\mathbf{T}'_{gg,n}\mathbf{T}_{hh,n} = o_p(1)$.

Part 3: (Verification of Assumption 9 for $\boldsymbol{\Upsilon}_{g,n} = (\boldsymbol{\Psi}_{gg,n}^{\rho\rho})^{-1}$ and $\widetilde{\boldsymbol{\Upsilon}}_{g,n} = (\widehat{\boldsymbol{\Psi}}_{gg,n}^{\rho\rho})^{-1}$ with $\widehat{\boldsymbol{\Psi}}_{gg,n}^{\rho\rho} = (\widehat{\psi}_{rs,gg,n}^{\rho\rho})$). Observe that the assumption that $\lambda_{\min}(\boldsymbol{\Psi}_n) \geq c$ for some $c > 0$ implies that also $\lambda_{\min}(\boldsymbol{\Psi}_{gg,n}^{\rho\rho}) \geq c$. Recall furthermore that by Lemma C.1 we have $\tilde{\boldsymbol{\rho}}_{g,n} - \boldsymbol{\rho}_{g,n} = o_p(1)$. It now follows directly from Theorem 3 that $\widehat{\boldsymbol{\Psi}}_{gg,n}^{\rho\rho} - \boldsymbol{\Psi}_{gg,n}^{\rho\rho} = o_p(1)$, $(\widehat{\boldsymbol{\Psi}}_{gg,n}^{\rho\rho})^{-1} - (\boldsymbol{\Psi}_{gg,n}^{\rho\rho})^{-1} = o_p(1)$, and $\boldsymbol{\Psi}_{gg,n}^{\rho\rho} = O(1)$,

$(\Psi_{gg,n}^{\rho\rho})^{-1} = O(1)$, which verifies Assumption 9.

Part 4: (Limiting Distribution of $\widehat{\delta}_{g,n}$ and $\widehat{\rho}_{g,n}$) Recall that Assumptions 1-8 are maintained. Thus, in light of the above discussion, all assumptions of Theorem 4 are satisfied. Next observe that since $\mathbf{T}_{gh,n} = 0$ for $g \neq h$ the expression for Ω_n given in (28) simplify to:

$$\Omega_n = \begin{bmatrix} \Omega_n^{\delta\delta} & \Omega_n^{\delta\rho} \\ \Omega_n^{\delta\rho'} & \Omega_n^{\rho\rho} \end{bmatrix}$$

with

$$\begin{aligned} \Omega_n^{\delta\delta} &= \Psi_n^{\delta\delta}, \\ \Omega_n^{\delta\rho} &= \Psi_n^{\delta\rho} \text{diag}_{g=1}^G ((\Psi_{gg,n}^{\rho\rho})^{-1} \mathbf{J}_{g,n} (\mathbf{J}'_{g,n} (\Psi_{gg,n}^{\rho\rho})^{-1} \mathbf{J}_{g,n})^{-1}), \\ \Omega_n^{\rho\rho} &= \text{diag}_{g=1}^G ((\mathbf{J}'_{g,n} (\Psi_{gg,n}^{\rho\rho})^{-1} \mathbf{J}_{g,n})^{-1} \mathbf{J}'_{g,n} (\Psi_{gg,n}^{\rho\rho})^{-1}) \begin{bmatrix} \Psi_{11,n}^{\rho\rho} & \cdots & \Psi_{1G,n}^{\rho\rho} \\ \vdots & \ddots & \vdots \\ \Psi_{G1,n}^{\rho\rho} & \cdots & \Psi_{GG,n}^{\rho\rho} \end{bmatrix} \\ &\quad \times \text{diag}_{g=1}^G ((\Psi_{gg,n}^{\rho\rho})^{-1} \mathbf{J}_{g,n} (\mathbf{J}'_{g,n} (\Psi_{gg,n}^{\rho\rho})^{-1} \mathbf{J}_{g,n})^{-1}). \end{aligned}$$

By Theorem 4, $n^{1/2}[(\widehat{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_n)', (\widehat{\boldsymbol{\rho}}_n - \boldsymbol{\rho}_n)']' \xrightarrow{d} N(\mathbf{0}, \Omega_n)$, and as a specialization,

$$n^{1/2} \begin{bmatrix} \widehat{\boldsymbol{\delta}}_{g,n} - \boldsymbol{\delta}_{g,n} \\ \widehat{\boldsymbol{\rho}}_{g,n} - \boldsymbol{\rho}_{g,n} \end{bmatrix} \xrightarrow{d} N \begin{bmatrix} \Omega_{gg,n}^{\delta\delta} & \Omega_{gg,n}^{\delta\rho} \\ \Omega_{gg,n}^{\delta\rho'} & \Omega_{gg,n}^{\rho\rho} \end{bmatrix} \quad (\text{C.1})$$

with

$$\begin{aligned} \Omega_{gg,n}^{\delta\delta} &= \Psi_{gg,n}^{\delta\delta} = \sigma_{gg} n^{-1} \mathbf{T}'_{gg,n} \mathbf{T}_{gg,n}, \\ \Omega_{gg,n}^{\delta\rho} &= \Psi_{gg,n}^{\delta\rho} (\Psi_{gg,n}^{\rho\rho})^{-1} \mathbf{J}_{g,n} (\mathbf{J}'_{g,n} (\Psi_{gg,n}^{\rho\rho})^{-1} \mathbf{J}_{g,n})^{-1}, \\ &= \sigma_{gg} n^{-1} \mathbf{T}'_{gg,n} \mathbf{T}_{gg,n} [\boldsymbol{\alpha}_{g,1,n}, \dots, \boldsymbol{\alpha}_{g,S,n}] (\Psi_{gg,n}^{\rho\rho})^{-1} \mathbf{J}_{g,n} (\mathbf{J}'_{g,n} (\Psi_{gg,n}^{\rho\rho})^{-1} \mathbf{J}_{g,n})^{-1}, \\ \Omega_{gg,n}^{\rho\rho} &= (\mathbf{J}'_{g,n} (\Psi_{gg,n}^{\rho\rho})^{-1} \mathbf{J}_{g,n})^{-1}. \end{aligned}$$

Observe from part 3 of the proof that

$$n^{-1} \widehat{\mathbf{T}}'_{gg,n} \widehat{\mathbf{T}}_{gg,n} = \left[n^{-1} \widehat{\mathbf{Z}}'_{*g,n} (\widehat{\boldsymbol{\rho}}_{g,n}) \widehat{\mathbf{Z}}_{*g,n} (\widehat{\boldsymbol{\rho}}_{g,n}) \right]^{-1}$$

and that $n^{-1} \widehat{\mathbf{T}}'_{gg,n} \widehat{\mathbf{T}}_{gg,n} - n^{-1} \mathbf{T}'_{gg,n} \mathbf{T}_{gg,n} = o_p(1)$. The asymptotic normality result of the theorem now follows immediately from (C.1), observing that

Theorem 4 also established the consistency of the VC estimators. \blacksquare

Proof of Theorem 6: The proof is again based on the generic limit theory developed in Theorem 4. For clarity we divide the proof, analogous to the proof of Theorem 5, into several parts.

Part 1: (Verification of Assumption 10 for $\widehat{\boldsymbol{\delta}}_n$) In light of Lemma A.6 we have $n^{1/2}[\widehat{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_n] = n^{-1/2}\mathbf{T}'_n\boldsymbol{\varepsilon}_n + o_p(1)$ with

$$\begin{aligned}\mathbf{T}_n &= (\mathbf{I}_G \otimes \mathbf{H}_n)\mathbf{P}_n, \\ \mathbf{P}_n &= [\boldsymbol{\Sigma}^{-1} \otimes \mathbf{Q}_{\mathbf{H}\mathbf{H}}^{-1}] \text{diag}_{g=1}^G [\mathbf{Q}_{\mathbf{H}\mathbf{Z},g^*}(\boldsymbol{\rho}_{g,n})] \\ &\quad \times \{ \text{diag}_{g=1}^G [\mathbf{Q}'_{\mathbf{H}\mathbf{Z},g^*}(\boldsymbol{\rho}_{g,n})] [\boldsymbol{\Sigma}^{-1} \otimes \mathbf{Q}_{\mathbf{H}\mathbf{H}}^{-1}] \text{diag}_{g=1}^G [\mathbf{Q}_{\mathbf{H}\mathbf{Z},g^*}(\boldsymbol{\rho}_{g,n})] \}^{-1}.\end{aligned}\tag{C.2}$$

Now let $\mathbf{T}_{gh,n}$ and $\mathbf{P}_{gh,n}$ denote the (g, h) -th block of \mathbf{T}_n and \mathbf{P}_n , then $\mathbf{T}_{gh,n} = \mathbf{F}_{gh,n}\mathbf{P}_{gh,n}$ with $\mathbf{F}_{gh,n} = \mathbf{H}_n$. The remaining conditions of Assumption 10 are then seen to hold in light of Lemma A.6.

Part 2: (Specialized Expressions for $\boldsymbol{\Psi}_n$ and the Corresponding Estimator) Specialized expressions for each of the submatrices $\boldsymbol{\Psi}_n^{\delta\delta}$, $\boldsymbol{\Psi}_n^{\delta\rho}$, $\boldsymbol{\Psi}_n^{\rho\rho}$ of $\boldsymbol{\Psi}_n$ defined in (26) are readily found by substituting into those expressions the formulae for \mathbf{T}_n given in (C.2), and by observing that $\sum_{u=1}^G \sum_{v=1}^G \sigma_{uv,n} \mathbf{T}'_{gu,n} \mathbf{T}_{hv,n}$ represents the (g, h) -th block of $\boldsymbol{\Psi}_n^{\delta\delta} = \mathbf{T}'_n(\boldsymbol{\Sigma} \otimes \mathbf{I})\mathbf{T}_n$.

Next let

$$\begin{aligned}\widehat{\mathbf{T}}_n &= (\mathbf{I}_G \otimes \mathbf{H}_n)\widehat{\mathbf{P}}_n, \\ \widehat{\mathbf{P}}_n &= [\widehat{\boldsymbol{\Sigma}}_n^{-1} \otimes (n^{-1}\mathbf{H}'_n\mathbf{H}_n)^{-1}] \text{diag} [n^{-1}\mathbf{H}'_n\mathbf{Z}_{*g,n}(\widehat{\boldsymbol{\rho}}_{g,n})] \\ &\quad \times [n^{-1}\widehat{\mathbf{Z}}'_{*n}(\widehat{\boldsymbol{\rho}}_n)(\widehat{\boldsymbol{\Sigma}}_n^{-1} \otimes \mathbf{I}_n)\mathbf{Z}_{*n}(\widehat{\boldsymbol{\rho}}_n)]^{-1}.\end{aligned}$$

Let $\widehat{\mathbf{T}}_{gh,n}$ and $\widehat{\mathbf{P}}_{gh,n}$ denote the (g, h) -th block of $\widehat{\mathbf{T}}_n$ and $\widehat{\mathbf{P}}_n$, respectively. Then, clearly, $\widehat{\mathbf{T}}_{gh,n} = \mathbf{H}_n\widehat{\mathbf{P}}_{gh,n}$. Next observe that

$$\widehat{\boldsymbol{\Psi}}_n^{\delta\delta} = [n^{-1}\widehat{\mathbf{Z}}'_{*n}(\widehat{\boldsymbol{\rho}}_n)(\widehat{\boldsymbol{\Sigma}}_n^{-1} \otimes \mathbf{I}_n)\mathbf{Z}_{*n}(\widehat{\boldsymbol{\rho}}_n)]^{-1} = n^{-1}\widehat{\mathbf{T}}_n'(\widehat{\boldsymbol{\Sigma}}_n \otimes \mathbf{I}_n)\widehat{\mathbf{T}}_n,$$

and thus the (g, h) -th block of $\widehat{\boldsymbol{\Psi}}_n^{\delta\delta}$ is given by $\widehat{\boldsymbol{\Psi}}_{gh,n}^{\delta\delta} = \sum_{u=1}^G \sum_{v=1}^G \widehat{\sigma}_{uv,n} \widehat{\mathbf{T}}'_{gu,n} \widehat{\mathbf{T}}_{hv,n}$. From this we see that the estimators $\widehat{\boldsymbol{\Psi}}_n^{\delta\delta}$, $\widehat{\boldsymbol{\Psi}}_n^{\delta\rho}$, $\widehat{\boldsymbol{\Psi}}_n^{\rho\rho}$, $\widehat{\boldsymbol{\Omega}}_n^{\delta\delta}$, $\widehat{\boldsymbol{\Omega}}_n^{\delta\rho}$, $\widehat{\boldsymbol{\Omega}}_n^{\rho\rho}$ are special

cases of the class of VC estimators considered by Theorem 4. Also note that in light of Lemma A.6 we have that

$$n^{-1} \widehat{\mathbf{T}}'_{gh,n} \widehat{\mathbf{T}}_{gl,n} - \mathbf{T}'_{gh,n} \mathbf{T}_{gl,n} = o_p(1),$$

which verifies that also this condition of Theorem 4 holds.

Part 3: (Verification of Assumption 9 for $\Upsilon_{g,n} = (\Psi_{gg,n}^{\rho\rho})^{-1}$ and $\widetilde{\Upsilon}_{g,n} = (\widehat{\Psi}_{gg,n}^{\rho\rho})^{-1}$ with $\widehat{\Psi}_{gg,n}^{\rho\rho} = (\widehat{\psi}_{rs,gg,n}^{\rho\rho})$) Observe that the assumption that $\lambda_{\min}(\Psi_n) \geq c$ for some $c > 0$ implies that also $\lambda_{\min}(\Psi_{gg,n}^{\rho\rho}) \geq c$. Recall furthermore that by Lemma C.1 we have $\widehat{\rho}_{g,n} - \rho_{g,n} = o_p(1)$. It now follows directly from Theorem 3 that $\widehat{\Psi}_{gg,n}^{\rho\rho} - \Psi_{gg,n}^{\rho\rho} = o_p(1)$, $(\widehat{\Psi}_{gg,n}^{\rho\rho})^{-1} - (\Psi_{gg,n}^{\rho\rho})^{-1} = o_p(1)$, and $\Psi_{gg,n}^{\rho\rho} = O(1)$, $(\Psi_{gg,n}^{\rho\rho})^{-1} = O(1)$, which verifies Assumption 9.

Part 4: (Limiting Distribution of $\widehat{\delta}_{g,n}$ and $\widehat{\rho}_{g,n}$) Recall that Assumptions 1-8 are maintained. Thus, in light of the above discussion, all assumptions of Theorem 4 are satisfied. It thus follows from that theorem that

$$\begin{bmatrix} \widehat{\delta}_n - \delta_n \\ \widehat{\rho}_n - \rho_n \end{bmatrix} \xrightarrow{d} N \begin{bmatrix} \Omega_n^{\delta\delta} & \Omega_n^{\delta\rho} \\ \Omega_n^{\delta\rho'} & \Omega_n^{\rho\rho} \end{bmatrix}, \quad (\text{C.3})$$

where

$$\Omega_n^{\delta\delta} = \Psi_n^{\delta\delta}, \quad \Omega_n^{\delta\rho} = \Psi_n^{\delta\rho} \text{diag}_{g=1}^G(\mathfrak{J}_{g,n}), \quad \Omega_n^{\rho\rho} = \text{diag}_{g=1}^G(\mathfrak{J}'_{g,n}) \Psi_n^{\rho\rho} \text{diag}_{g=1}^G(\mathfrak{J}_{g,n})$$

with

$$\begin{aligned} \Psi_n^{\delta\rho} &= \Psi_n^{\delta\delta} \text{diag}_{g=1}^G[\alpha_{g,1,n}, \dots, \alpha_{g,S,n}], \\ \mathfrak{J}_{g,n} &= (\Psi_n^{\rho\rho})^{-1} \mathbf{J}_{g,n} [\mathbf{J}'_{g,n} (\Psi_{gg,n}^{\rho\rho})^{-1} \mathbf{J}_{g,n}]^{-1}. \end{aligned}$$

The asymptotic normality result of the theorem now follows immediately from (C.3), observing that Theorem 4 also establishes the consistency of the VC estimators. \blacksquare

Online Supplementary Material

Drukker, D.M., P.H. Egger, and I.R. Prucha (2021): Supplement to “Simultaneous Equations Models with Higher-Order Spatial or Social Network Interactions”, *Econometric Theory* Supplementary Material. To view, please visit: [[doi will be inserted here by typesetter]]

References

- [1] Anselin, L. (1988) *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers.
- [2] Anselin, L. (2010) Thirty years of spatial econometrics. *Papers in Regional Science* 89, 3-25.
- [3] Anselin, L. and R. Florax (1995) *New Directions in Spatial Econometrics*. Springer.
- [4] Audretsch, D.B. and M.P. Feldmann (1996) R&D spillovers and the geography of innovation and production. *American Economic Review* 86, 630-640.
- [5] Ballester, C., A. Calvó-Armengol and Y. Zenou (2006) Who’s who in networks. Wanted: The key player. *Econometrica* 74, 1403-1417.
- [6] Baltagi, B.H. and Y. Deng (2015) EC3SLS estimator for a simultaneous system of spatial autoregressive equations with random effects. *Econometric Reviews* 34, 659-694.
- [7] Baltagi, B.H., P. Egger and M. Pfaffermayr (2007) Estimating models of complex FDI: Are there third-country effects? *Journal of Econometrics* 140, 260-281.
- [8] Baltagi, B.H., P. Egger and M. Pfaffermayr (2008) Estimating regional trade agreement effects on FDI in an interdependent world. *Journal of Econometrics* 145, 194-208.
- [9] Behrens, K., W. Koch and C. Ertur (2012) Dual gravity: Using spatial econometrics to control for multilateral resistance. *Journal of Applied Econometrics* 27, 773-794.

- [10] Bell, K.P. and N.E. Bockstael (2000) Applying the Generalized-Moments estimation approach to spatial problems involving microlevel data. *Review of Economics and Statistics* 82, 72-82.
- [11] Blonigen, B.A., R.B. Davies, G.R. Waddell and H.T. Naughton (2007) FDI in space: Spatial autoregressive relationships in foreign direct investment. *European Economic Review* 51, 1303-1325.
- [12] Blume, L.E., W.A. Brock, S.N. Durlauf and Y.M. Ioannides (2011) Identification of social interactions, In: J. Benhabib, A. Bisin, and M.O. Jackson (eds.), *Handbook of Social Economics*, Vol. 1B, pp. 853-964. Elsevier Publisher.
- [13] Calvó-Armengol, A., E. Patacchini and Y. Zenou (2009) Peer effects and social networks in education. *Review of Economic Studies* 76, 1239-1267.
- [14] Case, A., J. Hines Jr. and H. Rosen (1993) Budget spillovers and fiscal policy independence: Evidence from the States. *Journal of Public Economics* 52, 285-307.
- [15] Das, D., H.H. Kelejian and I.R. Prucha (2003) Finite sample properties of estimators of spatial autoregressive models with autoregressive disturbances. *Papers in Regional Science* 82, 1-26.
- [16] Dhrymes, P.J. (1978) *Introductory Econometrics*. Springer Verlag.
- [17] Ertur, C. and W. Koch (2007) Growth, technological interdependence and spatial externalities: Theory and evidence. *Journal of Applied Econometrics* 22, 1033-1062.
- [18] Cliff, A. and J. Ord (1973) *Spatial Autocorrelation*. Pion.
- [19] Cliff, A. and J. Ord (1981) *Spatial Processes, Models and Applications*. Pion.
- [20] Cohen, J.P. and C.J. Morrison Paul (2007) Higher order spatial autocorrelation in a system of equations: The impacts of transportation infrastructure on capital asset values. *Journal of Regional Science* 47, 457-478.
- [21] Cohen-Cole, E., X. Liu and Y. Zenou (2018) Multivariate choice and identification of social interactions. *Journal of Applied Econometrics* 33, 165-178.

- [22] Conley, T.G. and E. Ligon (2002) Economic distance and cross-country spillovers. *Journal of Economic Growth* 7, 157-187.
- [23] Devereux, M.P., B. Lockwood and M. Redoano (2008) Do countries compete over corporate taxes? *Journal of Public Economics* 92, 1210-1235.
- [24] Egger, P., M. Pfaffermayr and H. Winner (2005) Commodity taxation in a 'linear' world: A spatial panel data approach. *Regional Science and Urban Economics* 35, 527-541.
- [25] Holtz-Eakin, D. (1994) Public sector capital and the productivity puzzle. *Review of Economics and Statistics* 76, 12-21.
- [26] Horn, R.A. and C.R. Johnson (1985) *Matrix Analysis*. Cambridge University Press.
- [27] Kapoor, M., H.H. Kelejian and I.R. Prucha (2007) Panel data models with spatially correlated error components. *Journal of Econometrics* 140, 97-130.
- [28] Kelejian, H.H. and I.R. Prucha (1998) A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *Journal of Real Estate Finance and Economics* 17, 99-121.
- [29] Kelejian, H.H. and I.R. Prucha (1999) A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review* 40, 509-533.
- [30] Kelejian, H.H. and I.R. Prucha (2004) Estimation of simultaneous systems of spatially interrelated cross sectional equations. *Journal of Econometrics* 118, 27-50.
- [31] Kelejian, H.H. and I.R. Prucha (2007) HAC estimation in a spatial framework. *Journal of Econometrics* 140, 131-154.
- [32] Kelejian, H.H. and I.R. Prucha (2010) Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics* 157, 53-67.
- [33] Kuersteiner, G.M. and I.R. Prucha (2020) Dynamic spatial panel models: Networks, common shocks, and sequential exogeneity. *Econometrica* 88, 2109-2146.

- [34] Lee, J., P.C.B. Phillips and F. Rossi (2021) Specification testing in spatial autoregression. Working Paper.
- [35] Lee, L.-F. (2003) Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. *Econometric Reviews* 22, 307-335.
- [36] Lee, L.-F. (2007a) GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics* 137, 489-514.
- [37] Lee, L.-F. (2007b) Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. *Journal of Econometrics* 140, 333-374.
- [38] Lee, L.-F. and X. Liu (2010) Efficient GMM estimation of high order spatial autoregressive models with autoregressive disturbances. *Econometric Theory* 26, 187-230.
- [39] Liu, X. (2014) Identification and efficient estimation of simultaneous equations network models. *Journal of Business & Economic Statistics* 32, 516-536.
- [40] Liu, X. (2019) Simultaneous equations with binary outcomes and social interactions. *Econometric Reviews* 38, 921-937.
- [41] Liu, X. (2020) GMM identification and estimation of peer effects in a system of simultaneous equations. *Journal of Spatial Econometrics* 1, 2-27.
- [42] Liu, X. and P. Saraiva (2019) GMM estimation of spatial autoregressive models in a system of simultaneous equations with heteroskedasticity. *Econometric Reviews* 38, 359-385.
- [43] Livanis, G., C.B. Moss, V.C. Breneman and R.F. Nehring (2006) Urban sprawl and farmland prices. *American Journal of Agricultural Economics* 88, 915-929.
- [44] Pinkse, J. and M.E. Slade (1998) Contracting in space: An application of spatial statistics to discrete-choice models. *Journal of Econometrics* 85, 125-154.
- [45] Pinkse, J., M.E. Slade and C. Brett (2002) Spatial price competition: A semiparametric approach. *Econometrica* 70, 1111-1153.

- [46] Pötscher, B.M. and I.R. Prucha (1997) *Dynamic Nonlinear Econometric Models, Asymptotic Theory*. Springer Verlag.
- [47] Wang, W., L.-F. Lee and Y. Bao (2018) GMM estimation of the spatial autoregressive model in a system of interrelated networks. *Regional Science and Urban Economics* 69, 167-198.
- [48] Wang, L., K. Li and Z. Wang (2014) Quasi maximum likelihood estimation for simultaneous spatial autoregressive models. MPRA Working Paper 59901.
- [49] Whittle, P. (1954) On stationary processes in the plane. *Biometrika* 41, 434-449.
- [50] Yang, K. and L.-F. Lee (2017) Identification and QML estimation of multivariate and simultaneous equations spatial autoregressive models. *Journal of Econometrics* 196, 196-214.
- [51] Yang, K. and L.-F. Lee (2019) Identification and estimation of spatial dynamic panel simultaneous equations models. *Regional Science and Urban Economics* 76, 32-46.