# SIMULTANEOUS EQUATIONS MODELS WITH HIGHER-ORDER SPATIAL OR SOCIAL NETWORK INTERACTIONS

DAVID M. DRUKKER
*Department of Economics and International Business, Sam Houston State University*

PETER H. EGGER
*ETH Zurich*

INGMAR R. PRUCHA
*Department of Economics, University of Maryland*

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 estimation approach introduced in Kelejian and Prucha (1998, *Journal of Real Estate Finance and Economics* 17, 99–121; 1999, *International Economic Review* 40, 509–533; 2004, *Journal of Econometrics* 118, 27–50). 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.

## 1. INTRODUCTION

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

**1154**

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.[1]

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.[2]

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.[3] Most of the literature focused on single-equation models where a single dependent variable, say, $y_i$, is determined for units $i = 1, \ldots, n$. However, in economics, it is frequent that the outcomes for several dependent variables, say, $y_{i1}, \ldots, y_{iG}$, are determined jointly by a system of equations for units $i = 1, \ldots, n$. 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

---

[1] For 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 Jr., and Rosen (1993) report on spatial budget spillovers. The results in Behrens, Koch, and Ertur (2012) suggest that bilateral trade flows exhibit spatial interdependence, and Baltagi, Egger, and Pfaffermayr (2007), Baltagi, Egger, and Pfaffermayr, 2008 and Blonigen et al. (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 (2007); Calvó-Armengol, Patacchini, and Zenou (2009); Blume et al. (2011); Cohen-Cole, Liu, and Zenou (2018) and Liu (2014).

[2] As 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 dependence is molded via random Markov fields and conditional autoregressive models; see, e.g., Cressie (1993).

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

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 multiequation 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 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., although 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.[4] 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 multiple 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 Appendixes 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

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

If $\|\mathbf{A}_n\|_1 \le c$ and $\|\mathbf{A}_n\|_\infty \le c$ 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

---

[4]The 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.

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 $\mathbf{A}_n$, $\|\mathbf{A}_n\| = [tr(\mathbf{A}'_n\mathbf{A}_n)]^{1/2}$, where $tr$ denotes the trace operator. Let $\mathbf{A}_{n,g}$, $g = 1, \ldots G$, 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.[5] 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 ($g = 1, \ldots, G$):

$$\mathbf{y}_{g,n} = \sum_{l=1}^G b_{lg,n}\mathbf{y}_{l,n} + \sum_{k=1}^K c_{kg,n}\mathbf{x}_{k,n} + \sum_{l=1}^G \left[\sum_{s=1}^p \lambda_{lg,s,n}\mathbf{W}_{s,n}\right]\mathbf{y}_{l,n} + \mathbf{u}_{g,n},$$

$$\mathbf{u}_{g,n} = \left[\sum_{r=1}^q \rho_{g,r,n}\mathbf{M}_{r,n}\right]\mathbf{u}_{g,n} + \boldsymbol{\varepsilon}_{g,n}, \tag{1}$$

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,[6] $\mathbf{u}_{g,n}$ 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 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,

---

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

[6] In treating the exogenous variables as nonstochastic, the analysis may be viewed as conditional on the 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

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

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 Section 1, 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.[7] 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 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 $\xi_j$ 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.[8] 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

---

[7] The 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.

[8] Extensions of the estimation methodology will be discussed later.

autoregressive model of order $p$ with spatially autoregressive disturbances of order $q$, for short, a simultaneous SARAR(p,q) model.[9] 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

$$
\begin{aligned}
\mathbf{Y}_n &= \mathbf{Y}_n \mathbf{B}_n + \mathbf{X}_n \mathbf{C}_n + \overline{\mathbf{Y}}_n \mathbf{\Lambda}_n + \mathbf{U}_n, \\
\mathbf{U}_n &= \overline{\mathbf{U}}_n \mathbf{R}_n + \mathbf{E}_n
\end{aligned}
\tag{2}
$$

with

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

and where the parameter matrices $\mathbf{B}_n = (b_{lg,n})_{G \times G}$, $\mathbf{C}_n = (c_{lg,n})_{K \times G}$, $\mathbf{\Lambda}_n = (\lambda_{lg,s,n})_{pG \times G}$, and $\mathbf{R}_n = (\rho_{g,r,n})_{qG \times G}$ are defined conformably.[10]

## 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

---

[9]For 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).

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

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[11]:

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

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

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

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.[12]

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 $y_{i1}, \ldots, y_{iG}$, 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

---

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

[12] The 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.

$s$th group of peers, and assume that the utility of individual $i$ is of the following linear-quadratic form:

$$u(y_{i1}, \ldots, y_{iG}) = v(y_{i1}, \ldots, y_{iG}, \bar{y}_{i1,1}, \ldots \bar{y}_{i1,p}, \ldots, \bar{y}_{iG,1}, \ldots, \bar{y}_{iG,p})$$

$$= \underbrace{\sum_{g=1}^{G} \pi_{ig}^* y_{ig} + \sum_{g=1}^{G} y_{ig} \sum_{l=1}^{G} \sum_{s=1}^{p} \lambda_{lg,s}^* \bar{y}_{il,s}}_{payoff} \quad \underbrace{-\frac{1}{2} \sum_{g=1}^{G} b_g^* y_{ig}^2 - \sum_{g=1}^{G} \sum_{l=1, l \neq g}^{G} b_{lg}^* y_{ig} y_{il}}_{cost}. \tag{3}$$

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

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

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

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

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

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

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 $i = 1, \ldots, n$ 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

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

and let

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

Observing that $vec(\bar{\mathbf{Y}}_n) = (\mathbf{I}_G \otimes \mathbf{W}_n) \mathbf{y}_n$ and $vec(\bar{\mathbf{U}}_n) = (\mathbf{I}_G \otimes \mathbf{M}_n) \mathbf{u}_n$, and that $vec(\mathbf{A}_1 \mathbf{A}_2) = (\mathbf{A}_2' \otimes \mathbf{I}) vec(\mathbf{A}_1)$ 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

$$\mathbf{y}_n = \mathbf{B}_n^* \mathbf{y}_n + \mathbf{C}_n^* \mathbf{x}_n + \mathbf{u}_n,$$
$$\mathbf{u}_n = \mathbf{R}_n^* \mathbf{u}_n + \boldsymbol{\varepsilon}_n, \tag{6}$$

where $\mathbf{B}_n^* = \left[ (\mathbf{B}_n' \otimes \mathbf{I}_n) + (\boldsymbol{\Lambda}_n' \otimes \mathbf{I}_n)(\mathbf{I}_G \otimes \mathbf{W}_n) \right]$, $\mathbf{C}_n^* = (\mathbf{C}_n' \otimes \mathbf{I}_n)$, 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

$$\mathbf{y}_n = \left( \mathbf{I}_{nG} - \mathbf{B}_n^* \right)^{-1} \left[ \mathbf{C}_n^* \mathbf{x}_n + \mathbf{u}_n \right],$$
$$\mathbf{u}_n = \left( \mathbf{I}_{nG} - \mathbf{R}_n^* \right)^{-1} \boldsymbol{\varepsilon}_n. \tag{7}$$

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 nonzero elements of the $g$th column of $\mathbf{B}_n$, $\mathbf{C}_n$, $\boldsymbol{\Lambda}_n$, and $\mathbf{R}_n$, respectively, and let $\mathbf{Y}_{g,n}$, $\mathbf{X}_{g,n}$, $\overline{\mathbf{Y}}_{g,n}$, and $\overline{\mathbf{U}}_{g,n}$ 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 ($g = 1, \ldots, G$):

$$\mathbf{y}_{g,n} = \mathbf{Z}_{g,n} \boldsymbol{\delta}_{g,n} + \mathbf{u}_{g,n},$$
$$\mathbf{u}_{g,n} = \overline{\mathbf{U}}_{g,n} \boldsymbol{\rho}_{g,n} + \boldsymbol{\varepsilon}_{g,n}, \tag{8}$$

where $\mathbf{Z}_{g,n} = [\mathbf{Y}_{g,n}, \mathbf{X}_{g,n}, \overline{\mathbf{Y}}_{g,n}]$ 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 $s = 1, \ldots, p$ and $r = 1, \ldots, q$: (a) All diagonal elements of $\mathbf{W}_{s,n}$ and $\mathbf{M}_{r,n}$ are zero. (b) $\left\| \mathbf{W}_{s,n} \right\|_1 \leq c$, $\left\| \mathbf{M}_{r,n} \right\|_1 \leq c$ for some finite constant $c$ which does not depend on $n$, and $\left\| \mathbf{W}_{s,n} \right\|_\infty = 1$, $\left\| \mathbf{M}_{r,n} \right\|_\infty = 1$.

**Assumption 2.** (a) The matrices $\mathbf{I}_{nG} - \mathbf{B}_n^*$ are nonsingular. (b) The spatial autoregressive parameters satisfy $\sup_n \sum_{r \in \mathbf{I}_{g,\rho}} \left| \rho_{g,r,n} \right| < 1$ for $g = 1, \ldots, G$, where $\mathbf{I}_{g,\rho} = \{r_{g,1}, \ldots, r_{g,q_g}\} \subseteq \{1, \ldots, q\}$ 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 $\left\| \mathbf{W}_{s,n} \right\|_\infty = 1$ and $\left\| \mathbf{M}_{r,n} \right\|_\infty = 1$ implies a normalization for

the parameters. For interpretation, let $\mathbf{W}^*_{s,n}$ be some spatial weights matrix with $\left\|\mathbf{W}^*_{s,n}\right\|_\infty \neq 1$ and let $\lambda^*_{lg,s,n}$ be the corresponding spatial autoregressive parameter on $\mathbf{W}^*_{s,n}\mathbf{y}_{l,n}$. Now define $\mathbf{W}_{s,n} = \mathbf{W}^*_{s,n}/\left\|\mathbf{W}^*_{s,n}\right\|_\infty$ and $\lambda_{lg,s,n} = \lambda^*_{lg,s,n}\left\|\mathbf{W}^*_{s,n}\right\|_\infty$, then $\left\|\mathbf{W}_{s,n}\right\|_\infty = 1$ and $\lambda_{lg,s,n}\mathbf{W}_{s,n} = \lambda^*_{lg,s,n}\mathbf{W}^*_{s,n}$. Thus the normalizations $\left\|\mathbf{W}_{s,n}\right\|_\infty = 1$ and $\left\|\mathbf{M}_{r,n}\right\|_\infty = 1$ 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).[13]

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 $\mathbf{R}^*_n = diag^G_{g=1}\left[\mathbf{R}^*_{g,n}\right]$ 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 $\left\|\mathbf{R}^*_n\right\|_\infty \leq \max_g \sum_{r\in\mathbf{I}_{g,\rho}}\left|\rho_{g,r,n}\right| < 1$, 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 $\sup_n\left\|\mathbf{R}^*_{g,n}\right\|_\infty < 1$, which implies that the row sums of the matrices $\left[\mathbf{I}_n - \mathbf{R}^*_{g,n}\right]^{-1}$ are uniformly bounded in absolute value. To see this, observe that $\left\|\left[\mathbf{I}_n - \mathbf{R}^*_{g,n}\right]^{-1}\right\|_\infty \leq 1/\left[1 - \left\|\mathbf{R}^*_{g,n}\right\|_\infty\right] \leq 1/\left[1 - \sup_n\left\|\mathbf{R}^*_{g,n}\right\|_\infty\right] < \infty$. 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 $\mathbf{X}_n$. 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 $\mathbf{V}_n = [\mathbf{v}_{1.n}, \ldots, \mathbf{v}_{G,n}]$ be an $n \times G$ matrix of basic innovations and let $\mathbf{v}_n = vec(\mathbf{V}_n)$.

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

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

---

[13]The 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.

where $\boldsymbol{\Sigma}'_\star$ is a nonsingular $G \times G$ matrix and the random variables $\{v_{ig,n} : i = 1, \ldots, n, g = 1, \ldots, G\}$ 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,

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

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 $[\mathbf{Y}_n, \mathbf{X}_n, \overline{\mathbf{Y}}_n]$ 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

$$\mathbf{y}_{*g,n} = \mathbf{Z}_{*g,n}\boldsymbol{\delta}_{g,n} + \boldsymbol{\varepsilon}_{g,n} \tag{10}$$

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

$$\mathbf{y}_{*n} = \mathbf{Z}_{*n}\boldsymbol{\delta}_n + \boldsymbol{\varepsilon}_n \tag{11}$$

with $\mathbf{y}_{*n} = \left[\mathbf{y}'_{*1,n}, \ldots, \mathbf{y}'_{*G,n}\right]'$, $\mathbf{Z}_{*n} = diag_{g=1}^G \left[\mathbf{Z}_{*g,n}\right]$, $\boldsymbol{\delta}_n = \left[\boldsymbol{\delta}'_{1,n}, \ldots, \boldsymbol{\delta}'_{G,n}\right]'$ and where $\boldsymbol{\varepsilon}_n$ is as defined above.

## 3. MOMENT CONDITIONS

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

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

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 ($g = 1, \ldots, G$)

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

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

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 nonstochastic. 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.[14]

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 $\underline{\mathbf{H}}_n = \left[\mathbf{X}_n, \mathbf{M}_{1,n}\mathbf{X}_n, \ldots, \mathbf{M}_{p,n}\mathbf{X}_n\right]$.

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 $\left[\mathbf{X}_{g,n}, \mathbf{M}_{1,n}\mathbf{X}_{g,n}, \ldots, \mathbf{M}_{p,n}\mathbf{X}_{g,n}\right]$.

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

(a) $\mathbf{Q}_{HH} = \lim_{n \to \infty} n^{-1}\mathbf{H}_n'\mathbf{H}_n$ is finite and nonsingular.
(b) $\mathbf{Q}_{HZ,g} = \text{plim}_{n \to \infty} n^{-1}\mathbf{H}_n'\mathbf{Z}_{g,n}$ and $\mathbf{Q}_{HMZ,r,g} = \text{plim}_{n \to \infty} n^{-1}\mathbf{H}_n'\mathbf{M}_{r,n}\mathbf{Z}_{g,n}$ 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
   $\lambda_{\min}\left[\mathbf{Q}_{HZ,g*}(\boldsymbol{\rho}_{g,n})'\mathbf{Q}_{HH}^{-1}\mathbf{Q}_{HZ,g*}(\boldsymbol{\rho}_{g,n})\right] \geq c$ 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.

---

[14]We 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.

The best instruments for $\mathbf{Y}_{g,n}$ and $\overline{\mathbf{Y}}_{g,n}$ are given by their conditional means. Observe that in light of (6) we have

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

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

$$\overline{\mathbf{X}}_{1,n} = \left[\mathbf{W}_{j_1,n}\mathbf{X}_n\right]_{j_1=1}^p,$$
$$\overline{\mathbf{X}}_{2,n} = \left[\mathbf{W}_{j_1,n}\mathbf{W}_{j_2,n}\mathbf{X}_n\right]_{j_1,j_2=1}^p,$$
$$\vdots$$
$$\overline{\mathbf{X}}_{R,n} = \left[\mathbf{W}_{j_1,n}\mathbf{W}_{j_2,n}\ldots\mathbf{W}_{j_R,n}\mathbf{X}_n\right]_{j_1,j_2,\ldots,j_R=1}^p,$$

and let $\mathbf{H}_{R,n} = [\mathbf{X}_n, \overline{\mathbf{X}}_{1,n}, \ldots, \overline{\mathbf{X}}_{R,n}]$. Now suppose $\mathbf{W}_{s_1}, \ldots, \mathbf{W}_{s_g}$ are the spatial weights matrices appearing in $\overline{\mathbf{Y}}_{g,n}$, then by including in $\mathbf{H}_n$ the linearly independent columns of $\left[\mathbf{H}_{R,n}, \mathbf{W}_{s_1}\mathbf{H}_{R,n}, \ldots, \mathbf{W}_{s_g}\mathbf{H}_{R,n}\right]$, 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 $\mathbf{M}_{r_1}, \ldots, \mathbf{M}_{r_g}$ 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 $\mathbf{M}_{r_1}\left[\mathbf{H}_{R,n}, \mathbf{W}_{s_1}\mathbf{H}_{R,n}, \ldots, \mathbf{W}_{s_g}\mathbf{H}_{R,n}\right], \ldots, \mathbf{M}_{r_g}\left[\mathbf{H}_{R,n}, \mathbf{W}_{s_1}\mathbf{H}_{R,n}, \ldots, \mathbf{W}_{s_g}\mathbf{H}_{R,n}\right]$ 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 $\boldsymbol{\delta}_{g,n}$ 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 $\mathbf{A}_{s,n}$, $s = 1, \ldots, S$, are bounded uniformly in absolute value by some finite constant and, furthermore, all diagonal elements of $\mathbf{A}_{s,n}$ are zero for any $s = 1, \ldots, S$.

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

$$\mathbf{M}_{r,n}, \quad \mathbf{M}'_{r,n}\mathbf{M}_{r,n} - diag(\mathbf{M}'_{r,n}\mathbf{M}_{r,n}),$$
$$\mathbf{W}_{s,n}, \quad \mathbf{W}'_{s,n}\mathbf{W}_{s,n} - diag(\mathbf{W}'_{s,n}\mathbf{W}_{s,n}), \quad \mathbf{M}'_{s,n}\mathbf{W}_{s,n} - diag(\mathbf{M}'_{s,n}\mathbf{W}_{s,n}).$$

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

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

where

$$\underset{S \times 1}{\boldsymbol{\gamma}_{g,n}} = \begin{bmatrix} \boldsymbol{\gamma}_{1,g,n} \\ \vdots \\ \boldsymbol{\gamma}_{S,g,n} \end{bmatrix},$$

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

with

$$\boldsymbol{\gamma}_{s,g,n} = n^{-1}E\mathbf{u}'_{g,n}\overline{\mathbf{A}}_{s,n}\mathbf{u}_{g,n},$$
$$\boldsymbol{\Gamma}_{s1,g,n} = n^{-1}(2E\mathbf{u}'_{g,n}\mathbf{M}'_{r_{g,1},n}\overline{\mathbf{A}}_{s,n}\mathbf{u}_{g,n}, \ldots, 2E\mathbf{u}'_{g,n}\mathbf{M}'_{r_{g,q_g},n}\overline{\mathbf{A}}_{s,n}\mathbf{u}_{g,n}),$$
$$\boldsymbol{\Gamma}_{s2,g,n} = -n^{-1}(E\mathbf{u}'_{g,n}\mathbf{M}'_{r_{g,1},n}\overline{\mathbf{A}}_{s,n}\mathbf{M}_{r_{g,1},n}\mathbf{u}_{g,n}, \ldots, E\mathbf{u}'_{g,n}\mathbf{M}'_{r_{g,q_g},n}\overline{\mathbf{A}}_{s,n}\mathbf{M}_{r_{g,q_g},n}\mathbf{u}_{g,n}),$$
$$\boldsymbol{\Gamma}_{s3,g,n} = -n^{-1}(2E\mathbf{u}'_{g,n}\mathbf{M}'_{r_{g,1},n}\overline{\mathbf{A}}_{s,n}\mathbf{M}_{r_{g,2},n}\mathbf{u}_{g,n}, \ldots, 2E\mathbf{u}'_{g,n}\mathbf{M}'_{r_{g,q_g-1},n}\overline{\mathbf{A}}_{s,n}\mathbf{M}_{r_{g,q_g},n}\mathbf{u}_{g,n}),$$

$$\mathbf{r}_{1,g,n} = (\rho_{g,r_{g,1},n}, \ldots, \rho_{g,r_{g,q_g},n})',$$

$$\mathbf{r}_{2,g,n} = (\rho^2_{g,r_{g,1},n}, \ldots, \rho^2_{g,r_{g,q_g},n})',$$

$$\mathbf{r}_{3,g,n} = (\rho_{g,r_{g,1},n}\rho_{g,r_{g,2},n}, \ldots, \rho_{g,r_{g,1},n}\rho_{g,r_{g,q_g},n}, \ldots, \rho_{g,r_{g,q_g-1},n}\rho_{g,r_{g,q_g},n})',$$

recalling the definition of the index set $\mathbf{I}_{g,\rho} = \{r_{g,1}, \ldots, r_{g,q_g}\} \subseteq \{1, \ldots, q\}$ and where $q_g^* = 2q_g + q_g(q_g - 1)/2.$[15]

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

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

## 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, Kelejian, and Prucha (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 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.

---

[15]To 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 $n$, $\mathbf{r}_{1,g} = (\rho_{g,2}, \rho_{g,4}, \rho_{g,5})'$, $\mathbf{r}_{2,g} = (\rho^2_{g,2}, \rho^2_{g,4}, \rho^2_{g,5})'$ 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 $\boldsymbol{\Gamma}_{s1,g}$, $\boldsymbol{\Gamma}_{s2,g}$, and $\boldsymbol{\Gamma}_{s3,g}$ are indexed analogously.

## 4.1. Consistency of GMM Estimator for $\rho$

Let $\widetilde{\boldsymbol{\Upsilon}}_{gg,n}$ 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

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

$$(16)$$

with $a_\rho \geq 1$. We note that the objective function for $\widetilde{\boldsymbol{\rho}}_{g,n}$ remains well defined even for values of $\overline{\boldsymbol{\rho}}_g$ for which $\mathbf{I}_n - \mathbf{R}_{g,n}^{*'}(\overline{\boldsymbol{\rho}}_g)$ 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 $\widetilde{\boldsymbol{\rho}}_{g,n}$.

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

**Assumption 9.** $\widetilde{\boldsymbol{\Upsilon}}_{g,n} - \boldsymbol{\Upsilon}_{g,n} = o_p(1)$, where $\boldsymbol{\Upsilon}_{g,n}$ is an $S \times S$ nonstochastic 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 $\boldsymbol{\rho}_{g,n}$ is identifiable unique w.r.t. the nonstochastic analog of the objective function of the GMM estimator[16]

$$E\mathbf{m}_{g,n}^{\rho}(\overline{\boldsymbol{\rho}}_g, \boldsymbol{\delta}_{g,n})' \boldsymbol{\Upsilon}_{g,n} E\mathbf{m}_{g,n}^{\rho}(\overline{\boldsymbol{\rho}}_g, \boldsymbol{\delta}_{g,n})$$
$$= \left[ \boldsymbol{\gamma}_{g,n} - \boldsymbol{\Gamma}_{g,n} \mathbf{r}_{g,n}(\overline{\boldsymbol{\rho}}_{g,n}) \right]' \boldsymbol{\Upsilon}_{g,n} \left[ \boldsymbol{\gamma}_{g,n} - \boldsymbol{\Gamma}_{g,n} \mathbf{r}_{g,n}(\overline{\boldsymbol{\rho}}_{g,n}) \right].$$

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 $\widetilde{\boldsymbol{\Upsilon}}_{g,n}$ is positive definite with probability trending to one. Of course, Assumption 9 is satisfied for $\widetilde{\boldsymbol{\Upsilon}}_{g,n} = \boldsymbol{\Upsilon}_{g,n} = \mathbf{I}_S$. In this case, the estimator defined by (16) reduces to a nonlinear least-squares estimator. Choices of $\widetilde{\boldsymbol{\Upsilon}}_{g,n}$ which result in efficient estimates of $\boldsymbol{\rho}_{g,n}$ will be discussed below in conjunction with the asymptotic normality result.

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

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

$$\widetilde{\boldsymbol{\rho}}_{g,n} - \boldsymbol{\rho}_{g,n} \xrightarrow{p} \mathbf{0} \quad as \quad n \to \infty.$$

---

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

### 4.2. Asymptotic Distribution of GMM Estimator for $\rho$

The limiting distribution of the GMM estimator $\widetilde{\rho}_{g,n}$ will generally depend on the estimator $\widetilde{\delta}_{g,n}$ 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 two-stage least squares (2SLS) and three-stage least squares (3SLS) estimators defined in the next section. To keep the discussion general, we maintain the following assumption regarding $\widetilde{\delta}_{g,n}$.

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

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

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 nonstochastic matrices whose elements are uniformly bounded in absolute value by a finite constant, and where $p_{\delta_g}$ 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 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, $\sum_{h=1}^{G} \mathbf{T}'_{1h,n} \varepsilon_{h,n}$ reduces to $\mathbf{T}'_{11,n} \varepsilon_{1,n}$. Also note that under Assumptions 4 and 10, we have $n^{1/2}(\widetilde{\delta}_{g,n} - \delta_{g,n}) = O_p(1)$, 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

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

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

$$\psi^{\rho\rho}_{rs,gg,n} = \sigma^2_{gg}(2n)^{-1} tr\left[\left(\mathbf{A}_{r,n} + \mathbf{A}'_{r,n}\right)\left(\mathbf{A}_{s,n} + \mathbf{A}'_{s,n}\right)\right] \tag{18}$$

$$+ n^{-1} \boldsymbol{\alpha}'_{g,r,n} \left[\sum_{h=1}^{G}\sum_{l=1}^{G} \sigma_{hl}\mathbf{T}'_{gh,n}\mathbf{T}_{gl,n}\right]\boldsymbol{\alpha}_{g,s,n},$$

with

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

As shown in the proof of the subsequent theorem, $\boldsymbol{\Psi}^{\rho\rho}_{gg,n}$ is the asymptotic VC matrix of the sample moment vector $\mathbf{m}^{\rho}_{g,n}(\boldsymbol{\rho}_g, \widetilde{\delta}_{g,n})$. 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 $\widetilde{\delta}_{g,n} = \delta_{g,n}$ or $\mathbf{T}_{gh,n} = \mathbf{0}$, in which case the second term in (18) is zero.

THEOREM 2 (Asymptotic Normality). *Let $\widetilde{\boldsymbol{\rho}}_{g,n} = \widetilde{\boldsymbol{\rho}}_{g,n}(\widetilde{\boldsymbol{\Upsilon}}_{g,n})$ 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*

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

*where*

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

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

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

The above theorem implies that the difference between the cumulative distribution function of $n^{1/2}(\widetilde{\boldsymbol{\rho}}_{g,n} - \boldsymbol{\rho}_{g,n})$ and that of $N\left[\mathbf{0}, \boldsymbol{\Omega}_{\widetilde{\boldsymbol{\rho}}_{g,n}}\right]$ converges pointwise to zero, which justifies the use of the latter distribution as an approximation of the former.[17]

**Remark.** Clearly, $\boldsymbol{\Omega}_{gg,n}^{\rho\rho}((\boldsymbol{\Psi}_{gg,n}^{\rho\rho})^{-1}) = \left[\mathbf{J}_{g,n}'(\boldsymbol{\Psi}_{gg,n}^{\rho\rho})^{-1}\mathbf{J}_{g,n}\right]^{-1}$ 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 $\widetilde{\boldsymbol{\Upsilon}}_{g,n}$ 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, $\lambda_{\max}\left(\boldsymbol{\Psi}_{gg,n}^{\rho\rho}\right) \leq c_\Psi^{**}$ 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 $g, h = 1, \ldots, G$ define*

$$\widetilde{\sigma}_{gh,n} = n^{-1}\widetilde{\boldsymbol{\varepsilon}}_{g,n}'\widetilde{\boldsymbol{\varepsilon}}_{h,n} \tag{22}$$

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

---

[17]This follows from Corollary F4 in Pötscher and Prucha (1997). Compare also the discussion on pp. 86–87 in that reference.

We note that in the above lemma $\widetilde{\boldsymbol{\rho}}_{g,n}$ and $\widetilde{\boldsymbol{\delta}}_{g,n}$ can be any consistent estimators. Now, let $\widetilde{\boldsymbol{\Gamma}}_{g,n}$ be defined as in (15) and corresponding to (17) define

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

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

$$\widetilde{\psi}_{rs,gg,n}^{\rho\rho} = \widetilde{\sigma}_{gg,n}^2 (2n)^{-1} tr\left[\left(\mathbf{A}_{r,n} + \mathbf{A}_{r,n}'\right)\left(\mathbf{A}_{s,n} + \mathbf{A}_{s,n}'\right)\right] \tag{24}$$

$$+ n^{-1}\widetilde{\boldsymbol{\alpha}}_{g,r,n}' \left[\sum_{h=1}^{G}\sum_{l=1}^{G} \widetilde{\sigma}_{hl,n}\widetilde{\mathbf{T}}_{gh,n}'\widetilde{\mathbf{T}}_{gl,n}\right]\widetilde{\boldsymbol{\alpha}}_{g,s,n},$$

with

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

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

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

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

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

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

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

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

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

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

## 4.3. Joint Asymptotic Distribution of Estimators for $\rho$ and $\delta$

In the following, let $\boldsymbol{\delta}_n = [\boldsymbol{\delta}_{1,n}', \ldots, \boldsymbol{\delta}_{G,n}']'$ and $\boldsymbol{\rho}_n = [\boldsymbol{\rho}_{1,n}', \ldots, \boldsymbol{\rho}_{G,n}']'$, and let $\widetilde{\boldsymbol{\delta}}_n = [\widetilde{\boldsymbol{\delta}}_{1,n}', \ldots, \widetilde{\boldsymbol{\delta}}_{G,n}']'$ and $\widetilde{\boldsymbol{\rho}}_n = [\widetilde{\boldsymbol{\rho}}_{1,n}', \ldots, \widetilde{\boldsymbol{\rho}}_{G,n}']'$ denote the corresponding estimators as defined in the previous subsections. In this section, we derive the joint limiting distribution of $\widetilde{\boldsymbol{\delta}}_n$ and $\widetilde{\boldsymbol{\rho}}_n$. 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 $\boldsymbol{\eta}_n' = [\boldsymbol{\eta}_{1,n}', \ldots, \boldsymbol{\eta}_{G,n}']'$ and $\boldsymbol{\xi}_n' = [\boldsymbol{\xi}_{1,n}', \ldots, \boldsymbol{\xi}_{G,n}']'$, 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

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

where

$$\boldsymbol{\Psi}_n^{\rho\rho} = E\boldsymbol{\xi}_n \boldsymbol{\xi}_n' = \begin{bmatrix} \boldsymbol{\Psi}_{11,n}^{\rho\rho} & \cdots & \boldsymbol{\Psi}_{1G,n}^{\rho\rho} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Psi}_{G1,n}^{\rho\rho} & \cdots & \boldsymbol{\Psi}_{GG,n}^{\rho\rho} \end{bmatrix},$$

$$\boldsymbol{\Psi}_n^{\delta\delta} = E\boldsymbol{\eta}_n \boldsymbol{\eta}_n' = n^{-1} \mathbf{T}_n' (\boldsymbol{\Sigma} \otimes \mathbf{I}_n) \mathbf{T}_n,$$

$$\boldsymbol{\Psi}_n^{\delta\rho} = E\boldsymbol{\eta}_n \boldsymbol{\xi}_n' = n^{-1} \mathbf{T}_n' (\boldsymbol{\Sigma} \otimes \mathbf{I}_n) \mathbf{T}_n \, diag_{g=1}^{G} \left[ \boldsymbol{\alpha}_{g,1,n}, \ldots, \boldsymbol{\alpha}_{g,S,n} \right],$$

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

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

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

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

with

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

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

$$\widetilde{\boldsymbol{\Psi}}_n^{\delta\delta} = n^{-1}\widetilde{\mathbf{T}}_n'(\widetilde{\boldsymbol{\Sigma}}_n\otimes\mathbf{I}_n)\widetilde{\mathbf{T}}_n,$$
$$\widetilde{\boldsymbol{\Psi}}_n^{\delta\rho} = n^{-1}\widetilde{\mathbf{T}}_n'(\widetilde{\boldsymbol{\Sigma}}_n\otimes\mathbf{I}_n)\widetilde{\mathbf{T}}_n diag_{g=1}^G\left(\left[\widetilde{\boldsymbol{\alpha}}_{g,1,n},\ldots,\widetilde{\boldsymbol{\alpha}}_{g,S,n}\right]\right),$$

then our estimator for $\boldsymbol{\Psi}_n$ can be formulated as

$$\widetilde{\boldsymbol{\Psi}}_n = \begin{bmatrix} \widetilde{\boldsymbol{\Psi}}_n^{\delta\delta} & \widetilde{\boldsymbol{\Psi}}_n^{\delta\rho} \\ \widetilde{\boldsymbol{\Psi}}_n^{\delta\rho'} & \widetilde{\boldsymbol{\Psi}}_n^{\rho\rho} \end{bmatrix}. \tag{27}$$

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

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

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

*with*

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

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

$$\boldsymbol{\Omega}_n = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & diag_{g=1}^G\left((\mathbf{J}_{g,n}'\boldsymbol{\Upsilon}_{g,n}\mathbf{J}_{g,n})^{-1}\mathbf{J}_{g,n}'\boldsymbol{\Upsilon}_{g,n}\right) \end{bmatrix}\boldsymbol{\Psi}_n \tag{28}$$
$$\times \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & diag_{g=1}^G\left(\boldsymbol{\Upsilon}_{g,n}\mathbf{J}_{g,n}(\mathbf{J}_{g,n}'\boldsymbol{\Upsilon}_{g,n}\mathbf{J}_{g,n})^{-1}\right) \end{bmatrix},$$
$$\widetilde{\boldsymbol{\Omega}}_n = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & diag_{g=1}^G\left((\widetilde{\mathbf{J}}_{g,n}'\widetilde{\boldsymbol{\Upsilon}}_{g,n}\widetilde{\mathbf{J}}_{g,n})^{-1}\widetilde{\mathbf{J}}_{g,n}'\widetilde{\boldsymbol{\Upsilon}}_{g,n}\right) \end{bmatrix}\widetilde{\boldsymbol{\Psi}}_n$$
$$\times \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & diag_{g=1}^G\left(\widetilde{\boldsymbol{\Upsilon}}_{g,n}\widetilde{\mathbf{J}}_{g,n}(\widetilde{\mathbf{J}}_{g,n}'\widetilde{\boldsymbol{\Upsilon}}_{g,n}\widetilde{\mathbf{J}}_{g,n})^{-1}\right) \end{bmatrix}.$$

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

Theorem 4 implies that the difference between the joint cumulative distribution function of the estimators of all model parameters $n^{1/2}[(\widetilde{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_n)', (\widetilde{\boldsymbol{\rho}}_n - \boldsymbol{\rho}_n)']'$, 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.[18] The theorem also states that $\widetilde{\mathbf{\Omega}}_n$ is a consistent estimator of $\mathbf{\Omega}_n$. 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 $n^{1/2}[(\widetilde{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_n)', (\widetilde{\boldsymbol{\rho}}_n - \boldsymbol{\rho}_n)']'$.

## 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, 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 2SLS (GS2SLS) estimator of $\boldsymbol{\delta}_{g,n}$ and a GMM estimator of $\boldsymbol{\rho}_{g,n}$ based on GS2SLS residuals. WLOG we focus on the estimation of the parameters of the $g$th equation.

**Step 1a: 2SLS estimator of $\boldsymbol{\delta}_{g,n}$** 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 $\boldsymbol{\delta}_{g,n}$. The 2SLS estimator, say $\widetilde{\boldsymbol{\delta}}_{g,n}$, is then defined as

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

where $\widetilde{\mathbf{Z}}_{g,n} = \mathbf{P}_{\mathbf{H}_n}\mathbf{Z}_{g,n} = (\widetilde{\mathbf{Y}}_{g,n}, \mathbf{X}_{g,n}, \overline{\widetilde{\mathbf{Y}}}_{g,n})$, $\widetilde{\mathbf{Y}}_{g,n} = \mathbf{P}_{\mathbf{H}_n}\mathbf{Y}_{g,n}$, $\overline{\widetilde{\mathbf{Y}}}_{g,n} = \mathbf{P}_{\mathbf{H}_n}\overline{\mathbf{Y}}_{g,n}$, and where $\mathbf{P}_{\mathbf{H}_n} = \mathbf{H}_n(\mathbf{H}'_n\mathbf{H}_n)^{-1}\mathbf{H}'_n$.[19]

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

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

$$\widetilde{\boldsymbol{\rho}}_{g,n} = \underset{\overline{\boldsymbol{\rho}}_g : \sum_{r \in \mathbf{I}_{g,\rho}} |\overline{\rho}_{g,r}| \in [-a_\rho, a_\rho]}{argmin} \mathbf{m}^{\rho}_{g,n}(\overline{\boldsymbol{\rho}}_{g,n}, \widetilde{\boldsymbol{\delta}}_{g,n})' \mathbf{m}^{\rho}_{g,n}(\overline{\boldsymbol{\rho}}_{g,n}, \widetilde{\boldsymbol{\delta}}_{g,n}) \tag{30}$$

with $a_\rho \geq 1$.

**Step 2a: GS2SLS estimator of $\boldsymbol{\delta}_{g,n}$** Analogous to Kelejian and Prucha (1998), we next compute a GS2SLS estimator of $\boldsymbol{\delta}_{g,n}$, $\widehat{\boldsymbol{\delta}}_{g,n}(\widetilde{\boldsymbol{\rho}}_{g,n})$. This estimator is defined

---

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

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

as the 2SLS estimator of the $g$th equation of the spatially Cochrane–Orcutt transformed model (10) with $\boldsymbol{\rho}_{g,n}$ replaced by $\widetilde{\boldsymbol{\rho}}_{g,n}$, i.e.,

$$\widehat{\boldsymbol{\delta}}_{g,n} = \widehat{\boldsymbol{\delta}}_{g,n}(\widetilde{\boldsymbol{\rho}}_{g,n}) = [\widehat{\mathbf{Z}}_{*g,n}(\widetilde{\boldsymbol{\rho}}_{g,n})'\mathbf{Z}_{*g,n}(\widetilde{\boldsymbol{\rho}}_{g,n})]^{-1}\widehat{\mathbf{Z}}_{*g,n}(\widetilde{\boldsymbol{\rho}}_{g,n})'\mathbf{y}_{*g,n}(\widetilde{\boldsymbol{\rho}}_{g,n}), \tag{31}$$

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

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

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

### Step 2b: Efficient GMM estimator of $\rho_{g,n}$ 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 $\mathbf{m}_{g,n}^{\rho}(\overline{\boldsymbol{\rho}}_{g,n}, \widehat{\boldsymbol{\delta}}_{g,n})$ 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

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

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

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

with

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

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.

## 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 $\boldsymbol{\Psi}_n^{\delta\rho}$, $\boldsymbol{\Psi}_n^{\rho\rho}$ and $\boldsymbol{\Omega}_n^{\delta\delta}$, $\boldsymbol{\Omega}_n^{\delta\rho}$, $\boldsymbol{\Omega}_n^{\rho\rho}$ as implied by the specific structure of $\widehat{\boldsymbol{\delta}}_{g,n}$ and $\widehat{\boldsymbol{\rho}}_{g,n}$. More specifically, let

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

with

$$\widehat{\boldsymbol{\Psi}}_{gg,n}^{\delta\rho} = \widehat{\boldsymbol{\Psi}}_{gg,n}^{\delta\delta}\left[\widehat{\boldsymbol{\alpha}}_{g,1,n},\ldots,\widehat{\boldsymbol{\alpha}}_{g,S,n}\right],$$

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

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 $\boldsymbol{\Psi}_n$ are bounded away from zero.[20] Then, $\widehat{\boldsymbol{\rho}}_{g,n}$ is efficient among the class of GMM estimators based on GS2SLS residuals, and*

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

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 GS3SLS estimator.

**Step 3a: GS3SLS estimator of $\boldsymbol{\delta}_n$** The GS3SLS estimator of $\boldsymbol{\delta}_n$ based on the stacked spatially Cochrane–Orcutt transformed model (11) is given by

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

where $\mathbf{y}_{*n}(\widehat{\boldsymbol{\rho}}_n) = \left[\mathbf{y}_{*1,n}'(\widehat{\boldsymbol{\rho}}_{1,n}),\ldots,\mathbf{y}_{*G,n}'(\widehat{\boldsymbol{\rho}}_{G,n})\right]'$, $\mathbf{Z}_{*n}(\widehat{\boldsymbol{\rho}}_n) = diag_{g=1}^G\left[\mathbf{Z}_{*g,n}(\widehat{\boldsymbol{\rho}}_{g,n})\right]$, $\widehat{\mathbf{Z}}_{*n}(\widehat{\boldsymbol{\rho}}_n) = diag_{g=1}^G\left[\widehat{\mathbf{Z}}_{*g,n}(\widehat{\boldsymbol{\rho}}_{g,n})\right]$ with $\widehat{\mathbf{Z}}_{*g,n}(\widehat{\boldsymbol{\rho}}_{g,n}) = \mathbf{P}_{\mathbf{H}_n}\mathbf{Z}_{*g,n}(\widehat{\boldsymbol{\rho}}_{g,n})$, 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 $\widehat{\widehat{\boldsymbol{\delta}}}_n$):

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

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

---

[20]Explicit expressions for the submatrices composing $\boldsymbol{\Psi}_n$ specialized to $\widehat{\boldsymbol{\rho}}_{g,n}$ and $\widehat{\boldsymbol{\delta}}_{g,n}$ are given in the proof of the theorem.

**Step 3b: GMM estimator of** $\rho_{g,n}$ **based on GS3SLS residuals** In a final step, we compute a further GMM estimator of $\rho_{g,n}$ based on the GS3SLS residuals $\widehat{\overline{\mathbf{u}}}_{g,n} = \mathbf{y}_{g,n} - \mathbf{Z}_{g,n}\widehat{\overline{\boldsymbol{\delta}}}_{g,n}$, where $\widehat{\overline{\boldsymbol{\delta}}}_{g,n}$ denotes the $g$th component of $\widehat{\overline{\boldsymbol{\delta}}}_n$. Let $\mathbf{m}_{g,n}^{\rho}(\overline{\boldsymbol{\rho}}_{g,n}, \widehat{\overline{\boldsymbol{\delta}}}_{g,n})$ denote the corresponding sample moment vector as defined in (15). Then, the corresponding GMM estimator for $\rho_{g,n}$ based on GS3SLS residuals is given by

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

where $\widehat{\overline{\boldsymbol{\Psi}}}_{gg,n}^{\rho\rho}$ is an estimator of the VC matrix $\boldsymbol{\Psi}_{gg,n}^{\rho\rho}$ of the limiting distribution of the normalized sample moments $n^{1/2}\mathbf{m}_{g,n}^{\rho}(\boldsymbol{\rho}_{g,n}, \widehat{\overline{\boldsymbol{\delta}}}_{g,n})$. Towards presenting the asymptotic distribution of $\widehat{\overline{\boldsymbol{\rho}}}_{1,n}, \ldots, \widehat{\overline{\boldsymbol{\rho}}}_{G,n}$ 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 $\widehat{\overline{\boldsymbol{\Psi}}}_n^{\rho\rho}$ and $\widehat{\overline{\boldsymbol{\Psi}}}_{gh,n}^{\rho\rho}$ denote the estimators for $\boldsymbol{\Psi}_n^{\rho\rho}$ and $\boldsymbol{\Psi}_{gh,n}^{\rho\rho}$, respectively, then the $(r,s)$th element of $\widehat{\overline{\boldsymbol{\Psi}}}_{gh,n}^{\rho\rho}$ is defined as

$$\widehat{\overline{\psi}}_{rs,gh,n}^{\rho\rho} = (2n)^{-1}\widehat{\sigma}_{gh,n}^2 tr\left[(\mathbf{A}_{r,n} + \mathbf{A}_{r,n}')(\mathbf{A}_{s,n} + \mathbf{A}_{s,n}')\right] \tag{35}$$
$$+ \widehat{\overline{\boldsymbol{\alpha}}}_{g,r,n}' \widehat{\overline{\boldsymbol{\Psi}}}_{gh,n}^{\delta\delta} \widehat{\overline{\boldsymbol{\alpha}}}_{h,s,n},$$

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

## 5.4. Asymptotic Properties of Full Information Estimators

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

$$\widehat{\overline{\boldsymbol{\Omega}}}_n^{\delta\delta} = \widehat{\overline{\boldsymbol{\Psi}}}_n^{\delta\delta},$$
$$\widehat{\overline{\boldsymbol{\Omega}}}_n^{\delta\rho} = \widehat{\overline{\boldsymbol{\Psi}}}_n^{\delta\rho} diag_{g=1}^G\left(\widehat{\overline{\boldsymbol{\mathfrak{J}}}}_{g,n}\right), \qquad \widehat{\overline{\boldsymbol{\Omega}}}_n^{\rho\rho} = diag_{g=1}^G\left(\widehat{\overline{\boldsymbol{\mathfrak{J}}}}_{g,n}'\right)\widehat{\overline{\boldsymbol{\Psi}}}_n^{\rho\rho} diag_{g=1}^G\left(\widehat{\overline{\boldsymbol{\mathfrak{J}}}}_{g,n}\right),$$

with

$$\widehat{\overline{\boldsymbol{\Psi}}}_n^{\delta\rho} = \widehat{\overline{\boldsymbol{\Psi}}}_n^{\delta\delta} diag_{g=1}^G\left([\widehat{\overline{\boldsymbol{\alpha}}}_{g,1,n}, \ldots, \widehat{\overline{\boldsymbol{\alpha}}}_{g,S,n}]\right),$$
$$\widehat{\overline{\boldsymbol{\mathfrak{J}}}}_{g,n} = \left(\widehat{\overline{\boldsymbol{\Psi}}}_{gg,n}^{\rho\rho}\right)^{-1}\widehat{\mathbf{J}}_{g,n}\left[\widehat{\mathbf{J}}_{g,n}'\left(\widehat{\overline{\boldsymbol{\Psi}}}_{gg,n}^{\rho\rho}\right)^{-1}\widehat{\mathbf{J}}_{g,n}\right]^{-1},$$

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 $\widehat{\overline{\boldsymbol{\rho}}}_n$ and $\widehat{\overline{\boldsymbol{\delta}}}_n$.

THEOREM 6 (Joint Asymptotic Normality of $\widehat{\widehat{\boldsymbol{\rho}}}_n$ and $\widehat{\widehat{\boldsymbol{\delta}}}_n$). *Suppose Assumptions 1–8 hold, and that the smallest eigenvalues of $\boldsymbol{\Psi}_n$ are bounded away from zero.*[21] *Then,*

$$
\begin{bmatrix} (\widehat{\widehat{\boldsymbol{\delta}}}_n - \boldsymbol{\delta}_n) \\ (\widehat{\widehat{\boldsymbol{\rho}}}_n - \boldsymbol{\rho}_n) \end{bmatrix} \sim AN \left[ 0, n^{-1} \begin{pmatrix} \widehat{\widehat{\boldsymbol{\Omega}}}_n^{\delta\delta} & \widehat{\widehat{\boldsymbol{\Omega}}}_n^{\delta\rho} \\ \widehat{\widehat{\boldsymbol{\Omega}}}_n^{\delta\rho\prime} & \widehat{\widehat{\boldsymbol{\Omega}}}_n^{\rho\rho} \end{pmatrix} \right].
$$

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

In the above theorem, the estimators for the asymptotic VC matrix of $\widehat{\widehat{\boldsymbol{\rho}}}_n$ and $\widehat{\widehat{\boldsymbol{\delta}}}_n$ 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 $\widehat{\widehat{\boldsymbol{\rho}}}_{g,n}$ (or any other consistent estimator).

## 6. LIMITED AND FULL INFORMATION ONE-STEP ESTIMATORS

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

Toward 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).[22] Let

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

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

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

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

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

---

[21] Explicit expressions for the submatrices $\boldsymbol{\Psi}_n$ are given in the proof of the theorem.

[22] The purpose of switching the superscripts from $\delta$ and $\rho$ 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 $\boldsymbol{\delta}_{g,n}$.

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

$$(\widehat{\boldsymbol{\delta}}_{g,n}^{o}, \widehat{\boldsymbol{\rho}}_{g,n}^{o}) = \underset{\overline{\boldsymbol{\rho}}_{g,n}, \overline{\boldsymbol{\delta}}_{g,n}}{\arg\min} \quad \mathbf{m}_{g,n}(\overline{\boldsymbol{\rho}}_{g,n}, \overline{\boldsymbol{\delta}}_{g,n})' \begin{bmatrix} \widetilde{\mathbf{\Psi}}_{gg,n}^{LL} & \mathbf{0} \\ \mathbf{0} & \widetilde{\mathbf{\Psi}}_{gg,n}^{QQ} \end{bmatrix}^{-1} \mathbf{m}_{g,n}(\overline{\boldsymbol{\rho}}_{g,n}, \overline{\boldsymbol{\delta}}_{g,n}). \quad \textbf{(36)}$$

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 GS2SLS (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

$$\widehat{\mathbf{S}}_{g,n}^{o} = \begin{bmatrix} \widehat{\mathbf{S}}_{g,n}^{o,LL} & \widehat{\mathbf{S}}_{g,n}^{o,LQ} \\ \widehat{\mathbf{S}}_{g,n}^{o,QL} & \widehat{\mathbf{S}}_{g,n}^{o,QQ} \end{bmatrix},$$

$$\widehat{\mathbf{S}}_{g,n}^{o,LL} = \widehat{\sigma}_{gg,n}^{o-1}[n^{-1}\widehat{\mathbf{Z}}_{*g,n}(\widehat{\boldsymbol{\rho}}_{g,n}^{o})'\widehat{\mathbf{Z}}_{*g,n}(\widehat{\boldsymbol{\rho}}_{g,n}^{o})] + \widehat{\sigma}_{gg,n}^{o-2}\widehat{\boldsymbol{\alpha}}_{g,n}^{o}(\mathbf{K}_{n}^{QQ})^{-1}\widehat{\boldsymbol{\alpha}}_{g,n}^{o\prime},$$

$$\widehat{\mathbf{S}}_{g,n}^{o,LQ} = \widehat{\mathbf{S}}_{g,n}^{o,QL\prime} = -\widehat{\sigma}_{gg,n}^{o-2}\widehat{\boldsymbol{\alpha}}_{g,n}^{o}(\mathbf{K}_{n}^{QQ})^{-1}\mathbf{J}_{g,n}(\widehat{\boldsymbol{\rho}}_{g,n}^{o}),$$

$$\widehat{\mathbf{S}}_{g,n}^{o,QQ} = \widehat{\sigma}_{gg,n}^{o-2}\mathbf{J}_{g,n}'(\widehat{\boldsymbol{\rho}}_{g,n}^{o})\left(\mathbf{K}_{n}^{QQ}\right)^{-1}\mathbf{J}_{g,n}(\widehat{\boldsymbol{\rho}}_{g,n}^{o}),$$

where $\widehat{\sigma}_{gg,n}^{o} = n^{-1}\widehat{\boldsymbol{\varepsilon}}_{g,n}^{o\prime}\widehat{\boldsymbol{\varepsilon}}_{g,n}^{o}$ 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 $\widehat{\boldsymbol{\alpha}}_{g,n}^{o} = [\widehat{\boldsymbol{\alpha}}_{g,1,n}^{o}, \ldots, \widehat{\boldsymbol{\alpha}}_{g,S,n}^{o}]$ 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 $\widehat{\mathbf{u}}_{g,n}^{o} = \mathbf{y}_{g,n} - \mathbf{Z}_{g,n}\widehat{\boldsymbol{\delta}}_{n}^{o}$.

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.

Toward 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

$$\boldsymbol{\Psi}_n^{LL} = \boldsymbol{\Sigma}_n \otimes \left[ n^{-1} \mathbf{H}_n' \mathbf{H}_n \right],$$

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

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

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

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

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

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

$$\widehat{\widehat{\mathbf{S}}}_n^o = \begin{bmatrix} \widehat{\widehat{\mathbf{S}}}_n^{LL} & \widehat{\widehat{\mathbf{S}}}_n^{LQ} \\ \widehat{\widehat{\mathbf{S}}}_n^{QL} & \widehat{\widehat{\mathbf{S}}}_n^{QQ} \end{bmatrix},$$

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

$$\widehat{\mathbf{S}}_{g,n}^{o,LQ} = \widehat{\mathbf{S}}_{g,n}^{o,LQ\prime} = -diag_{g=1}^{G}[\widehat{\widehat{\boldsymbol{\alpha}}}_{g,n}^{o}][(\widehat{\widehat{\boldsymbol{\Sigma}}}_{SQ,n}^{o})^{-1} \otimes (\mathbf{K}_{n}^{QQ})^{-1}]diag_{g=1}^{G}[\mathbf{J}_{g,n}(\widehat{\widehat{\boldsymbol{\rho}}}_{n}^{o})],$$

$$\widehat{\mathbf{S}}_{g,n}^{o,QQ} = diag_{g=1}^{G}[\mathbf{J}_{g,n}'(\widehat{\widehat{\boldsymbol{\rho}}}_{n}^{o})][(\widehat{\widehat{\boldsymbol{\Sigma}}}_{SQ,n}^{o})^{-1} \otimes (\mathbf{K}_{n}^{QQ})^{-1}]diag_{g=1}^{G}[\mathbf{J}_{g,n}(\widehat{\widehat{\boldsymbol{\rho}}}_{n}^{o})],$$

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

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

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

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 $\boldsymbol{\alpha}_{g,n} = 0$, 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 information generalized method of moments estimators, which utilize both linear and quadratic moment conditions. We consider both two- 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* $\boldsymbol{\mu}_n = (\mu_{1,n}, \ldots, \mu_{m_n,n})$ *and* $\boldsymbol{\eta}_n = (\eta_{1,n}, \ldots, \eta_{m_n,n})'$ *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* $\boldsymbol{\xi}_n = (\xi_{1,n}, \ldots, \xi_{m_n,n})' = \boldsymbol{\mu}_n + \mathbf{A}_n \boldsymbol{\eta}_n$. *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*

$$E \left| z_{ij,n} \right|^4 \leq C < \infty, \tag{A.1}$$

*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*

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

*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*

$$\mathbf{P}_{gg} = \mathbf{Q}_{\mathbf{HH}}^{-1}\mathbf{Q}_{\mathbf{HZ},g}[\mathbf{Q}'_{\mathbf{HZ},g}\mathbf{Q}_{\mathbf{HH}}^{-1}\mathbf{Q}_{\mathbf{HZ},g}]^{-1},$$

$$\mathbf{F}_{gg,n} = \left(\mathbf{I}_n - \mathbf{R}_{g,n}^{*'}\right)^{-1}\mathbf{H}_n.$$

(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*

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

(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*

$$\mathbf{P}_{gg,n}^* = \mathbf{Q}_{\mathbf{HH}}^{-1}\mathbf{Q}_{\mathbf{HZ},g*}(\boldsymbol{\rho}_{g,n})[\mathbf{Q}'_{\mathbf{HZ},g*}(\boldsymbol{\rho}_{g,n})\mathbf{Q}_{\mathbf{HH}}^{-1}\mathbf{Q}_{\mathbf{HZ},g*}(\boldsymbol{\rho}_{g,n})]^{-1},$$

$$\mathbf{F}_{gg,n}^* = \mathbf{H}_n.$$

(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*

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

(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 and 6 hold. Consider the GS3SLS estimator*

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

*where* $\widehat{\mathbf{Z}}_{*n}(\widehat{\boldsymbol{\rho}}_n) = diag_{g=1}^G\left[\widehat{\mathbf{Z}}'_{*g,n}(\widehat{\boldsymbol{\rho}}_{g,n})\right]$ *with* $\widehat{\mathbf{Z}}_{*g,n}(\widehat{\boldsymbol{\rho}}_n) = \mathbf{P}_{\mathbf{H}_n}\mathbf{Z}_{*g,n}(\widehat{\boldsymbol{\rho}}_n)$, *and* $\widehat{\boldsymbol{\rho}}_n = \left[\widehat{\boldsymbol{\rho}}'_{1,n},\ldots,\widehat{\boldsymbol{\rho}}'_{G,n}\right]'$ *is any consistent estimator for* $\widehat{\boldsymbol{\rho}}_n$ *and* $\widehat{\boldsymbol{\Sigma}}_n$ *is any consistent estimator for* $\boldsymbol{\Sigma}$. *Then*
(a) $n^{1/2}[\widehat{\widehat{\boldsymbol{\delta}}}_n - \boldsymbol{\delta}_n] = n^{-1/2}\mathbf{T}_n^{**'}\boldsymbol{\varepsilon}_n + o_p(1)$ *with* $\mathbf{T}_n^{**} = \mathbf{F}_n^{**}\mathbf{P}_n^{**}$, *and where*

$$\mathbf{P}_n^{**} = \left[\boldsymbol{\Sigma}^{-1}\otimes\mathbf{Q}_{\mathbf{HH}}^{-1}\right]diag\left[\mathbf{Q}_{\mathbf{HZ},g*}(\boldsymbol{\rho}_{g,n})\right]$$
$$\times\left\{diag\left[\mathbf{Q}'_{\mathbf{HZ},g*}(\boldsymbol{\rho}_{g,n})\right]\left[\boldsymbol{\Sigma}^{-1}\otimes\mathbf{Q}_{\mathbf{HH}}^{-1}\right]diag\left[\mathbf{Q}_{\mathbf{HZ},g*}(\boldsymbol{\rho}_{g,n})\right]\right\}^{-1}$$

*and*

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

*(b)* $n^{-1/2}\mathbf{T}_n^{**\prime}\boldsymbol{\varepsilon}_n = O_p(1)$.

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

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

*(d)* $\lambda_{\min}(n^{-1}\mathbf{T}_n^{**\prime}\mathbf{T}_n^{**}) \geq c$ *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 $\mathfrak{u}_n$ and corresponding forms based on a predictor $\widetilde{\mathfrak{u}}_n$.

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

$$\underset{n \times n}{\mathfrak{R}_n} \underset{n \times 1}{\mathfrak{u}_n} = \underset{n \times 1}{\mathfrak{e}_n},$$

where $\mathfrak{R}_n$ is a nonstochastic nonsigular $n \times n$ 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 $\mathfrak{e}_n = (\mathfrak{e}_{1,n}, \ldots, \mathfrak{e}_{n,n})'$ have the following properties: For each $n \geq 1$ the random variables $\mathfrak{e}_{1,n}, \ldots, \mathfrak{e}_{n,n}$ are totally independent with $E\mathfrak{e}_{i,n} = 0$, $E(\mathfrak{e}_{i,n}^2) = \sigma^2 > 0$, and $\sup_{1 \leq i \leq n, n \geq 1} E|\mathfrak{e}_{i,n}|^{4+\upsilon} < \infty$ for some $\upsilon > 0$.

**Assumption A.2.** The predictor $\widetilde{\mathfrak{u}}_n$ for $\mathfrak{u}_n$ satisfies that

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

where $\mathfrak{D}_n = (\mathfrak{d}_{ij,n})$ is an $n \times p_\Delta$ random matrix and $\Delta_n$ 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,

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

**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 $\mathfrak{u}_n$ 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 $\mathfrak{u}_n = \mathbf{u}_{g,n}$, $\mathfrak{D}_n = -\mathbf{Z}_{g,n}$, $\mathfrak{R}_n = \mathbf{I} - \sum_{r \in \mathbf{I}_{g,\rho}} \rho_{g,r,n}\mathbf{M}_{r,n}$, $\mathfrak{e}_n = \boldsymbol{\varepsilon}_{g,n}$, and $\Delta_n = \widetilde{\boldsymbol{\delta}}_{g,n} - \boldsymbol{\delta}_{g,n}$, where $\widetilde{\boldsymbol{\delta}}_{g,n}$ is some estimator for the parameter vector $\boldsymbol{\delta}_{g,n}$. Observe that under Assumptions 1-4, and given $\widetilde{\boldsymbol{\delta}}_{g,n}$ 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) $n^{-1}E\left|\mathfrak{u}_n' \mathbf{A}_n^* \mathfrak{u}_n\right| = O(1)$, $var(n^{-1}\mathfrak{u}_n' \mathbf{A}_n^* \mathfrak{u}_n) = o(1)$ and

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

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

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

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

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

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

We next state an additional assumption regarding $\Delta_n$, 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 $n^{-1/2}\widetilde{\mathfrak{u}}_n' \mathbf{A}_n^* \widetilde{\mathfrak{u}}_n$ 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 $\mathfrak{u}_n' \mathbf{A}_n^* \mathfrak{u}_n = \mathfrak{u}_n' \mathfrak{R}_n' \mathbf{A}_n \mathfrak{R}_n \mathfrak{u}_n = \mathfrak{e}_n' \mathbf{A}_n \mathfrak{e}_n$.

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

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

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 $\sup_n n^{-1}\sum_{i=1}^{n}\left|f_{his,n}\right|^\eta < \infty$ with $\eta > 2$ for $h = 1, \ldots, G, s = 1, \ldots, p_F$, 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 $\Delta_n$ is asymptotically linear satisfying Assumption A.4, and suppose $\mathfrak{e}_n = \boldsymbol{\varepsilon}_{g,n}$. 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*

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

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

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

$$E\left[n^{-1/2}\boldsymbol{\varepsilon}'_{g,n}\mathbf{A}_n\boldsymbol{\varepsilon}_{g,n}+n^{-1/2}\sum_{h=1}^{G}\mathbf{a}'_{h,n}\boldsymbol{\varepsilon}_{h,n}\right]=0,$$

$$var\left[n^{-1/2}\boldsymbol{\varepsilon}'_{g,n}\mathbf{A}_n\boldsymbol{\varepsilon}_{g,n}+n^{-1/2}\sum_{h=1}^{G}\mathbf{a}'_{h,n}\boldsymbol{\varepsilon}_{h,n}\right]$$

$$=n^{-1}\sigma^2_{gg}tr\left[\mathbf{A}_n(\mathbf{A}_n+\mathbf{A}'_n)\right]+n^{-1}\sum_{h=1}^{G}\sum_{l=1}^{G}\sigma_{hl}\mathbf{a}'_{h,n}\mathbf{a}_{l,n}.$$

## B. Appendix: Proofs for Section 4

**Proof of Theorem 1.** The existence and measurability of $\widetilde{\rho}_{g,n}$ 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 nonstochastic counterpart are given by, respectively,

$$R_n(\omega,\overline{\boldsymbol{\rho}}_g)=\mathbf{m}^{\rho}_{g,n}(\overline{\boldsymbol{\rho}}_g,\widetilde{\boldsymbol{\delta}}_{g,n})'\widetilde{\boldsymbol{\Upsilon}}_{g,n}\mathbf{m}^{\rho}_{g,n}(\overline{\boldsymbol{\rho}}_g,\widetilde{\boldsymbol{\delta}}_{g,n})$$

$$=\left[\widetilde{\boldsymbol{\Gamma}}_{g,n}\mathbf{r}_{g,n}(\overline{\boldsymbol{\rho}}_g)-\widetilde{\boldsymbol{\gamma}}_{g,n}\right]'\widetilde{\boldsymbol{\Upsilon}}_{g,n}\left[\widetilde{\boldsymbol{\Gamma}}_{g,n}\mathbf{r}_{g,n}(\overline{\boldsymbol{\rho}}_g)-\widetilde{\boldsymbol{\gamma}}_{g,n}\right],$$

$$\overline{R}_n(\overline{\boldsymbol{\rho}}_g)=\left[\boldsymbol{\Gamma}_{g,n}\mathbf{r}_{g,n}(\overline{\boldsymbol{\rho}}_g)-\boldsymbol{\gamma}_n\right]'\boldsymbol{\Upsilon}_n\left[\boldsymbol{\Gamma}_{g,n}\mathbf{r}_{g,n}(\overline{\boldsymbol{\rho}}_g)-\boldsymbol{\gamma}_n\right],$$

where the quantities appearing in the above equation are defined before the theorem in the text. To prove the consistency of $\widetilde{\rho}_{g,n}$ 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 $\overline{R}_n$. Observe that $\overline{R}_n(\overline{\boldsymbol{\rho}}_g)\geq 0$ and that $\overline{R}_n(\boldsymbol{\rho}_{g,n})=0$, 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

$$\overline{R}_n(\overline{\boldsymbol{\rho}}_g)-\overline{R}_n(\boldsymbol{\rho}_{g,n})=\overline{R}_n(\overline{\boldsymbol{\rho}}_g)$$

$$=\left[\boldsymbol{\Gamma}_{g,n}\mathbf{r}_{g,n}(\overline{\boldsymbol{\rho}}_g)-\boldsymbol{\Gamma}_{g,n}\mathbf{r}_{g,n}(\boldsymbol{\rho}_{g,n})\right]'\boldsymbol{\Upsilon}_n\left[\boldsymbol{\Gamma}_{g,n}\mathbf{r}_{g,n}(\overline{\boldsymbol{\rho}}_g)-\boldsymbol{\Gamma}_{g,n}\mathbf{r}_{g,n}(\boldsymbol{\rho}_{g,n})\right]$$

$$=\left[\mathbf{r}_{g,n}(\overline{\boldsymbol{\rho}}_g)-\mathbf{r}_{g,n}(\boldsymbol{\rho}_{g,n})\right]'\boldsymbol{\Gamma}'_{g,n}\boldsymbol{\Upsilon}_n\boldsymbol{\Gamma}_{g,n}\left[\mathbf{r}_{g,n}(\overline{\boldsymbol{\rho}}_g)-\mathbf{r}_{g,n}(\boldsymbol{\rho}_{g,n})\right]$$

$$\geq\lambda_{\min}(\boldsymbol{\Upsilon}_n)\lambda_{\min}(\boldsymbol{\Gamma}'_{g,n}\boldsymbol{\Gamma}_{g,n})\left[\mathbf{r}_{g,n}(\overline{\boldsymbol{\rho}}_g)-\mathbf{r}_{g,n}(\boldsymbol{\rho}_{g,n})\right]'\left[\mathbf{r}_{g,n}(\overline{\boldsymbol{\rho}}_g)-\mathbf{r}_{g,n}(\boldsymbol{\rho}_{g,n})\right]$$

$$\geq\lambda_*\sum_{s=1}^{q}\left[\overline{\rho}_{g,s}-\rho_{g,s,n}\right]^2$$

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

$$\inf_{\overline{\boldsymbol{\rho}}_g:\sum_{r\in\mathbf{I}_{g,\rho}}|\overline{\rho}_{g,r}|\in[-a_\rho,a_\rho]\text{ and }\|\overline{\boldsymbol{\rho}}_g-\boldsymbol{\rho}_{g,n}\|\geq\varepsilon}[\overline{R}_n(\overline{\boldsymbol{\rho}}_g)-\overline{R}_n(\boldsymbol{\rho}_{g,n})]$$

$$\geq\inf_{\overline{\boldsymbol{\rho}}_g:\sum_{r\in\mathbf{I}_{g,\rho}}|\overline{\rho}_{g,r}|\in[-a_\rho,a_\rho]\text{ and }\|\overline{\boldsymbol{\rho}}_g-\boldsymbol{\rho}_{g,n}\|\geq\varepsilon}\lambda_*\sum_{r\in\mathbf{I}_{g,\rho}}\left[\overline{\rho}_{g,r}-\rho_{g,r,n}\right]^2=\lambda_*\varepsilon^2>0,$$

which proves that $\boldsymbol{\rho}_{g,n}$ is identifiably unique. Next let $\boldsymbol{\Phi}_n = [\boldsymbol{\Gamma}_n, -\boldsymbol{\gamma}_n]$ and $\widetilde{\boldsymbol{\Phi}}_n = [\widetilde{\boldsymbol{\Gamma}}_n, -\widetilde{\boldsymbol{\gamma}}_n]$, then

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

Next, observe that the elements of $\boldsymbol{\Phi}_n$ and $\widetilde{\boldsymbol{\Phi}}_n$ are all of the form $n^{-1} E\mathbf{u}_n' A_n \mathbf{u}_n$ and $n^{-1} \widetilde{\mathbf{u}}_n' A_n \widetilde{\mathbf{u}}_n$, 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 $\widetilde{\mathbf{u}}_{g,n} - \mathbf{u}_{g,n} = -\mathbf{Z}_{g,n}(\widetilde{\boldsymbol{\delta}}_{g,n} - \boldsymbol{\delta}_{g,n})$. 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 $4 + \upsilon$ absolute moments for some $\upsilon > 0$. 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 $n^{1/2}(\widetilde{\boldsymbol{\delta}}_{g,n} - \boldsymbol{\delta}_{g,n}) = O_p(1)$. It now follows immediately from Lemma A.7(a) that $\widetilde{\boldsymbol{\Phi}}_n - \boldsymbol{\Phi}_n \xrightarrow{p} 0$, that the elements of $\boldsymbol{\Phi}_n$ are $O(1)$ and, consequently, that the elements of $\widetilde{\boldsymbol{\Phi}}_n$ are $O_p(1)$. The elements of $\widetilde{\boldsymbol{\Upsilon}}_n$ and $\boldsymbol{\Upsilon}_n$ have the analogous properties in light of condition (b) in the theorem. Given this, it follows from the above inequality that $R_n(\omega, \overline{\boldsymbol{\rho}}_g) - \overline{R}_n(\overline{\boldsymbol{\rho}}_g)$ converges to zero uniformly over the optimization space, i.e.,

$$
\sup_{\overline{\boldsymbol{\rho}}_g : \sum_{r \in \mathbf{I}_{g,\rho}} |\overline{\rho}_{g,r}| \in [-a_\rho, a_\rho]} \left| R_n(\omega, \overline{\boldsymbol{\rho}}_g) - \overline{R}_n(\overline{\boldsymbol{\rho}}_g) \right|
$$
$$
\leq \left\| \widetilde{\boldsymbol{\Phi}}_n' \widetilde{\boldsymbol{\Upsilon}}_n \widetilde{\boldsymbol{\Phi}}_n - \boldsymbol{\Phi}_n' \boldsymbol{\Upsilon}_n \boldsymbol{\Phi}_n \right\| \left\{ 1 + [2q + q(q-1)/2] a_\rho^4 \right\} \xrightarrow{p} 0
$$

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

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

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

Substituting the mean value theorem expression

$$
\mathbf{m}_{g,n}^\rho(\widetilde{\boldsymbol{\rho}}_{g,n}, \widetilde{\boldsymbol{\delta}}_{g,n}) = \mathbf{m}_{g,n}^\rho(\boldsymbol{\rho}_{g,n}, \widetilde{\boldsymbol{\delta}}_{g,n}) + \frac{\partial \mathbf{m}_{g,n}^\rho(\check{\boldsymbol{\rho}}_{g,n}, \widetilde{\boldsymbol{\delta}}_{g,n})}{\partial \boldsymbol{\rho}_g} (\widetilde{\boldsymbol{\rho}}_{g,n} - \boldsymbol{\rho}_{g,n}),
$$

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

$$\frac{\partial \mathbf{m}_{g,n}^{\rho}(\widetilde{\boldsymbol{\rho}}_{g,n}, \widetilde{\boldsymbol{\delta}}_{g,n})}{\partial \boldsymbol{\rho}_g'} \widetilde{\boldsymbol{\Upsilon}}_{g,n} \frac{\partial \mathbf{m}_{g,n}^{\rho}(\check{\boldsymbol{\rho}}_{g,n}, \widetilde{\boldsymbol{\delta}}_{g,n})}{\partial \boldsymbol{\rho}_g} n^{1/2}(\widetilde{\boldsymbol{\rho}}_{g,n} - \boldsymbol{\rho}_{g,n}) \tag{B.1}$$

$$= -\frac{\partial \mathbf{m}_{g,n}^{\rho}(\widetilde{\boldsymbol{\rho}}_{g,n}, \widetilde{\boldsymbol{\delta}}_{g,n})}{\partial \boldsymbol{\rho}_g'} \widetilde{\boldsymbol{\Upsilon}}_{g,n} \mathbf{m}_{g,n}^{\rho}(\boldsymbol{\rho}_{g,n}, \widetilde{\boldsymbol{\delta}}_{g,n}).$$

Observe that

$$\frac{\partial \mathbf{m}_{g,n}^{\rho}(\boldsymbol{\rho}_g, \widetilde{\boldsymbol{\delta}}_{g,n})}{\partial \boldsymbol{\rho}_g} = -\widetilde{\boldsymbol{\Gamma}}_{g,n} \frac{\partial \mathbf{r}_{g,n}(\boldsymbol{\rho}_g)}{\partial \boldsymbol{\rho}_g} \tag{B.2}$$

and consider

$$\widetilde{\boldsymbol{\Xi}}_{g,n} = \frac{\partial \mathbf{m}_{g,n}^{\rho}(\widetilde{\boldsymbol{\rho}}_{g,n}, \widetilde{\boldsymbol{\delta}}_{g,n})}{\partial \boldsymbol{\rho}_g'} \widetilde{\boldsymbol{\Upsilon}}_{g,n} \frac{\partial \mathbf{m}_{g,n}^{\rho}(\check{\boldsymbol{\rho}}_{g,n}, \widetilde{\boldsymbol{\delta}}_{g,n})}{\partial \boldsymbol{\rho}_g} \tag{B.3}$$

$$= \frac{\partial \mathbf{r}_{g,n}(\widetilde{\boldsymbol{\rho}}_{g,n})}{\partial \boldsymbol{\rho}_g'} \widetilde{\boldsymbol{\Gamma}}_{g,n}' \widetilde{\boldsymbol{\Upsilon}}_{g,n} \widetilde{\boldsymbol{\Gamma}}_{g,n} \frac{\partial \mathbf{r}_{g,n}(\check{\boldsymbol{\rho}}_{g,n})}{\partial \boldsymbol{\rho}_g},$$

$$\boldsymbol{\Xi}_{g,n} = \frac{\partial \mathbf{r}_{g,n}(\boldsymbol{\rho}_{g,n})}{\partial \boldsymbol{\rho}_g'} \boldsymbol{\Gamma}_{g,n}' \boldsymbol{\Upsilon}_{g,n} \boldsymbol{\Gamma}_{g,n} \frac{\partial \mathbf{r}_{g,n}(\boldsymbol{\rho}_{g,n})}{\partial \boldsymbol{\rho}_g}.$$

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

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

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

$$\lambda_{\min}(\boldsymbol{\Xi}_{g,n}) \geq \lambda_{\min}(\boldsymbol{\Upsilon}_{g,n}) \lambda_{\min}(\boldsymbol{\Gamma}_{g,n}' \boldsymbol{\Gamma}_{g,n}) \lambda_{\min} \left\{ \frac{\partial \mathbf{r}_{g,n}(\boldsymbol{\rho}_{g,n})}{\partial \overline{\boldsymbol{\rho}}_g'} \frac{\partial \mathbf{r}_{g,n}(\boldsymbol{\rho}_{g,n})}{\partial \overline{\boldsymbol{\rho}}_g} \right\} \geq \lambda_{\Xi}^*$$

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

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

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

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

as $n \to \infty$.

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

$$n^{1/2}(\widetilde{\boldsymbol{\rho}}_{g,n} - \boldsymbol{\rho}_{g,n}) = \left[ I - \widetilde{\boldsymbol{\Xi}}_{g,n}^{+} \widetilde{\boldsymbol{\Xi}}_{g,n} \right] n^{1/2}(\widetilde{\boldsymbol{\rho}}_{g,n} - \boldsymbol{\rho}_{g,n})$$
$$- \widetilde{\boldsymbol{\Xi}}_{g,n}^{+} \frac{\partial \mathbf{m}_{g,n}^{\rho}(\widetilde{\boldsymbol{\rho}}_{g,n}, \widetilde{\boldsymbol{\delta}}_{g,n})}{\partial \boldsymbol{\rho}_g'} \widetilde{\boldsymbol{\Upsilon}}_{g,n} n^{1/2} \mathbf{m}_{g,n}^{\rho}(\boldsymbol{\rho}_{g,n}, \widetilde{\boldsymbol{\delta}}_{g,n}).$$

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

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

Observe that

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

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

$$n^{-1/2} \widetilde{\mathbf{u}}_{g,n}' \left[ \mathbf{I}_n - \mathbf{R}_{g,n}^{*\prime}(\boldsymbol{\rho}_{g,n}) \right] \mathbf{A}_{s,n} \left[ \mathbf{I}_n - \mathbf{R}_{g,n}^{*}(\boldsymbol{\rho}_{g,n}) \right] \widetilde{\mathbf{u}}_{g,n}$$

with

$$\left[ \mathbf{I}_n - \mathbf{R}_{g,n}^{*}(\boldsymbol{\rho}_{g,n}) \right] \mathbf{u}_{g,n} = \boldsymbol{\varepsilon}_{g,n},$$
$$\widetilde{\mathbf{u}}_{g,n} - \mathbf{u}_{g,n} = -\mathbf{Z}_{g,n}(\widetilde{\boldsymbol{\delta}}_{g,n} - \boldsymbol{\delta}_{g,n}),$$
$$n^{1/2}(\widetilde{\boldsymbol{\delta}}_{g,n} - \boldsymbol{\delta}_{g,n}) = n^{-1/2} \sum_{h=1}^{G} \mathbf{T}_{gh,n}' \boldsymbol{\varepsilon}_{h,n} + o_p(1).$$

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

$$n^{-1/2} \widetilde{\mathbf{u}}_{g,n}' \left[ \mathbf{I}_n - \mathbf{R}_{g,n}^{*\prime}(\boldsymbol{\rho}_{g,n}) \right] \mathbf{A}_{s,n} \left[ \mathbf{I}_n - \mathbf{R}_{g,n}^{*}(\boldsymbol{\rho}_{g,n}) \right] \widetilde{\mathbf{u}}_{g,n}$$
$$= n^{-1/2} \frac{1}{2} \boldsymbol{\varepsilon}_{g,n}' (\mathbf{A}_{s,n} + \mathbf{A}_{s,n}') \boldsymbol{\varepsilon}_{g,n} + n^{-1/2} \boldsymbol{\alpha}_{g,s,n}' \sum_{h=1}^{G} \mathbf{T}_{gh,n}' \boldsymbol{\varepsilon}_{h,n} + o_p(1),$$

where $\boldsymbol{\alpha}_{g,s,n} = -n^{-1} E \mathbf{Z}_{g,n}' \left[ \mathbf{I}_n - \mathbf{R}_{g,n}^{*\prime}(\boldsymbol{\rho}_{g,n}) \right] (\mathbf{A}_{s,n} + \mathbf{A}_{s,n}') \boldsymbol{\varepsilon}_{g,n}$. 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

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

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

$$\psi_{rs,gg,n}^{\rho\rho} = \frac{1}{2n}\sigma_{gg}^2 tr\left[(\mathbf{A}_{r,n} + \mathbf{A}_{r,n}')(\mathbf{A}_{s,n} + \mathbf{A}_{s,n}')\right] \tag{B.9}$$

$$+ \frac{1}{n}\boldsymbol{\alpha}_{g,r,n}'\left[\sum_{h=1}^{G}\sum_{l=1}^{G}\sigma_{hl}\mathbf{T}_{gh,n}'\mathbf{T}_{gl,n}\right]\boldsymbol{\alpha}_{g,s,n}.$$

We note that $\lambda_{\min}(\boldsymbol{\Psi}_{gg,n}^{\rho\rho}) \geq const > 0$ 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

$$\boldsymbol{\xi}_{g,n} = -(\boldsymbol{\Psi}_{gg,n}^{\rho\rho})^{-1/2}n^{-1/2}\begin{bmatrix}\frac{1}{2}\boldsymbol{\varepsilon}_{g,n}'(\mathbf{A}_{1,n}+\mathbf{A}_{1,n}')\boldsymbol{\varepsilon}_{g,n}+\boldsymbol{\alpha}_{g,1,n}'\sum_{h=1}^{G}\mathbf{T}_{gh,n}'\boldsymbol{\varepsilon}_{h,n}\\ \vdots \\ \frac{1}{2}\boldsymbol{\varepsilon}_{g,n}'(\mathbf{A}_{S,n}+\mathbf{A}_{S,n}')\boldsymbol{\varepsilon}_{g,n}+\boldsymbol{\alpha}_{g,S,n}'\sum_{h=1}^{G}\mathbf{T}_{gh,n}'\boldsymbol{\varepsilon}_{h,n}\end{bmatrix}$$

$$\xrightarrow{d} N(\mathbf{0},\mathbf{I}_S). \tag{B.10}$$

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

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

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 $n^{1/2}(\widetilde{\boldsymbol{\rho}}_{g,n} - \boldsymbol{\rho}_{g,n}) = O_p(1)$. 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 $\lambda_{\min}(\boldsymbol{\Xi}_{g,n}^{-1}) = 1/\lambda_{\max}(\boldsymbol{\Xi}_{g,n})$. Hence

$$\lambda_{\min}\left\{\boldsymbol{\Xi}_{g,n}^{-1}\mathbf{J}_{g,n}'\boldsymbol{\Upsilon}_{g,n}\boldsymbol{\Psi}_{gg,n}^{\rho\rho}\boldsymbol{\Upsilon}_{g,n}\mathbf{J}_{g,n}\boldsymbol{\Xi}_{g,n}^{-1}\right\}$$

$$\geq \lambda_{\min}\left(\boldsymbol{\Psi}_{gg,n}^{\rho\rho}\right)\lambda_{\min}\left(\boldsymbol{\Upsilon}_{g,n}\right)\lambda_{\min}(\boldsymbol{\Xi}_{g,n}^{-1}\mathbf{J}_{g,n}'\boldsymbol{\Upsilon}_{g,n}\mathbf{J}_{g,n}\boldsymbol{\Xi}_{g,n}^{-1})$$

$$\geq \lambda_{\min}\left(\boldsymbol{\Psi}_{gg,n}^{\rho\rho}\right)\lambda_{\min}\left(\boldsymbol{\Upsilon}_{g,n}\right)/\lambda_{\Xi}^{**} \geq const > 0.$$

This establishes the last claim of the theorem.                                    □

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

$$\widetilde{\boldsymbol{\varepsilon}}_{g,n} = \left[\mathbf{I}_n - \mathbf{R}_{g,n}^*(\widetilde{\boldsymbol{\rho}}_{g,n})\right]\widetilde{\mathbf{u}}_{g,n}$$

$$= \boldsymbol{\varepsilon}_{g,n} - \mathbf{R}_{g,n}^*(\widetilde{\boldsymbol{\rho}}_{g,n} - \boldsymbol{\rho}_{g,n})\mathbf{u}_{g,n} - \left[\mathbf{I}_n - \mathbf{R}_{g,n}^*(\widetilde{\boldsymbol{\rho}}_{g,n})\right]\mathbf{Z}_{g,n}\boldsymbol{\Delta}_{g,n}.$$

Consequently

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

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 $\boldsymbol{\varepsilon}_{h,n}=\left[\mathbf{I}_n-\mathbf{R}^*_{g,n}(\boldsymbol{\rho}_{g,n})\right]\mathbf{u}_{g,n}$, 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 $\widetilde{\sigma}_{gh,n}-\sigma_{gh}=o_p(1)$ as claimed. $\qquad\square$

**Proof of Theorem 3.** We first demonstrate that $\widetilde{\boldsymbol{\Psi}}^{\rho\rho}_{gg,n}-\boldsymbol{\Psi}^{\rho\rho}_{gg,n}=o_p(1)$, where the elements of $\widetilde{\boldsymbol{\Psi}}^{\rho\rho}_{gg,n}$ and $\boldsymbol{\Psi}^{\rho\rho}_{gg,n}$ are defined in (18) and (24). By Lemma 1, we have $\widetilde{\sigma}_{gh,n}-\sigma_{gh}=o_p(1)$. Furthermore, by assumption $n^{-1}\widetilde{\mathbf{T}}'_{gh,n}\widetilde{\mathbf{T}}_{gl,n}-n^{-1}\mathbf{T}'_{gh,n}\mathbf{T}_{gl,n}=o_p(1)$, 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 $\widetilde{\boldsymbol{\alpha}}_{g,r,n}-\boldsymbol{\alpha}_{g,r,n}=o_p(1)$, where $\boldsymbol{\alpha}_{g,r,n}=O(1)$ and $\widetilde{\boldsymbol{\alpha}}_{g,r,n}=O_p(1)$. Hence, clearly $\widetilde{\psi}^{\rho\rho}_{rs,gg,n}-\psi^{\rho\rho}_{rs,gg,n}=o_p(1)$, $\psi^{\rho\rho}_{rs,gg,n}=O(1)$ and $\widetilde{\psi}^{\rho\rho}_{rs,gg,n}=O_p(1)$, as well as $\widetilde{\boldsymbol{\Psi}}^{\rho\rho}_{gg,n}-\boldsymbol{\Psi}^{\rho\rho}_{gg,n}=o_p(1)$, $\boldsymbol{\Psi}^{\rho\rho}_{gg,n}=O(1)$ and $\widetilde{\boldsymbol{\Psi}}^{\rho\rho}_{gg,n}=O_p(1)$. Observing that $\lambda_{\min}(\boldsymbol{\Psi}^{\rho\rho}_{gg,n})\geq c^*_{\Psi}>0$ it follows further that $(\widetilde{\boldsymbol{\Psi}}^{\rho\rho}_{gg,n})^{-1}-(\boldsymbol{\Psi}^{\rho\rho}_{gg,n})^{-1}=o_p(1)$, $(\boldsymbol{\Psi}^{\rho\rho}_{gg,n})^{-1}=O(1)$ and $(\widetilde{\boldsymbol{\Psi}}^{\rho\rho}_{gg,n})^{-1}=O_p(1)$.

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

We shall utilize the following lemma.

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

$$E(\boldsymbol{\varepsilon}'\mathbf{A}\varepsilon + \mathbf{a}'\boldsymbol{\varepsilon})=\mathbf{0}$$

$$cov(\boldsymbol{\varepsilon}'\mathbf{A}\boldsymbol{\varepsilon} + \mathbf{a}'\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon}'\mathbf{B}\boldsymbol{\varepsilon} + \mathbf{b}'\boldsymbol{\varepsilon})=2tr\left[\mathbf{A}(\boldsymbol{\Sigma}\otimes\mathbf{I})\mathbf{B}(\boldsymbol{\Sigma}\otimes\mathbf{I})\right] + \mathbf{a}'(\boldsymbol{\Sigma}\otimes\mathbf{I})\mathbf{b}$$

$$= 2\sum_{h=1}^{G}\sum_{l=1}^{G}\sigma_{hl}^2 tr\left[\mathbf{A}_{hh}\mathbf{B}_{ll}\right] + \sum_{h=1}^{G}\sum_{l=1}^{G}\sigma_{hl}\mathbf{a}_h'\mathbf{b}_l.$$

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

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

Next, observe that $\xi_{gs,n} = n^{-1/2}\left[\boldsymbol{\varepsilon}_n'\mathbf{B}\boldsymbol{\varepsilon}_n + \mathbf{b}'\boldsymbol{\varepsilon}_n\right]$ with $\mathbf{B} = diag(\mathbf{0}, \ldots, \mathbf{B}_{gg}, \ldots, \mathbf{0})$, 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 $\mathbf{T}_{g,n} = [\mathbf{T}_{g1,n}', \ldots, \mathbf{T}_{gG,n}']'$. It now follows from Lemma B.1 that

$$cov(\boldsymbol{\eta}_n, \xi_{gs,n}) = cov(n^{-1/2}\mathbf{T}_n'\boldsymbol{\varepsilon}_n, n^{-1/2}\left[\boldsymbol{\varepsilon}_n'\mathbf{B}\boldsymbol{\varepsilon}_n + \mathbf{b}'\boldsymbol{\varepsilon}_n\right]) = cov(n^{-1/2}\boldsymbol{\eta}_n, n^{-1/2}\mathbf{b}'\boldsymbol{\varepsilon}_n)$$

$$= n^{-1}\mathbf{T}_n'(\boldsymbol{\Sigma}\otimes\mathbf{I}_n)\mathbf{T}_{g,n}\boldsymbol{\alpha}_{g,s,n},$$

and thus

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

and

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

as claimed.

Define $\xi_{gr,n} = n^{-1/2}\left[\boldsymbol{\varepsilon}_n'\mathbf{A}\varepsilon_n + \mathbf{a}'\boldsymbol{\varepsilon}_n\right]$ with $\mathbf{A} = diag(\mathbf{0}, \ldots, \mathbf{A}_{gg}, \ldots, \mathbf{0})$ 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

$$cov(\xi_{gr,n}, \xi_{hs,n}) = \sigma_{gh,n}^2 (2n)^{-1}tr\left[\left(\mathbf{A}_{r,n} + \mathbf{A}_{r,n}'\right)\left(\mathbf{A}_{s,n} + \mathbf{A}_{s,n}'\right)\right]$$

$$+ \boldsymbol{\alpha}_{g,r,n}'\left[\sum_{u=1}^{G}\sum_{v=1}^{G}\sigma_{uv,n}\mathbf{T}_{gu,n}'\mathbf{T}_{hv,n}\right]\boldsymbol{\alpha}_{h,s,n},$$

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

In light of Assumption 10 and Theorem 2, we have

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

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

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

which proves the first part of the theorem.

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

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

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

$$\lambda_{\min}(\boldsymbol{\Omega}_n) \geq \lambda_{\min}(\boldsymbol{\Psi}_n)\lambda_{\min}\left\{ diag_g[(\mathbf{J}_{g,n}'\boldsymbol{\Upsilon}_{g,n}\mathbf{J}_{g,n})^{-1}] \right\} \geq const > 0$$

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

## C. Appendix: Proofs for Section 5

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

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

**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 $\widehat{\boldsymbol{\delta}}_n = (\widehat{\boldsymbol{\delta}}_{1,n}', \ldots, \widehat{\boldsymbol{\delta}}_{G,n}')'$ and $\widehat{\boldsymbol{\rho}}_n = (\widehat{\boldsymbol{\rho}}_{1,n}', \ldots, \widehat{\boldsymbol{\rho}}_{G,n}')'$. The limiting distribution of $\widehat{\boldsymbol{\delta}}_{g,n}$ and $\widehat{\boldsymbol{\rho}}_{g,n}$ 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 $\widehat{\boldsymbol{\delta}}_{g,n}$) In light of Lemma A.5, we have $n^{1/2}[\widehat{\boldsymbol{\delta}}_{g,n} - \boldsymbol{\delta}_{g,n}] = n^{-1/2}\mathbf{T}'_{gg,n}\boldsymbol{\varepsilon}_{g,n} + o_p(1)$ with

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

and thus

$$n^{-1}\mathbf{T}'_{gg,n}\mathbf{T}_{hh,n} = [\mathbf{Q}'_{\mathbf{HZ},g*}(\boldsymbol{\rho}_{g,n})\mathbf{Q}_{\mathbf{HH}}^{-1}\mathbf{Q}_{\mathbf{HZ},g*}(\boldsymbol{\rho}_{g,n})]^{-1}\mathbf{Q}'_{\mathbf{HZ},g*}(\boldsymbol{\rho}_{g,n})\mathbf{Q}_{\mathbf{HH}}^{-1}$$
$$\times (n^{-1}\mathbf{H}'_n\mathbf{H}_n)\mathbf{Q}_{\mathbf{HH}}^{-1}\mathbf{Q}_{\mathbf{HZ},h*}(\boldsymbol{\rho}_{h,n})[\mathbf{Q}'_{\mathbf{HZ},h*}(\boldsymbol{\rho}_{h,n})\mathbf{Q}_{\mathbf{HH}}^{-1}\mathbf{Q}_{\mathbf{HZ},h*}(\boldsymbol{\rho}_{h,n})]^{-1}.$$

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

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

$$\boldsymbol{\Psi}_{gh,n}^{\delta\delta} = \sigma_{gh}n^{-1}\mathbf{T}'_{gg,n}\mathbf{T}_{hh,n},$$
$$\boldsymbol{\Psi}_{gh,n}^{\delta\rho} = \sigma_{gh}n^{-1}\mathbf{T}'_{gg,n}\mathbf{T}_{hh,n}[\boldsymbol{\alpha}_{h,1,n}, \ldots, \boldsymbol{\alpha}_{h,S,n}],$$

and the elements of $\boldsymbol{\Psi}_{gh,n}^{\rho\rho}$, i.e., the elements of the $(g,h)$th block of $\boldsymbol{\Psi}_n^{\rho\rho}$ are given by

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

Now consider the estimator

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

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

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

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 $\widehat{\boldsymbol{\varepsilon}}_{g,n} = \mathbf{y}_{*g,n}(\widetilde{\boldsymbol{\rho}}_{g,n}) - \mathbf{Z}_{*g,n}(\widetilde{\boldsymbol{\rho}}_{g,n})\widehat{\boldsymbol{\delta}}_{g,n}$. Then the components of the estimator for $\boldsymbol{\Psi}_n$ defined in (27) simplify to

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

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

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

Note that

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

and

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

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

**Part 4**: (Limiting Distribution of $\widehat{\boldsymbol{\delta}}_{g,n}$ and $\widehat{\boldsymbol{\rho}}_{g,n}$) 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 $\boldsymbol{\Omega}_n$ given in (28) simplify to:

$$\boldsymbol{\Omega}_n = \begin{bmatrix} \boldsymbol{\Omega}^{\delta\delta}_n & \boldsymbol{\Omega}^{\delta\rho}_n \\ \boldsymbol{\Omega}^{\delta\rho'}_n & \boldsymbol{\Omega}^{\rho\rho}_n \end{bmatrix}$$

with

$$\boldsymbol{\Omega}^{\delta\delta}_n = \boldsymbol{\Psi}^{\delta\delta}_n,$$
$$\boldsymbol{\Omega}^{\delta\rho}_n = \boldsymbol{\Psi}^{\delta\rho}_n \, diag^G_{g=1}\left((\boldsymbol{\Psi}^{\rho\rho}_{gg,n})^{-1}\mathbf{J}_{g,n}(\mathbf{J}'_{g,n}(\boldsymbol{\Psi}^{\rho\rho}_{gg,n})^{-1}\mathbf{J}_{g,n})^{-1}\right),$$

$$\boldsymbol{\Omega}^{\rho\rho}_n = diag^G_{g=1}\left((\mathbf{J}'_{g,n}(\boldsymbol{\Psi}^{\rho\rho}_{gg,n})^{-1}\mathbf{J}_{g,n})^{-1}\mathbf{J}'_{g,n}(\boldsymbol{\Psi}^{\rho\rho}_{gg,n})^{-1}\right) \begin{bmatrix} \boldsymbol{\Psi}^{\rho\rho}_{11,n} & \cdots & \boldsymbol{\Psi}^{\rho\rho}_{1G,n} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Psi}^{\rho\rho}_{G1,n} & \cdots & \boldsymbol{\Psi}^{\rho\rho}_{GG,n} \end{bmatrix}$$
$$\times diag^G_{g=1}\left((\boldsymbol{\Psi}^{\rho\rho}_{gg,n})^{-1}\mathbf{J}_{g,n}(\mathbf{J}'_{g,n}(\boldsymbol{\Psi}^{\rho\rho}_{gg,n})^{-1}\mathbf{J}_{g,n})^{-1}\right).$$

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

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

with

$$\boldsymbol{\Omega}^{\delta\delta}_{gg,n} = \boldsymbol{\Psi}^{\delta\delta}_{gg,n} = \sigma_{gg}n^{-1}\mathbf{T}'_{gg,n}\mathbf{T}_{gg,n},$$
$$\boldsymbol{\Omega}^{\delta\rho}_{gg,n} = \boldsymbol{\Psi}^{\delta\rho}_{gg,n}(\boldsymbol{\Psi}^{\rho\rho}_{gg,n})^{-1}\mathbf{J}_{g,n}(\mathbf{J}'_{g,n}(\boldsymbol{\Psi}^{\rho\rho}_{gg,n})^{-1}\mathbf{J}_{g,n})^{-1},$$
$$= \sigma_{gg}n^{-1}\mathbf{T}'_{gg,n}\mathbf{T}_{gg,n}[\boldsymbol{\alpha}_{g,1,n}, \ldots, \boldsymbol{\alpha}_{g,S,n}](\boldsymbol{\Psi}^{\rho\rho}_{gg,n})^{-1}\mathbf{J}_{g,n}(\mathbf{J}'_{g,n}(\boldsymbol{\Psi}^{\rho\rho}_{gg,n})^{-1}\mathbf{J}_{g,n})^{-1},$$
$$\boldsymbol{\Omega}^{\rho\rho}_{gg,n} = (\mathbf{J}'_{g,n}(\boldsymbol{\Psi}^{\rho\rho}_{gg,n})^{-1}\mathbf{J}_{g,n})^{-1}.$$

Observe from part 3 of the proof that

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

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.     □

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

$$\mathbf{T}_n = (\mathbf{I}_G \otimes \mathbf{H}_n)\mathbf{P}_n,$$
$$\mathbf{P}_n = \left[\boldsymbol{\Sigma}^{-1} \otimes \mathbf{Q}_{\mathbf{HH}}^{-1}\right]diag_{g=1}^{G}\left[\mathbf{Q}_{\mathbf{HZ},g*}(\boldsymbol{\rho}_{g,n})\right]$$
$$\times \left\{diag_{g=1}^{G}\left[\mathbf{Q}'_{\mathbf{HZ},g*}(\boldsymbol{\rho}_{g,n})\right]\left[\boldsymbol{\Sigma}^{-1} \otimes \mathbf{Q}_{\mathbf{HH}}^{-1}\right]diag_{g=1}^{G}\left[\mathbf{Q}_{\mathbf{HZ},g*}(\boldsymbol{\rho}_{g,n})\right]\right\}^{-1}. \quad \textbf{(C.2)}$$

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 formulas 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

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

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

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

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

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

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 $\widetilde{\Upsilon}_{g,n} = (\widehat{\widehat{\Psi}}_{gg,n}^{\rho\rho})^{-1}$ with $\widehat{\widehat{\Psi}}_{gg,n}^{\rho\rho} = (\widehat{\widehat{\psi}}_{rs,gg,n}^{\rho\rho}))$ 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 $\widehat{\widehat{\rho}}_{g,n} - \rho_{g,n} = o_p(1)$. It now follows directly from Theorem 3 that $\widehat{\widehat{\Psi}}_{gg,n}^{\rho\rho} - \Psi_{gg,n}^{\rho\rho} = o_p(1)$, $(\widehat{\widehat{\Psi}}_{gg,n}^{\rho\rho})^{-1} - (\Psi_{gg,n}^{\rho\rho})^{-1} = o_p(1)$, 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 $\widehat{\widehat{\delta}}_{g,n}$ and $\widehat{\widehat{\rho}}_{g,n}$) 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

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

where

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

with

$$\Psi_n^{\delta\rho} = \Psi_n^{\delta\delta} diag_{g=1}^G[\alpha_{g,1,n}, \ldots, \alpha_{g,S,n}],$$

$$\mathfrak{J}_{g,n} = (\Psi_n^{\rho\rho})^{-1}\mathbf{J}_{g,n}\left[\mathbf{J}_{g,n}'(\Psi_{gg,n}^{\rho\rho})^{-1}\mathbf{J}_{g,n}\right]^{-1}.$$

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. $\square$

## SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit: 10.1017/S026646662200007X.

## *REFERENCES*

Anselin, L. (1988) *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers.

Anselin, L. (2010) Thirty years of spatial econometrics. *Papers in Regional Science* 89, 3–25.

Anselin, L. & R. Florax (1995) *New Directions in Spatial Econometrics*. Springer.

Audretsch, D.B. & M.P. Feldmann (1996) R&D spillovers and the geography of innovation and production. *American Economic Review* 86, 630–640.

Ballester, C., A. Calvó-Armengol, & Y. Zenou (2006) Who's who in networks. Wanted: The key player. *Econometrica* 74, 1403–1417.

Baltagi, B.H. & Y. Deng (2015) EC3SLS estimator for a simultaneous system of spatial autoregressive equations with random effects. *Econometric Reviews* 34, 659–694.

Baltagi, B.H., P. Egger, & M. Pfaffermayr (2007) Estimating models of complex FDI: Are there third-country effects? *Journal of Econometrics* 140, 260–281.

Baltagi, B.H., P. Egger, & M. Pfaffermayr (2008) Estimating regional trade agreement effects on FDI in an interdependent world. *Journal of Econometrics* 145, 194–208.

Behrens, K., W. Koch, & C. Ertur (2012) Dual gravity: Using spatial econometrics to control for multilateral resistance. *Journal of Applied Econometrics* 27, 773–794.

Bell, K.P. & N.E. Bockstael (2000) Applying the generalized-moments estimation approach to spatial problems involving microlevel data. *Review of Economics and Statistics* 82, 72–82.

Blonigen, B.A., R.B. Davies, G.R. Waddell, & H.T. Naughton (2007) FDI in space: Spatial autoregressive relationships in foreign direct investment. *European Economic Review* 51, 1303–1325.

Blume, L.E., W.A. Brock, S.N. Durlauf, & Y.M. Ioannides (2011) Identification of social interactions. In J. Benhabib, A. Bisin, & M.O. Jackson (eds.), *Handbook of Social Economics*, vol. 1B, pp. 853–964. Elsevier Publisher.

Calvó-Armengol, A., E. Patacchini, & Y. Zenou (2009) Peer effects and social networks in education. *Review of Economic Studies* 76, 1239–1267.

Case, A., J. HinesJr., & H. Rosen (1993) Budget spillovers and fiscal policy independence: Evidence from the States. *Journal of Public Economics* 52, 285–307.

Cliff, A. & J. Ord (1973) *Spatial Autocorrelation*. Pion.

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

Cohen, J.P. & 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.

Cohen-Cole, E., X. Liu, & Y. Zenou (2018) Multivariate choice and identification of social interactions. *Journal of Applied Econometrics* 33, 165–178.

Conley, T.G. & E. Ligon (2002) Economic distance and cross-country spillovers. *Journal of Economic Growth* 7, 157–187.

Cressie, N.A.C. (1993) *Statistics of Spatial Data*. Wiley.

Das, D., H.H. Kelejian, & I.R. Prucha (2003) Finite sample properties of estimators of spatial autoregressive models with autoregressive disturbances. *Papers in Regional Science* 82, 1–26.

Devereux, M.P., B. Lockwood, & M. Redoano (2008) Do countries compete over corporate taxes? *Journal of Public Economics* 92, 1210–1235.

Egger, P., M. Pfaffermayr, & H. Winner (2005) Commodity taxation in a 'linear' world: A spatial panel data approach. *Regional Science and Urban Economics* 35, 527–541.

Ertur, C. & W. Koch (2007) Growth, technological interdependence and spatial externalities: Theory and evidence. *Journal of Applied Econometrics* 22, 1033–1062.

Holtz-Eakin, D. (1994) Public sector capital and the productivity puzzle. *Review of Economics and Statistics* 76, 12–21.

Horn, R.A. & C.R. Johnson (1985) *Matrix Analysis*. Cambridge University Press.

Kapoor, M., H.H. Kelejian, & I.R. Prucha (2007) Panel data models with spatially correlated error components. *Journal of Econometrics* 140, 97–130.

Kelejian, H.H. & 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.

Kelejian, H.H. & I.R. Prucha (1999) A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review* 40, 509–533.

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

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

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

Lee, J., P.C.B. Phillips, & F. Rossi (2021) Specification Testing in Spatial Autoregression. Working Paper.

Lee, L.-F. (2003) Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. *Econometric Reviews* 22, 307–335.

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

Lee, L.-F. & X. Liu (2010) Efficient GMM estimation of high order spatial autoregrssive models with autoregressive disturbances. *Econometric Theory* 26, 187–230.

Liu, X. (2014) Identification and efficient estimation of simultaneous equations network models. *Journal of Business & Economic Statistics* 32, 516–536.

Liu, X. (2019) Simultaneous equations with binary outcomes and social interactions. *Econometric Reviews* 38, 921–937.

Liu, X. (2020) GMM identification and estimation of peer effects in a system of simultaneous equations. *Journal of Spatial Econometrics* 1, 2–27.

Liu, X. & P. Saraiva (2019) GMM estimation of spatial autoregressive models in a system of simultaneous equations with heteroskedasticity. *Econometric Reviews* 38, 359–385.

Pinkse, J. & M.E. Slade (1998) Contracting in space: An application of spatial statistics to discrete-choice models. *Journal of Econometrics* 85, 125–154.

Pinkse, J., M.E. Slade, & C. Brett (2002) Spatial price competition: A semiparametric approach. *Econometrica* 70, 1111–1153.

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

Wang, W., L.-F. Lee, & Y. Bao (2018) GMM estimation of the spatial autoregressive model in a system of interrelated networks. *Regional Science and Urban Economics* 69, 167–198.

Wang, L., K. Li, & Z. Wang (2014) Quasi Maximum Likelihood Estimation for Simultaneous Spatial Autoregressive Models. MPRA Working Paper 59901.

Whittle, P. (1954) On stationary processes in the plane. *Biometrica* 41, 434–449.

Yang, K. & L.-F. Lee (2017) Identification and QML estimation of multivariate and simultaneous equations spatial autoregressive models. *Journal of Econometrics* 196, 196–214.

Yang, K. & L.-F. Lee (2019) Identification and estimation of spatial dynamic panel simultaneous equations models. *Regional Science and Urban Economics* 76, 32–46.

# Online Supplementary Appendices for "Simultaneous Equations Models with Higher-Order Spatial or Social Network Interactions"[a]

David M. Drukker,[b] Peter H. Egger,[c] and Ingmar R. Prucha[d]

December 27, 2021 (This Version)

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

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

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

**Abstract**

This supplement provides additional technical material, expanded proofs, and reports on a Monte Carlo study that explores the small sample properties of our estimators. For the main Assumptions, notations and definitions of various quantities appearing in the sequel, the reader is referred to the main paper.

# D    Supplement to Appendix A

**Proof of Lemma A.1:** Let $C_A = \sup_n \max_{i=1}^{m_n} \sum_{j=1}^{n} |a_{ij,n}|$, $C_\mu = \sup_n \max_{i=1}^{m_n} E |\mu_{i,n}|^p$ and $C_\eta = \sup_n \max_{i=1}^{m_n} E |\eta_{i,n}|^p$. Clearly

$$|\xi_{i,n}| \le |\mu_{i,n}| + \sum_{j=1}^{m_n} |a_{ij,n}| \, |\eta_{jn}| \le |\mu_{i,n}| + C_A \sum_{j=1}^{m_n} b_{ij,n} \, |\eta_{jn}|$$

with $b_{ij,n} = |a_{ij,n}| / \left( \sum_{j=1}^{m_n} |a_{ij,n}| \right)$ if $\sum_{j=1}^{m_n} |a_{ij,n}| > 0$ and $b_{ij,n} = 0$ if $\sum_{j=1}^{m_n} |a_{ij,n}| = 0$. Clearly $E |\xi_{i,n}|^p \le C_\mu$ for $\sum_{j=1}^{m_n} |a_{ij,n}| = 0$. For $\sum_{j=1}^{m_n} |a_{ij,n}| > 0$ observe that since $0 \le b_{ij,n} \le 1$ and $\sum_{j=1}^{m_n} b_{ij,n} = 1$ it follows from Holder's inequality that $\sum_{j=1}^{m_n} b_{ij,n} |\eta_{j,n}| \le \left[ \sum_{j=1}^{m_n} b_{ij,n} |\eta_{j,n}|^p \right]^{1/p}$ and consequently

$$\begin{aligned}
E |\xi_{i,n}|^p & \le 2^p E |\mu_{i,n}|^p + 2^p C_A^p E \left[ \sum_{j=1}^{m_n} b_{ij,n} |\eta_{j,n}| \right]^p \\
& \le 2^p E |\mu_{i,n}|^p + 2^p C_A^p \sum_{j=1}^{m_n} b_{ij,n} E[|\eta_{j,n}|^p] \le 2^p C_\mu + 2^p C_A^p C_\eta < \infty
\end{aligned}$$

which proves the claim since $C_A$, $C_\mu$ and $C_\eta$ do not depend on $i$ and $n$.    ∎

**Proof of Lemma A.2:** By assumption $\mathbf{X}_n$ is non-stochastic with $\sup_n \sup_{i,k} |x_{ik,n}| < \infty$, and so (A.1) holds trivially if $z_{ij,n}$ corresponds to an element of $\mathbf{X}_n$. Next observe that by (7) and (9) we have

$$\begin{aligned}
\mathbf{y}_n & = \mathbf{a}_n + \mathbf{A}_n \boldsymbol{\nu}_n, \\
\mathbf{a}_n & = (\mathbf{I}_{nG} - \mathbf{B}_n^*)^{-1} \mathbf{C}_n^* \mathbf{x}_n, \\
\mathbf{A}_n & = (\mathbf{I}_{nG} - \mathbf{B}_n^*)^{-1} (\mathbf{I}_{nG} - \mathbf{R}_n^*)^{-1} (\boldsymbol{\Sigma}_\star' \otimes \mathbf{I}_n).
\end{aligned}$$

In light of Assumptions 1-3 the absolute elements of $\mathbf{a}_n$ are uniformly bounded, and furthermore the row and column sums of the absolute elements of $\mathbf{A}_n$ are uniformly bounded; compare, e.g., Remark A.1 in Kelejian and Prucha (2004). By Assumption 4 the elements of $\boldsymbol{\nu}_n$ are i.i.d. with finite fourth moments. Thus it follows immediately from Lemma A.1 that $\sup_n \sup_{i,l} E |y_{il,n}|^4 < \infty$. Next observe that the columns of $\overline{\mathbf{Y}}_n$ are of the form $\overline{\mathbf{y}}_{l,s,n} = \mathbf{W}_{s,n} \mathbf{y}_{l,n}$. Since by As-

sumption 1 the row and column sums of the absolute elements of $\mathbf{W}_{s,n}$ are uniformly bounded it follows further from Lemma A.1 that $\sup_n \sup_{i,l,s} E\left|\overline{y}_{il,s,n}\right|^4 < \infty$, which completes the proof. ∎

**Proof of Lemma A.3:** In light of the proof of Lemma A.2, and observing that $\mathbf{u}_n = (\mathbf{I}_{nG} - \mathbf{R}_n^*)^{-1}(\mathbf{\Sigma}_*' \otimes \mathbf{I}_n)\boldsymbol{\nu}_n$, it is readily seen that under the maintained assumptions $\mathbf{u}_{g,n}$ and all columns of $\underline{\mathbf{Z}}_n$ are of the generic form

$$\mathbf{C}_{g,n}\boldsymbol{\nu}_n, \ \mathbf{c}_{g,n} \text{ or } \mathbf{c}_{g,n} + \mathbf{C}_{g,n}\boldsymbol{\nu}_n, \tag{D.1}$$

where $\mathbf{c}_{g,n}$ is an $n \times 1$ nonstochastic vector with uniformly bounded elements and $\mathbf{C}_{g,n}$ is an $n \times nG$ nonstochastic matrix whose row and column sums are uniformly bounded in absolute value. By Assumption 4 the elements of the $nG \times 1$ vector $\boldsymbol{\nu}_n$ are i.i.d. $(0,1)$ with finite fourth moments. Given this, it is readily seen that $n^{-1}\mathbf{u}_{h,n}'\mathbf{A}_n\mathbf{u}_{g,n}$ and the elements of $n^{-1}\underline{\mathbf{Z}}_n'\mathbf{A}_n\mathbf{u}_{g,n}$ and $n^{-1}\underline{\mathbf{Z}}_n\mathbf{A}_n\underline{\mathbf{Z}}_n$ are of the generic form $d_n$, $n^{-1}\mathbf{d}_n'\boldsymbol{\nu}_n$ or $n^{-1}\boldsymbol{\nu}_n'\mathbf{D}_n\boldsymbol{\nu}_n$, or sums thereof, where $|d_n|$, the absolute elements of $\mathbf{d}_n$ and the row and column sums of the absolute elements of $\mathbf{D}_n$ are uniformly bounded by some finite constant, say $K$. In the following let $\overline{\mathbf{D}}_n = (\overline{d}_{lk,n}) = (\mathbf{D}_n + \mathbf{D}_n')/2$. Observe that $E\left[n^{-1}\mathbf{d}_n'\boldsymbol{\nu}_n\right] = 0$ and $E\left[n^{-1}\boldsymbol{\nu}_n'\mathbf{D}_n\boldsymbol{\nu}_n\right] = n^{-1}tr(\mathbf{D}_n) = O(1)$. Furthermore observe that $var(n^{-1}\mathbf{d}_n'\boldsymbol{\nu}_n) \le n^{-1}K^2 = o(1)$, and that in light of, e.g., Lemma A.1 in Kelejian and Prucha (2004) and Remark A.2 in Kapoor et al. (2007), we have $var(n^{-1}\boldsymbol{\nu}_n'\mathbf{D}_n\boldsymbol{\nu}_n) \le n^{-2}tr(\overline{\mathbf{D}}_n^2) + n^{-2}K\sum_{i,g}\left|Ev_{ig,n}^4 - 3\right| \le n^{-1}K_* = o(1)$ for some finite constant $K_*$. Thus clearly $n^{-1}\mathbf{u}_{h,n}'\mathbf{A}_n\mathbf{u}_{g,n} = O_p(1)$, $n^{-1}\underline{\mathbf{Z}}_n\mathbf{A}_n\mathbf{u}_{g,n} = O_p(1)$ and $n^{-1}\underline{\mathbf{Z}}_n\mathbf{A}_n\underline{\mathbf{Z}}_n = O_p1)$. The third claim in the lemma follows from Chebychev's inequality. ∎

**Proof of Lemma A.4:** Clearly,

$$n^{1/2}(\widetilde{\boldsymbol{\delta}}_{g,n} - \boldsymbol{\delta}_{g,n}) = \widetilde{\mathbf{P}}_{gg,n}' n^{-1/2}\mathbf{F}_{gg,n}'\boldsymbol{\varepsilon}_{g,n},$$

where $\widetilde{\mathbf{P}}_{gg,n}$ and $\mathbf{F}_{gg,n}$ are defined in the lemma. Given Assumption 6 clearly $\widetilde{\mathbf{P}}_{gg,n} = \mathbf{P}_{gg} + o_p(1)$ with $\mathbf{P}_{gg}$ being finite, which establishes (c). Since by Assumption 2 the row and column sums of $(\mathbf{I}_n - \mathbf{R}_{g,n}^{*'})^{-1}$ are uniformly bounded in absolute value, and since by Assumption 5 the elements of $\mathbf{H}_n$ are uniformly bounded in absolute value, it follows that the elements of $\mathbf{F}_{gg,n}$ are uniformly bounded in absolute value. By Assumption 4, $E(\boldsymbol{\varepsilon}_{g,n}) = 0$ and $E(\boldsymbol{\varepsilon}_{g,n}\boldsymbol{\varepsilon}_{g,n}') = \sigma_{gg}\mathbf{I}_n$. Therefore, $En^{-1/2}\mathbf{F}_{gg,n}'\boldsymbol{\varepsilon}_{g,n} = \mathbf{0}$ and the elements of

2

$VC(n^{-1/2}\mathbf{F}'_n\boldsymbol{\varepsilon}_{g,n}) = \sigma_{gg}n^{-1}\mathbf{F}'_{gg,n}\mathbf{F}_{gg,n}$ are also uniformly bounded in absolute value. Thus, by Chebyshev's inequality $n^{-1/2}\mathbf{F}'_{gg,n}\boldsymbol{\varepsilon}_{g,n} = O_p(1)$, and consequently $n^{1/2}(\widetilde{\boldsymbol{\delta}}_{g,n} - \boldsymbol{\delta}_{g,n}) = \mathbf{P}'_{gg}n^{-1/2}\mathbf{F}'_{gg,n}\boldsymbol{\varepsilon}_{g,n} + o_p(1)$ and $\mathbf{P}'_{gg}n^{-1/2}\mathbf{F}'_{gg,n}\boldsymbol{\varepsilon}_{g,n} = O_p(1)$. This establishes (a) and (b), recalling that $\mathbf{T}_{gg,n} = \mathbf{F}_{gg,n}\mathbf{P}_{gg,n}$. Next observe that

$$
\begin{aligned}
&\lambda_{\min}(n^{-1}\mathbf{T}'_{gg,n}\mathbf{T}_{gg,n}) \\
\geq\quad &\lambda_{\min}\left[\left(\mathbf{I}_n - \mathbf{R}^{*\prime}_{g,n}\right)^{-1}\left(\mathbf{I}_n - \mathbf{R}^{*}_{g,n}\right)^{-1}\right]\lambda_{\min}\left[n^{-1}\mathbf{H}'_n\mathbf{H}_n\right] \\
&\lambda_{\min}\left\{[\mathbf{Q}'_{\mathbf{HZ},g}\mathbf{Q}^{-1}_{\mathbf{HH}}\mathbf{Q}_{\mathbf{HZ},g}]^{-1}\mathbf{Q}'_{\mathbf{HZ},g}\mathbf{Q}^{-1}_{\mathbf{HH}}\mathbf{Q}^{-1}_{\mathbf{HH}}\mathbf{Q}_{\mathbf{HZ},g}[\mathbf{Q}'_{\mathbf{HZ},g}\mathbf{Q}^{-1}_{\mathbf{HH}}\mathbf{Q}_{\mathbf{HZ},g}]^{-1}\right\} \\
\geq\quad &c
\end{aligned}
$$

for some $c > 0$, since in light of Assumptions 1 and 2 the largest eigenvalue of $\left(\mathbf{I}_n - \mathbf{R}^{*}_{g,n}\right)\left(\mathbf{I}_n - \mathbf{R}^{*\prime}_{g,n}\right)$ is bounded from above, and thus the smallest eigenvalue of $\left(\mathbf{I}_n - \mathbf{R}^{*\prime}_{g,n}\right)^{-1}\left(\mathbf{I}_n - \mathbf{R}^{*}_{g,n}\right)^{-1}$ is bounded away from zero, and since $\lambda_{\min}\left[n^{-1}\mathbf{H}'_n\mathbf{H}_n\right] \geq [\lambda_{\min}(\mathbf{Q}_{\mathbf{HH}})]/2 > 0$ for $n$ sufficiently large in light of Assumption 6. This establishes (d). ∎

**Proof of Lemma A.5:** Note from (1) and (10) that

$$
\mathbf{y}_{*g,n}(\widehat{\boldsymbol{\rho}}_{g,n}) = \mathbf{Z}_{*g,n}(\widehat{\boldsymbol{\rho}}_{g,n})\boldsymbol{\delta}_{g,n} + \boldsymbol{\varepsilon}_{g,n} - (\widehat{\mathbf{R}}^{*}_{g,n} - \mathbf{R}_{g,n})\mathbf{u}_{g,n}
$$

and hence

$$
\begin{aligned}
&n^{1/2}[\widehat{\boldsymbol{\delta}}_{g,n} - \boldsymbol{\delta}_{g,n}] \\
=\quad &\left[n^{-1}\widehat{\mathbf{Z}}'_{*g,n}(\widehat{\boldsymbol{\rho}}_{g,n})\mathbf{Z}_{*g,n}(\widehat{\boldsymbol{\rho}}_{g,n})\right]^{-1}n^{-1/2}\widehat{\mathbf{Z}}'_{*g,n}(\widehat{\boldsymbol{\rho}}_{g,n})\left[\boldsymbol{\varepsilon}_{g,n} - (\widehat{\mathbf{R}}^{*}_{g,n} - \mathbf{R}_{g,n})\mathbf{u}_{g,n}\right] \\
=\quad &\widetilde{\mathbf{P}}^{*\prime}_{gg,n}\left[n^{-1/2}\mathbf{F}^{*\prime}_{gg,n}\boldsymbol{\varepsilon}_{g,n} - \sum_{r\in\mathbf{I}_{g,\rho}}(\widehat{\rho}_{g,r,n} - \rho_{g,r,n})n^{-1/2}\overline{\mathbf{F}}^{*\prime}_{gg,r,n}\boldsymbol{\varepsilon}_{g,n}\right],
\end{aligned}
$$

with $\widehat{\mathbf{R}}^{*}_{g,n} = \mathbf{R}^{*}_{g,n}(\widehat{\boldsymbol{\rho}}_{g,n})$, and where $\widetilde{\mathbf{P}}^{*}_{gg,n}$ is defined in the lemma, $\mathbf{F}^{*}_{gg,n} = \mathbf{H}_n$, and $\overline{\mathbf{F}}^{*}_{gg,r,n} = \left(\mathbf{I}_n - \mathbf{R}^{*\prime}_{g,n}\right)^{-1}\mathbf{M}'_{r,n}\mathbf{H}_n$. In light of Assumption 6, and since $\widehat{\boldsymbol{\rho}}_n$ is consistent, it follows that

$$
n^{-1}\widehat{\mathbf{Z}}'_{*g,n}(\widehat{\boldsymbol{\rho}}_{g,n})\mathbf{Z}_{*g,n}(\widehat{\boldsymbol{\rho}}_{g,n}) - \mathbf{Q}'_{\mathbf{HZ},g*}(\boldsymbol{\rho}_{g,n})\mathbf{Q}^{-1}_{\mathbf{HH}}\mathbf{Q}_{\mathbf{HZ},g*}(\boldsymbol{\rho}_{g,n}) = o_p(1).
$$

Since by Assumption 6 we have $\mathbf{Q}'_{\mathbf{HZ},g*}(\boldsymbol{\rho}_{g,n})\mathbf{Q}^{-1}_{\mathbf{HH}}\mathbf{Q}_{\mathbf{HZ},g*}(\boldsymbol{\rho}_{g,n}) = O(1)$ and

$[\mathbf{Q}'_{\mathbf{HZ},g*}(\boldsymbol{\rho}_{g,n})\mathbf{Q}^{-1}_{\mathbf{HH}}\mathbf{Q}_{\mathbf{HZ},g*}(\boldsymbol{\rho}_{g,n})]^{-1} = O(1)$ it follows that

$$[n^{-1}\widehat{\mathbf{Z}}'_{*g,n}(\widehat{\rho}_{g,n})\mathbf{Z}_{*g,n}(\widehat{\boldsymbol{\rho}}_{g,n})]^{-1} - [\mathbf{Q}'_{\mathbf{HZ},g*}(\boldsymbol{\rho}_{g,n})\mathbf{Q}^{-1}_{\mathbf{HH}}\mathbf{Q}_{\mathbf{HZ},g*}(\boldsymbol{\rho}_{g,n})]^{-1} = o_p(1);$$

compare, e.g., Pötscher and Prucha (1997), Lemma F1. In light of this it follows further that $\widetilde{\mathbf{P}}^*_{gg,n} - \mathbf{P}^*_{gg,n} = o_p(1)$ and $\mathbf{P}^*_{gg,n} = O(1)$, with $\mathbf{P}^*_{gg,n}$ defined in the lemma. By argumentation analogous to that in the proof of Lemma A.4 it is readily seen that $n^{-1/2}\mathbf{F}^{*\prime}_{gg,n}\boldsymbol{\varepsilon}_{g,n} = O_p(1)$ and $n^{-1/2}\overline{\mathbf{F}}^{*\prime}_{gg,r,n}\boldsymbol{\varepsilon}_{g,n} = O_p(1)$. Consequently $n^{1/2}[\widehat{\boldsymbol{\delta}}_{g,n}-\boldsymbol{\delta}_{g,n}] = \mathbf{P}^{*\prime}_{gg,n}n^{-1/2}\mathbf{F}^{*\prime}_{gg,n}\boldsymbol{\varepsilon}_{g,n}+o_p(1)$ and $\mathbf{P}^{*\prime}_{gg,n}n^{-1/2}\mathbf{F}^{*\prime}_{gg,n}\varepsilon_{g,n} = O_p(1)$, observing again that $\widehat{\boldsymbol{\rho}}_{g,n} - \boldsymbol{\rho}_{g,n} = o_p(1)$. This established (a)-(c) recalling that $\mathbf{T}^*_{gg,n} = \mathbf{F}^*_{gg,n}\mathbf{P}^*_{gg,n}$. Next observe that

$$\begin{aligned}
&\lambda_{\min}(n^{-1}\mathbf{T}^{*\prime}_{gg,n}\mathbf{T}^*_{gg,n})\\
\geq\ &\lambda_{\min}\left[\mathbf{Q}'^{-1/2}_{\mathbf{HH}}n^{-1}\mathbf{H}'_n\mathbf{H}_n\mathbf{Q}^{-1/2}_{\mathbf{HH}}\right]\\
&\lambda_{\min}\left\{[\mathbf{Q}'_{\mathbf{HZ},g*}(\rho_{g,n})\mathbf{Q}^{-1}_{\mathbf{HH}}\mathbf{Q}_{\mathbf{HZ},g*}(\boldsymbol{\rho}_{g,n})]^{-1}\right\}\\
\geq\ &\lambda_{\min}(\mathbf{Q}^{-1}_{\mathbf{HH}})\lambda_{\min}\left\{[\mathbf{Q}'_{\mathbf{HZ},g*}(\rho_{g,n})\mathbf{Q}^{-1}_{\mathbf{HH}}\mathbf{Q}_{\mathbf{HZ},g*}(\boldsymbol{\rho}_{g,n})]^{-1}\right\}\lambda_{\min}\left[n^{-1}\mathbf{H}'_n\mathbf{H}_n\right]\\
\geq\ &\lambda_{\min}(\mathbf{Q}^{-1}_{\mathbf{HH}})\lambda_{\min}\left\{[\mathbf{Q}'_{\mathbf{HZ},g*}(\rho_{g,n})\mathbf{Q}^{-1}_{\mathbf{HH}}\mathbf{Q}_{\mathbf{HZ},g*}(\boldsymbol{\rho}_{g,n})]^{-1}\right\}\left[\lambda_{\min}(\mathbf{Q}^{-1}_{\mathbf{HH}})\right]/2 \geq c_*
\end{aligned}$$

for some $c_* > 0$ in light of Assumption 6, and observing that $\lambda_{\min}\left[n^{-1}\mathbf{H}'_n\mathbf{H}_n\right] \geq [\lambda_{\min}(\mathbf{Q}_{\mathbf{HH}})]/2 > 0$ for $n$ sufficiently large. This establishes (d). $\blacksquare$

**Proof of Lemma A.6:** Note from (1) and (11) that

$$\mathbf{y}_{*n}(\widehat{\boldsymbol{\rho}}_n) = \mathbf{Z}_{*n}(\widehat{\boldsymbol{\rho}}_n)\boldsymbol{\delta}_n + \boldsymbol{\varepsilon}_n - (\widehat{\mathbf{R}}^*_n - \mathbf{R}_n)\mathbf{u}_n$$

and hence

$$\begin{aligned}
&n^{1/2}[\widehat{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_n]\\
=\ &\left[n^{-1}\widehat{\mathbf{Z}}'_{*n}(\widehat{\boldsymbol{\rho}}_n)(\widehat{\boldsymbol{\Sigma}}^{-1}_n \otimes \mathbf{I}_n)\mathbf{Z}_{*n}(\widehat{\boldsymbol{\rho}}_n)\right]^{-1} n^{-1/2}\widehat{\mathbf{Z}}'_{*n}(\widehat{\boldsymbol{\rho}}_n)(\widehat{\boldsymbol{\Sigma}}^{-1}_n \otimes \mathbf{I}_n)\\
&\times\left[\boldsymbol{\varepsilon}_n - (\widehat{\mathbf{R}}^*_n - \mathbf{R}_n)\mathbf{u}_n\right]\\
=\ &\left[n^{-1}\widehat{\mathbf{Z}}'_{*n}(\widehat{\boldsymbol{\rho}}_n)(\widehat{\boldsymbol{\Sigma}}^{-1}_n \otimes \mathbf{I}_n)\mathbf{Z}_{*n}(\widehat{\boldsymbol{\rho}}_n)\right]^{-1} diag\left[n^{-1}\mathbf{Z}'_{*g,n}(\widehat{\boldsymbol{\rho}}_{g,n})\mathbf{H}_n\right]\\
&\times\left[\widehat{\boldsymbol{\Sigma}}^{-1}_n \otimes (n^{-1}\mathbf{H}'_n\mathbf{H}_n)^{-1}\right](\mathbf{I}_G \otimes n^{-1/2}\mathbf{H}'_n)\left[\boldsymbol{\varepsilon}_n - (\widehat{\mathbf{R}}^*_n - \mathbf{R}_n)\mathbf{u}_n\right]\\
=\ &\widetilde{\mathbf{P}}^{**\prime}_n n^{-1/2}\mathbf{F}^{**\prime}_n\boldsymbol{\varepsilon}_n + \widetilde{\mathbf{P}}^{**\prime}_n\begin{bmatrix}\sum_{r\in\mathbf{I}_{1,\rho}}(\widehat{\rho}_{1,r,n} - \rho_{1,r,n})n^{-1/2}\overline{\mathbf{F}}^{**\prime}_{11,r,n}\varepsilon_{1,n}\\ \vdots\\ \sum_{r\in\mathbf{I}_{G,\rho}}(\widehat{\rho}_{G,r,n} - \rho_{G,r,n})n^{-1/2}\overline{\mathbf{F}}^{**\prime}_{GG,r,n}\varepsilon_{g,n}\end{bmatrix},
\end{aligned}$$

with $\widehat{\mathbf{R}}_n^* = diag_{g=1}^G \left[\mathbf{R}_{g,n}^*(\widehat{\boldsymbol{\rho}}_{g,n})\right]$, and where $\widetilde{\mathbf{P}}_n^{**}$ and $\mathbf{F}_n^{**}$ are defined in the lemma, and $\overline{\mathbf{F}}_{gg,r,n}^{**} = \left(\mathbf{I}_n - \mathbf{R}_{g,n}^{*\prime}\right)^{-1}\mathbf{M}_{r,n}'\mathbf{H}_n$. Observe that the $(g,h)$-th block of

$$n^{-1}\widehat{\mathbf{Z}}_{*n}'(\widehat{\boldsymbol{\rho}}_n)(\widehat{\boldsymbol{\Sigma}}_n^{-1}\otimes\mathbf{I}_n)\mathbf{Z}_{*n}(\widehat{\boldsymbol{\rho}}_n) - diag\left[\mathbf{Q}_{\mathbf{HZ},g*}'(\boldsymbol{\rho}_{g,n})\right]\left[\boldsymbol{\Sigma}^{-1}\otimes\mathbf{Q}_{\mathbf{HH}}^{-1}\right]diag\left[\mathbf{Q}_{\mathbf{HZ},g*}(\boldsymbol{\rho}_{g,n})\right]$$

is

$$\widehat{\sigma}_{gh,n}n^{-1}\widehat{\mathbf{Z}}_{*g,n}'(\widehat{\boldsymbol{\rho}}_{g,n})\mathbf{Z}_{*h,n}(\widehat{\boldsymbol{\rho}}_{h,n}) - \sigma_{gh}\mathbf{Q}_{\mathbf{HZ},g*}'(\boldsymbol{\rho}_{g,n})\mathbf{Q}_{\mathbf{HH}}^{-1}\mathbf{Q}_{\mathbf{HZ},h*}(\boldsymbol{\rho}_{h,n}) = o_p(1)$$

in light of Assumption 6, and since $\widehat{\boldsymbol{\rho}}_n$ and $\widehat{\boldsymbol{\Sigma}}_n$ are consistent. By Assumption 6 we have $diag\left[\mathbf{Q}_{\mathbf{HZ},g*}'(\boldsymbol{\rho}_{g,n})\right]\left[\boldsymbol{\Sigma}^{-1}\otimes\mathbf{Q}_{\mathbf{HH}}^{-1}\right]diag\left[\mathbf{Q}_{\mathbf{HZ},g*}(\boldsymbol{\rho}_{g,n})\right] = O(1)$ and

$$\lambda_{\min}\left\{diag\left[\mathbf{Q}_{\mathbf{HZ},g*}'(\boldsymbol{\rho}_{g,n})\right]\left[\boldsymbol{\Sigma}^{-1}\otimes\mathbf{Q}_{\mathbf{HH}}^{-1}\right]diag\left[\mathbf{Q}_{\mathbf{HZ},g*}(\boldsymbol{\rho}_{g,n})\right]\right\}$$
$$\geq \quad \lambda_{\min}\left\{\boldsymbol{\Sigma}^{-1}\otimes\mathbf{I}\right\}\lambda_{\min}\left\{diag\left[\mathbf{Q}_{\mathbf{HZ},g*}'(\boldsymbol{\rho}_{g,n})\mathbf{Q}_{\mathbf{HH}}^{-1}\mathbf{Q}_{\mathbf{HZ},g*}(\boldsymbol{\rho}_{g,n})\right]\right\} \geq c_*$$

for some $c_* > 0$. This in turn implies that

$$\left\{diag\left[\mathbf{Q}_{\mathbf{HZ},g*}'(\boldsymbol{\rho}_{g,n})\right]\left[\boldsymbol{\Sigma}^{-1}\otimes\mathbf{Q}_{\mathbf{HH}}^{-1}\right]diag\left[\mathbf{Q}_{\mathbf{HZ},g*}(\boldsymbol{\rho}_{g,n})\right]\right\}^{-1} = O(1).$$

Consequently

$$\left\{n^{-1}\widehat{\mathbf{Z}}_{*n}'(\widehat{\boldsymbol{\rho}}_n)(\widehat{\boldsymbol{\Sigma}}_n^{-1}\otimes\mathbf{I}_n)\mathbf{Z}_{*n}(\widehat{\boldsymbol{\rho}}_n)\right\}^{-1} -$$
$$\left\{diag\left[\mathbf{Q}_{\mathbf{HZ},g*}'(\boldsymbol{\rho}_{g,n})\right]\left[\boldsymbol{\Sigma}^{-1}\otimes\mathbf{Q}_{\mathbf{HH}}^{-1}\right]diag\left[\mathbf{Q}_{\mathbf{HZ},g*}(\boldsymbol{\rho}_{g,n})\right]\right\}^{-1} = o_p(1);$$

compare, e.g., Pötscher and Prucha (1997), Lemma F1. In light of this it is now readily seen that $\widetilde{\mathbf{P}}_n^{**} - \mathbf{P}_n^{**} = o_p(1)$ and $\mathbf{P}_n^{**} = O(1)$, with $\mathbf{P}_n^{**}$ defined in the lemma. By argumentation analogous to that in the proof of Lemma A.4 it is readily seen that $n^{-1/2}\mathbf{F}_n^{**\prime}\boldsymbol{\varepsilon}_n = O_p(1)$ and $n^{-1/2}\overline{\mathbf{F}}_{gg,r,n}^{*\prime}\boldsymbol{\varepsilon}_{g,n} = O_p(1)$. Consequently $n^{1/2}[\widehat{\widehat{\boldsymbol{\delta}}}_n - \boldsymbol{\delta}_n] = \mathbf{P}_n^{**\prime}n^{-1/2}\mathbf{F}_n^{**\prime}\varepsilon_n + o_p(1)$ with $\mathbf{P}_n^{**\prime}n^{-1/2}\mathbf{F}_n^{**\prime}\varepsilon_n = O_p(1)$, observing again that $\widehat{\boldsymbol{\rho}}_{g,n} - \boldsymbol{\rho}_{g,n} = o_p(1)$. This establishes (a)-(c), recalling that

$\mathbf{T}_n^{**} = \mathbf{F}_n^{**}\mathbf{P}_n^{**}$. Next observe that

$$\lambda_{\min}(n^{-1}\mathbf{T}_n^{**\prime}\mathbf{T}_n^{**})$$
$$\geq \lambda_{\min}(\Sigma^{-1})\lambda_{\min}\left[\mathbf{Q}_{\mathbf{HH}}^{\prime -1/2}n^{-1}\mathbf{H}_n^{\prime}\mathbf{H}_n\mathbf{Q}_{\mathbf{HH}}^{-1/2}\right]$$
$$\times\lambda_{\min}\left[\left\{diag\left[\mathbf{Q}_{\mathbf{HZ},g*}^{\prime}(\boldsymbol{\rho}_{g,n})\right]\left[\Sigma^{-1}\otimes\mathbf{Q}_{\mathbf{HH}}^{-1}\right]diag\left[\mathbf{Q}_{\mathbf{HZ},g*}(\boldsymbol{\rho}_{g,n})\right]\right\}^{-1}\right]$$
$$\geq \lambda_{\min}(\Sigma^{-1})\lambda_{\min}(\mathbf{Q}_{\mathbf{HH}}^{-1})\lambda_{\min}\left[n^{-1}\mathbf{H}_n^{\prime}\mathbf{H}_n\right]$$
$$\times\lambda_{\min}\left[\left\{diag\left[\mathbf{Q}_{\mathbf{HZ},g*}^{\prime}(\boldsymbol{\rho}_{g,n})\right]\left[\Sigma^{-1}\otimes\mathbf{Q}_{\mathbf{HH}}^{-1}\right]diag\left[\mathbf{Q}_{\mathbf{HZ},g*}(\boldsymbol{\rho}_{g,n})\right]\right\}^{-1}\right]$$
$$\geq \lambda_{\min}(\Sigma^{-1})\lambda_{\min}(\mathbf{Q}_{\mathbf{HH}}^{-1})[\lambda_{\min}(\mathbf{Q}_{\mathbf{HH}}^{-1})/2]$$
$$\times\lambda_{\min}\left[\left\{diag\left[\mathbf{Q}_{\mathbf{HZ},g*}^{\prime}(\boldsymbol{\rho}_{g,n})\right]\left[\Sigma^{-1}\otimes\mathbf{Q}_{\mathbf{HH}}^{-1}\right]diag\left[\mathbf{Q}_{\mathbf{HZ},g*}(\boldsymbol{\rho}_{g,n})\right]\right\}^{-1}\right] \geq c_*$$

for some $c_* > 0$ in light of Assumption 6, and observing that $\lambda_{\min}\left[n^{-1}\mathbf{H}_n^{\prime}\mathbf{H}_n\right] \geq [\lambda_{\min}(\mathbf{Q}_{\mathbf{HH}})]/2 > 0$ for $n$ sufficiently large. This establishes (d). ∎

**Proof of Lemma A.7:** Without loss of generality, assume that $\sigma^2 = 1$, since the model in Assumption A.1 can always be normalized accordingly.

We first prove part (a) of the lemma. Let $\vartheta_n = n^{-1}\mathfrak{u}_n^{\prime}\mathbf{A}_n^*\mathfrak{u}_n$ and $\widetilde{\vartheta}_n = n^{-1}\widetilde{\mathfrak{u}}_n^{\prime}\mathbf{A}_n^*\widetilde{\mathfrak{u}}_n$, then, in light of Assumption A.1, we have $\vartheta_n = n^{-1}\mathfrak{e}_n^{\prime}\mathbf{B}_n^*\mathfrak{e}_n$ with $\mathbf{B}_n^* = (1/2)\mathfrak{R}_n^{-1\prime}(\mathbf{A}_n^* + \mathbf{A}_n^{*\prime})\mathfrak{R}_n^{-1}$. Furthermore, by Assumption A.1, the row and column sums of the matrices $\mathfrak{R}_n$ are uniformly bounded in absolute value. Since this property is preserved under matrix addition and multiplication - see, e.g., Remark A.1 in Kelejian and Prucha (2004) - it follows that also the row and column sums of the matrices $\mathbf{B}_n^*$ and $\mathbf{B}_n^*\mathbf{B}_n^*$ are uniformly bounded in absolute value. In the following let $K < \infty$ be a common bound for the row and column sums of the absolute elements of $\mathbf{B}_n^*$ and $\mathbf{B}_n^*\mathbf{B}_n^*$, and of their respective elements. Then, using the triangle inequality and the Cauchy-Schwarz inequality, we have

$$E|\vartheta_n| = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}|b_{ij,n}^*|\,E|\mathfrak{e}_{i,n}|\,|\mathfrak{e}_{j,n}| \leq n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}|b_{ij,n}^*| \leq K.$$

Furthermore, utilizing the expression for the variance of linear quadratic forms given in Lemma A.1 in Kelejian and Prucha (2007) we have in light of Assumption A.1

$$var(\vartheta_n) = n^{-2}2tr(\mathbf{B}_n^*\mathbf{B}_n^*) + n^{-2}\sum_{i=1}^{n}b_{ii,n}^{*2}\left[E\mathfrak{e}_{i,n}^4 - 3\right]$$
$$\leq n^{-1}2K + n^{-1}K^2\sup_{i}\left[E\mathfrak{e}_{i,n}^4 - 3\right].$$

6

Given that the fourth moments of the $\mathfrak{e}_{i,n}$ are uniformly bounded in light of Assumption A.1, this establishes the first two claims of part (a) of the lemma.

We next prove the last claim of part (a) of the lemma. The above discussion implies that $\vartheta_n - E\vartheta_n = o_p(1)$. Hence it suffices to show that $\widetilde{\vartheta}_n - \vartheta_n = o_p(1)$. By Assumptions A.1 and A.2

$$\widetilde{\vartheta}_n - \vartheta_n = \phi_n + \psi_n$$

with

$$
\begin{aligned}
\phi_n &= n^{-1}\mathbf{\Delta}_n'\mathfrak{D}_n'\left(\mathbf{A}_n^* + \mathbf{A}_n^{*\prime}\right)\mathfrak{u}_n = n^{-1}\mathbf{\Delta}_n'\mathfrak{D}_n'\mathbf{C}_n^*\mathfrak{e}_n, \\
\psi_n &= n^{-1}\mathbf{\Delta}_n'\mathfrak{D}_n'\mathbf{A}_n^*\mathfrak{D}_n\mathbf{\Delta}_n,
\end{aligned}
$$

and $\mathbf{C}_n^* = (c_{ij,n}^*) = \left(\mathbf{A}_n^* + \mathbf{A}_n^{*\prime}\right)\mathfrak{R}_n^{-1}$. The row and column sums of the matrices $\mathbf{C}_n^*$ are again seen to be uniformly bounded in absolute value. Let $\overline{K} < \infty$ denote a uniform bound for the row and column sums of the absolute elements of the matrices $\mathbf{A}_n^*$ and $\mathbf{C}_n^*$, and let $\mathbf{c}_{i.,n}^*$ and $\mathfrak{d}_{i.,n}$ denote the $i$-th row of $\mathbf{C}_n^*$ and $\mathfrak{D}_n$, respectively.

To prove the claim we now show that both $\phi_n$ and $\psi_n$ are $o_p(1)$. Using the triangle and Hölder inequality we get

$$
\begin{aligned}
|\phi_n| &= \left| n^{-1}\sum_{i=1}^n \mathbf{\Delta}_n'\mathfrak{d}_{i.,n}'\mathbf{c}_{i.,n}^*\mathfrak{e}_n \right| \tag{D.2} \\
&\leq n^{-1}\|\mathbf{\Delta}_n\|\sum_{i=1}^n\|\mathfrak{d}_{i.,n}\|\sum_{j=1}^n\left|c_{ij,n}^*\right||\mathfrak{e}_{j,n}| \leq n^{-1}\|\mathbf{\Delta}_n\|\sum_{j=1}^n|\mathfrak{e}_{j,n}|\sum_{i=1}^n\|\mathfrak{d}_{i.,n}\|\left|c_{ij,n}^*\right| \\
&\leq n^{-1}\|\mathbf{\Delta}_n\|\sum_{j=1}^n|\mathfrak{e}_{j,n}|\left(\sum_{i=1}^n\|\mathfrak{d}_{i.,n}\|^p\right)^{1/p}\left(\sum_{i=1}^n\left|c_{ij,n}^*\right|^q\right)^{1/q} \\
&\leq \overline{K}n^{1/p-1/2}\left(n^{1/2}\|\mathbf{\Delta}_n\|\right)\left(n^{-1}\sum_{j=1}^n|\mathfrak{e}_{j,n}|\right)\left(n^{-1}\sum_{i=1}^n\|\mathfrak{d}_{i.,n}\|^p\right)^{1/p}
\end{aligned}
$$

for $p = 2+\delta$ and $1/p+1/q = 1$, and where $\delta > 0$ is as in Assumption A.2. The last inequality utilizes the observation of Remark C.1 in Kelejian and Prucha (2007). Since the $\mathfrak{e}_{j,n}$ are independent with bounded second moments, it follows that $n^{-1}\sum_{j=1}^n|\mathfrak{e}_{j,n}| = O_p(1)$. The terms $n^{1/2}\|\mathbf{\Delta}_n\|$ and $n^{-1}\sum_{i=1}^n\|\mathfrak{d}_{i.,n}\|^p$ are $O_p(1)$ by Assumption A.2. Since $n^{1/p-1/2} \to 0$ as $n \to \infty$ it follows that $\phi_n = o_p(1)$.

Again, using the triangle and Hölder inequality yields

$$
\begin{aligned}
|\psi_n| &= \left| n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbf{\Delta}_n' \mathfrak{d}_{i.,n}' a_{ij,n}^* \mathfrak{d}_{j.,n} \mathbf{\Delta}_n \right| \qquad\qquad\qquad\qquad\text{(D.3)} \\
&\leq n^{-1} \|\mathbf{\Delta}_n\|^2 \sum_{i=1}^{n} \|\mathfrak{d}_{i.,n}\| \sum_{j=1}^{n} \|\mathfrak{d}_{j.,n}\| \, |a_{ij,n}^*| \\
&\leq n^{-1} \|\mathbf{\Delta}_n\|^2 \sum_{i=1}^{n} \|\mathfrak{d}_{i.,n}\| \left( \sum_{j=1}^{n} \|\mathfrak{d}_{j.,n}\|^p \right)^{1/p} \left( \sum_{j=1}^{n} |a_{ij,n}^*|^q \right)^{1/q} \\
&\leq \overline{K} n^{1/p} \|\mathbf{\Delta}_n\|^2 \left( n^{-1} \sum_{i=1}^{n} \|\mathfrak{d}_{i.,n}\| \right) \left( n^{-1} \sum_{j=1}^{n} \|\mathfrak{d}_{j.,n}\|^p \right)^{1/p} \\
&\leq \overline{K} n^{1/p-1/2} n^{-1/2} (n^{1/2} \|\mathbf{\Delta}_n\|)^2 \left( n^{-1} \sum_{i=1}^{n} \|\mathfrak{d}_{i.,n}\|^p \right)^{2/p}
\end{aligned}
$$

with $p$ and $q$ as before. By Assumption A.2 both $n^{-1} \sum_{i=1}^{n} \|\mathfrak{d}_{i.,n}\|^p$ and $n^{1/2} \|\mathbf{\Delta}_n\|$ are $O_p(1)$. Since $n^{1/p-1/2} \to 0$ as $n \to \infty$ it follows that $\psi_n = o_p(1)$. From the last inequality we see also that $n^{1/2} \psi_n = o_p(1)$.

We next prove part (b) of the lemma. In the following let $\underline{\vartheta}_{s,n}$ denote the $s$-th element of $n^{-1} \mathfrak{D}_n' \mathbf{A}_n^* \mathfrak{u}_n$. Observe $E \mathfrak{u}_n \mathfrak{u}_n' = \mathfrak{R}_n^{-1} \mathfrak{R}_n^{-1\prime}$. Then given Assumptions A.1 and A.2 there exists a constant $\overline{\overline{K}} < \infty$ such that $E \mathfrak{u}_{i,n}^2 \leq \overline{\overline{K}}$ and $E |\mathfrak{d}_{js,n}|^p \leq \overline{\overline{K}}$. WLOG assume that the row and column sums of the matrices $\mathbf{A}_n^*$ are uniformly bounded by $\overline{\overline{K}}$. Utilizing the Cauchy-Schwarz and Lyapunov inequalities we then have $E |\mathfrak{u}_{i,n}| |\mathfrak{d}_{js,n}| \leq \left[ E \mathfrak{u}_{i,n}^2 \right]^{1/2} \left[ E \mathfrak{d}_{js,n}^2 \right]^{1/2} \leq \left[ E \mathfrak{u}_{i,n}^2 \right]^{1/2} (E |\mathfrak{d}_{js,n}|^p)^{1/p} \leq \overline{\overline{K}}^{1/2+1/p}$ with $p$ as before and, hence,

$$
E |\underline{\vartheta}_{s,n}| = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} |a_{ij,n}^*| \, E \left[ |\mathfrak{u}_{i,n}| |\mathfrak{d}_{js,n}| \right] \leq \overline{\overline{K}}^{1/2+1/p} n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} |a_{ij,n}^*| \leq \overline{\overline{K}}^{3/2+1/p} < \infty,
$$

which shows that indeed $E \left| n^{-1} \mathfrak{d}_{.s,n}' \mathbf{A}_n^* \mathfrak{u}_n \right| = O(1)$ where $\mathfrak{d}_{.s,n}$ denotes the $s$-th column of $\mathfrak{D}_n$. Of course, the argument also shows that $\boldsymbol{\alpha}_n^* = n^{-1} E \mathfrak{D}_n' (\mathbf{A}_n^* + \mathbf{A}_n^{*\prime}) \mathfrak{u}_n = O(1)$. Next observe that

$$
n^{-1} \mathfrak{D}_n' \mathbf{A}_n^* \widetilde{\mathfrak{u}}_n = n^{-1} \mathfrak{D}_n' \mathbf{A}_n^* \mathfrak{u}_n + \underline{\psi}_n,
$$

where $\underline{\psi}_n = n^{-1} \mathfrak{D}_n' \mathbf{A}_n^* \mathfrak{D}_n \mathbf{\Delta}_n$. By argumentation analogous to that employed

8

to demonstrate that $n^{1/2}\psi_n = o_p(1)$ it follows that also $\underline{\psi}_n = o_p(1)$, which completes the proof of part (b).

We next prove part (c). In light of the proof of part (a) we have

$$n^{-1/2}\widetilde{\mathfrak{u}}'_n \mathbf{A}^*_n \widetilde{\mathfrak{u}}_n = n^{-1/2}\mathfrak{u}'_n \mathbf{A}^*_n \mathfrak{u}_n + [n^{-1}\mathfrak{u}'_n(\mathbf{A}^*_n + \mathbf{A}^{*\prime}_n)\mathfrak{D}_n]n^{1/2}\mathbf{\Delta}_n + n^{1/2}\psi_n$$

with $n^{1/2}\psi_n = o_p(1)$. In light of part (b) and Assumption A.3 we have $n^{-1}\mathfrak{u}'_n(\mathbf{A}^*_n + \mathbf{A}^{*\prime}_n)\mathfrak{D}_n - \boldsymbol{\alpha}^{*\prime}_n = o_p(1)$. The claim follows since $n^{1/2}\mathbf{\Delta}_n = O_p(1)$ by Assumption A.2. ∎

**Remark A.1:** For future reference it proves helpful to note that in light of Remark A.1 in Kelejian and Prucha (2004) the constant $\overline{K}$ used in proving the last claim of part (a) of the above lemma can be chosen as $\overline{K} = 2c_P c_A$ where $c_P$ and $c_A$ denote a bound for the row and column sums of the absolute elements of $\mathfrak{R}^{*-1}_n$ and $\mathbf{A}^*_n$. Furthermore it proves helpful to observe that in light of (D.2) and (D.3)

$$\left|\widetilde{\vartheta}_n - \vartheta_n\right| \leq 2c_P c_A \varsigma_n,$$

where $\varsigma_n = o_p(1)$ does not depend on $\mathbf{A}^*_n$.

**Proof of Lemma A.8:** Given Assumption A.1 and the maintained assumptions on $\mathbf{A}_n$ it follows that the row and column sums of $\mathbf{A}^*_n = \mathfrak{R}'_n \mathbf{A}_n \mathfrak{R}_n$ are bounded uniformly in absolute value. Thus by Lemma A.7(c), and utilizing Assumption A.4, we have

$$
\begin{aligned}
& n^{-1/2}\widetilde{\mathfrak{u}}'_n \mathfrak{R}'_n \mathbf{A}_n \mathfrak{R}_n \widetilde{\mathfrak{u}}_n \\
= & \; n^{-1/2}\mathfrak{u}'_n \mathfrak{R}'_n \mathbf{A}_n \mathfrak{R}_n \mathfrak{u}_n + \boldsymbol{\alpha}'_n n^{1/2}\Delta_n + o_p(1) \\
= & \; n^{-1/2}\boldsymbol{\varepsilon}'_{g,n} \mathbf{A}_n \varepsilon_{g,n} + n^{-1/2}\boldsymbol{\alpha}'_n \left[\sum_{h=1}^{G} \mathbf{T}'_{h,n} \boldsymbol{\varepsilon}_{h,n} + o_p(1)\right] + o_p(1) \\
= & \; n^{-1/2}\boldsymbol{\varepsilon}'_{g,n} \mathbf{A}_n \varepsilon_{g,n} + n^{-1/2}\sum_{h=1}^{G} \mathbf{a}'_{h,n} \varepsilon_{h,n} + o_p(1).
\end{aligned}
$$

The last inequality holds since $\boldsymbol{\alpha}_n = O(1)$; see the remark in Lemma A.7(c). Given this and the maintained assumption on $\mathbf{P}_{h,n}$ it follows that $\mathbf{c}_{h,n} = (c_{h1,n}, \ldots, c_{hp_\Delta,n})' = \mathbf{P}_{h,n}\boldsymbol{\alpha}_n = O(1)$. Since $\mathbf{a}_{h,n} = \mathbf{F}_{h,n}\mathbf{c}_{h,n}$ we have

$$|a_{hi,n}|^\eta = \left|\sum_{s=1}^{p_F} f_{his,n} c_{hs,n}\right|^\eta \leq p_F^\eta K^\eta \sum_{s=1}^{p_\Delta} |f_{his,n}|^\eta,$$

9

using inequality (1.4.4.) in Bierens (1994). Thus

$$\sup_n n^{-1} \sum_{i=1}^n |a_{hi,n}|^\eta \leq p_F^\eta K^\eta \sum_{s=1}^{pF} \sup_n n^{-1} \sum_{i=1}^n |f_{his,n}|^\eta < \infty$$

in light of Assumption A.4(a). This proves part (a). Part (b) follows readily from, e.g., Lemma A.1 in Kelejian and Prucha (2010). ∎

# E   Supplement to Section 6

In the following we provide more details on the derivation of the results in the Section on Limited and Full Information One-Step Estimators. We first discuss limited information estimators. Let $\boldsymbol{\theta}_{g,n} = [\boldsymbol{\delta}'_{g,n}, \boldsymbol{\rho}'_{g,n}]'$ and $\widehat{\boldsymbol{\theta}}^o_{g,n} = [\widehat{\boldsymbol{\delta}}^{o\prime}_{g,n}, \widehat{\boldsymbol{\rho}}^{o\prime}_{g,n}]'$, then by argumentation analogous to those for two-step estimators we have $(\boldsymbol{\Phi}^o_{gg,n})^{-1/2}n^{1/2}(\widehat{\boldsymbol{\theta}}^o_{g,n} - \boldsymbol{\theta}_{g,n}) \xrightarrow{d} N[0, I]$ with

$$\boldsymbol{\Phi}^o_{gg,n} - \left[ \left[ \frac{\partial \mathbf{m}_{g,n}(\boldsymbol{\theta}_{g,n})}{\partial \boldsymbol{\theta}'_{g,n}} \right] \begin{bmatrix} \boldsymbol{\Psi}^{LL}_{gg,n} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi}^{QQ}_{gg,n} \end{bmatrix}^{-1} \left[ \frac{\partial \mathbf{m}_{g,n}(\boldsymbol{\theta}_{g,n})}{\partial \boldsymbol{\theta}_{g,n}} \right] \right]^{-1} = o_p(1),$$

where

$$\begin{aligned} \boldsymbol{\Psi}^{LL}_{gg,n} &= \sigma_{gg,n} \left[ n^{-1} \mathbf{H}'_n \mathbf{H}_n \right], \\ \boldsymbol{\Psi}^{QQ}_{gg,n} &= \sigma^2_{gg,n} \mathbf{K}^{QQ}_n, \end{aligned}$$

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

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

Recall that

$$\begin{aligned} \mathbf{y}_{g,n} &= \mathbf{Z}_{g,n} \boldsymbol{\delta}_{g,n} + \mathbf{u}_{g,n}, \\ \mathbf{u}_{g,n} &= \overline{\mathbf{U}}_{g,n} \boldsymbol{\rho}_{g,n} + \varepsilon_{g,n} = \mathbf{R}^*_{g,n}(\boldsymbol{\rho}_{g,n}) \mathbf{u}_{g,n} + \boldsymbol{\varepsilon}_{g,n}, \end{aligned}$$

with $\mathbf{R}^*_{g,n}(\boldsymbol{\rho}_{g,n}) = \sum_{r \in \mathbf{I}_{g,\rho}} \rho_{g,r,n} \mathbf{M}_{r,n}$. Consequently

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

and

$$\begin{aligned} \mathbf{H}'_n \boldsymbol{\varepsilon}_{g,n} &= \mathbf{H}'_n [\mathbf{I}_n - \mathbf{R}^*_{g,n}(\boldsymbol{\rho}_{g,n})][\mathbf{y}_{g,n} - \mathbf{Z}_{g,n} \boldsymbol{\delta}_{g,n}], \\ \boldsymbol{\varepsilon}'_{g,n} \mathbf{A}_{s,n} \boldsymbol{\varepsilon}_{g,n} &= \mathbf{u}'_{g,n} [\mathbf{I}_n - \mathbf{R}^{*\prime}_{g,n}(\boldsymbol{\rho}_{g,n})] \overline{\mathbf{A}}_{s,n} [\mathbf{I}_n - \mathbf{R}^*_{g,n}(\boldsymbol{\rho}_{g,n})] \mathbf{u}_{g,n} \\ &= \mathbf{u}'_{g,n} \overline{\mathbf{A}}_{s,n} \mathbf{u}_{g,n} - 2 \mathbf{u}'_{g,n} \mathbf{R}^{*\prime}_{g,n}(\boldsymbol{\rho}_{g,n}) \overline{\mathbf{A}}_{s,n} \mathbf{u}_{g,n} \\ &\quad + \mathbf{u}'_{g,n} \mathbf{R}^{*\prime}_{g,n}(\boldsymbol{\rho}_{g,n}) \overline{\mathbf{A}}_{s,n} \mathbf{R}^*_{g,n}(\boldsymbol{\rho}_{g,n}) \mathbf{u}_{g,n}. \end{aligned}$$

Thus

$$\frac{\partial}{\partial\boldsymbol{\delta}_{g,n}}\left[\mathbf{H}_n'\boldsymbol{\varepsilon}_{g,n}\right] = \mathbf{H}_n'[\mathbf{I}_n - \mathbf{R}_{g,n}^*(\boldsymbol{\rho}_{g,n})]\frac{\partial\mathbf{u}_{g,n}}{\partial\boldsymbol{\delta}_{g,n}} = -\mathbf{H}_n'[\mathbf{I}_n - \mathbf{R}_{g,n}^*(\boldsymbol{\rho}_{g,n})]\mathbf{Z}_{g,n},$$

$$\frac{\partial}{\partial\boldsymbol{\delta}_{g,n}}[\boldsymbol{\varepsilon}_{g,n}'\mathbf{A}_{s,n}\boldsymbol{\varepsilon}_{g,n}] = 2\mathbf{u}_{g,n}'[\mathbf{I}_n - \mathbf{R}_{g,n}^{*\prime}(\boldsymbol{\rho}_{g,n})]\overline{\mathbf{A}}_{s,n}[\mathbf{I}_n - \mathbf{R}_{g,n}^*(\boldsymbol{\rho}_{g,n})]\frac{\partial\mathbf{u}_{g,n}}{\partial\boldsymbol{\delta}_{g,n}}$$
$$= \quad -2\boldsymbol{\varepsilon}_{g,n}'\overline{\mathbf{A}}_{s,n}[\mathbf{I}_n - \mathbf{R}_{g,n}^*(\boldsymbol{\rho}_{g,n})]\mathbf{Z}_{g,n}$$
$$= \quad -\boldsymbol{\varepsilon}_{g,n}'(\mathbf{A}_{s,n} + \mathbf{A}_{s,n}')[\mathbf{I}_n - \mathbf{R}_{g,n}^*(\boldsymbol{\rho}_{g,n})]\mathbf{Z}_{g,n},$$

and

$$\frac{\partial}{\partial\boldsymbol{\rho}_{g,n}}\left[\mathbf{H}_n'\boldsymbol{\varepsilon}_{g,n}\right] = -[\mathbf{H}_n'\mathbf{M}_{r_{g,1},n}\mathbf{u}_{g,n},\ldots,\mathbf{H}_n'\mathbf{M}_{r_{g,q_g},n}\mathbf{u}_{g,n}],$$

$$\frac{\partial}{\partial\boldsymbol{\rho}_{g,n}}[\boldsymbol{\varepsilon}_{g,n}'\mathbf{A}_{s,n}\boldsymbol{\varepsilon}_{g,n}] = -2[\mathbf{u}_{g,n}'\mathbf{M}_{r_{g,1},n}'\overline{\mathbf{A}}_{s,n}\mathbf{u}_{g,n},...,\mathbf{u}_{g,n}'\mathbf{M}_{r_{g,q_g},n}'\overline{\mathbf{A}}_{s,n}\mathbf{u}_{g,n})]\frac{\partial\mathbf{r}_{1,g,n}}{\partial\boldsymbol{\rho}_{g,n}}$$
$$+[\mathbf{u}_{g,n}'\mathbf{M}_{r_{g,1},n}'\overline{\mathbf{A}}_{s,n}\mathbf{M}_{r_{g,1},n}\mathbf{u}_{g,n},...,\mathbf{u}_{g,n}'\mathbf{M}_{r_{g,q_g},n}'\overline{\mathbf{A}}_{s,n}\mathbf{M}_{r_{g,q_g},n}\mathbf{u}_{g,n}]\frac{\partial\mathbf{r}_{2,g,n}}{\partial\boldsymbol{\rho}_{g,n}}$$
$$+2[E\mathbf{u}_{g,n}'\mathbf{M}_{r_{g,1},n}'\overline{\mathbf{A}}_{s,n}\mathbf{M}_{r_{g,2},n}\mathbf{u}_{g,n},...,E\mathbf{u}_{g,n}'\mathbf{M}_{r_{g,q_g-1},n}'\overline{\mathbf{A}}_{s,n}\mathbf{M}_{r_{g,q_g},n}\mathbf{u}_{g,n}]\frac{\partial\mathbf{r}_{3,g,n}}{\partial\boldsymbol{\rho}_{g,n}}.$$

Recalling that

$$\mathbf{m}_{g,n}^{\delta}(\boldsymbol{\rho}_{g,n},\boldsymbol{\delta}_{g,n}) \quad = \quad n^{-1}\mathbf{H}_n'\boldsymbol{\varepsilon}_{g,n},$$

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

we have

$$\frac{\partial \mathbf{m}_{g,n}^{\delta}}{\partial \boldsymbol{\delta}_{g,n}} = -n^{-1}\mathbf{H}_n'\mathbf{Z}_{*g,n}(\boldsymbol{\rho}_{g,n}) = -n^{-1}\left[\mathbf{H}_n'\mathbf{Z}_{g,n} - \sum_{r\in\mathbf{I}_{g,\rho}}\rho_{g,r,n}\mathbf{H}_n'\mathbf{M}_{r,n}\mathbf{Z}_{g,n}\right]$$

$$= -\mathbf{Q}_{HZ,g*}(\boldsymbol{\rho}_{g,n}) + o_p(1),$$

$$\frac{\partial \mathbf{m}_{g,n}^{\delta}}{\partial \boldsymbol{\rho}_{g,n}} = -n^{-1}[\mathbf{H}_n'\mathbf{M}_{r_{g,1},n}\mathbf{u}_{g,n},\ldots,\mathbf{H}_n'\mathbf{M}_{r_{g,q_g},n}\mathbf{u}_{g,n}] = o_p(1),$$

$$\frac{\partial \mathbf{m}_{g,n}^{\rho}}{\partial \boldsymbol{\delta}_{g,n}} = \begin{bmatrix} -n^{-1}\boldsymbol{\varepsilon}_{g,n}'[\mathbf{A}_{1,n}+\mathbf{A}_{1,n}'][\mathbf{I}_n - \mathbf{R}_{g,n}^*(\boldsymbol{\rho}_{g,n})]\mathbf{Z}_{g,n} \\ \vdots \\ -n^{-1}\boldsymbol{\varepsilon}_{g,n}'[\mathbf{A}_{S,n}+\mathbf{A}_{S,n}'][\mathbf{I}_n - \mathbf{R}_{g,n}^*(\boldsymbol{\rho}_{g,n})]\mathbf{Z}_{g,n} \end{bmatrix} = \begin{bmatrix} \alpha_{g,1,n}' \\ \vdots \\ \alpha_{g,S,n}' \end{bmatrix} + o_p(1),$$

$$\frac{\partial \mathbf{m}_{g,n}^{\rho}}{\partial \boldsymbol{\rho}_{g,n}} = -\boldsymbol{\Gamma}_{g,n}(\boldsymbol{\delta}_{g,n})\frac{\partial \mathbf{r}_{g,n}(\boldsymbol{\rho}_{g,n})}{\partial \boldsymbol{\rho}_{g,n}} + o_p(1) = -\mathbf{J}_{g,n} + o_p(1),$$

with $\boldsymbol{\alpha}_{g,r,n} = -n^{-1}E\mathbf{Z}_{g,n}'(\mathbf{I}_n - \mathbf{R}_{g,n}^{*\prime}(\boldsymbol{\rho}_{g,n}))(\mathbf{A}_{r,n}+\mathbf{A}_{r,n}')\boldsymbol{\varepsilon}_{g,n}$ and $\mathbf{J}_{g,n} = \boldsymbol{\Gamma}_{g,n}\frac{\partial \mathbf{r}_{g,n}(\boldsymbol{\rho}_{g,n})}{\partial \boldsymbol{\rho}_{g,n}}$.
Consequently, we have

$$\frac{\partial \mathbf{m}_{g,n}}{\partial \boldsymbol{\theta}_{g,n}} = \mathbf{G}_{g,n} + o_p(1)$$

with

$$\mathbf{G}_{g,n} = \begin{bmatrix} \mathbf{G}_{g,n}^{LL} & \mathbf{G}_{g,n}^{LQ} \\ \mathbf{G}_{g,n}^{QL} & \mathbf{G}_{g,n}^{QQ} \end{bmatrix} = \begin{bmatrix} -\mathbf{Q}_{HZ,g*}(\boldsymbol{\rho}_{g,n}) & \mathbf{0} \\ \boldsymbol{\alpha}_{g,n}' & -\mathbf{J}_{g,n} \end{bmatrix}$$

where $\boldsymbol{\alpha}_{g,n} = [\alpha_{g,1,n},\ldots,\alpha_{g,S,n}]$. Furthermore

$$\mathbf{S}_{g,n} = \mathbf{G}_{g,n}'\begin{bmatrix} (\boldsymbol{\Psi}_{gg,n}^{LL})^{-1} & \mathbf{0} \\ \mathbf{0} & (\boldsymbol{\Psi}_{gg,n}^{QQ})^{-1} \end{bmatrix}\mathbf{G}_{g,n} = \begin{bmatrix} \mathbf{S}_{g,n}^{LL} & \mathbf{S}_{g,n}^{LQ} \\ \mathbf{S}_{g,n}^{QL} & \mathbf{S}_{g,n}^{QQ} \end{bmatrix}$$

where

$$\mathbf{S}_{g,n}^{LL} = \mathbf{G}_{g,n}^{LL\prime}(\boldsymbol{\Psi}_{gg,n}^{LL})^{-1}\mathbf{G}_{g,n}^{LL} + \mathbf{G}_{g,n}^{QL\prime}(\boldsymbol{\Psi}_{gg,n}^{QQ})^{-1}\mathbf{G}_{g,n}^{QL}$$

$$= \sigma_{gg,n}^{-1}\mathbf{Q}_{HZ,g*}(\boldsymbol{\rho}_{g,n})'(n^{-1}\mathbf{H}_n'\mathbf{H}_n)^{-1}\mathbf{Q}_{HZ,g*}(\boldsymbol{\rho}_{g,n}) + \sigma_{gg,n}^{-2}\boldsymbol{\alpha}_{g,n}(\mathbf{K}_n^{QQ})^{-1}\boldsymbol{\alpha}_{g,n}',$$

$$\mathbf{S}_{g,n}^{LQ} = \mathbf{S}_{g,n}^{LQ\prime} = \mathbf{G}_{g,n}^{QL\prime}(\boldsymbol{\Psi}_{gg,n}^{QQ})^{-1}\mathbf{G}_{g,n}^{QQ} = -\sigma_{gg,n}^{-2}\boldsymbol{\alpha}_{g,n}(\mathbf{K}_n^{QQ})^{-1}\mathbf{J}_{g,n},$$

$$\mathbf{S}_{g,n}^{QQ} = \mathbf{G}_{g,n}^{QQ\prime}(\boldsymbol{\Psi}_{gg,n}^{QQ})^{-1}\mathbf{G}_{g,n}^{QQ} = \sigma_{gg,n}^{-2}\mathbf{J}_{g,n}'(\mathbf{K}_n^{QQ})^{-1}\mathbf{J}_{g,n}.$$

The submatrices of $\mathbf{S}_{g,n}$ can be estimated consistently by

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

where $\widehat{\boldsymbol{\alpha}}_{g,n}^o = [\widehat{\alpha}_{g,1,n}^o, \ldots, \widehat{\alpha}_{g,S,n}^o]$ with

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

We next discuss full information estimators. Let $\boldsymbol{\theta}_n = [\boldsymbol{\delta}_n', \boldsymbol{\rho}_n']'$ and $\widehat{\widetilde{\boldsymbol{\theta}}}_n^o = [\widehat{\widetilde{\boldsymbol{\delta}}}_n^{o\prime}, \widehat{\widetilde{\boldsymbol{\rho}}}_n^{o\prime}]'$, then by argumentation analogous to those for two-step estimators we have $(\boldsymbol{\Phi}_n^o)^{-1/2}n^{1/2}(\widehat{\widetilde{\boldsymbol{\theta}}}_n^o - \boldsymbol{\theta}_n) \xrightarrow{d} N[0, I]$ with

$$
\boldsymbol{\Phi}_n^o - \left[\left[\frac{\partial \mathbf{m}_n(\boldsymbol{\theta}_n)}{\partial \boldsymbol{\theta}_n'}\right]\left[\begin{array}{cc} \boldsymbol{\Psi}_n^{LL} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi}_n^{QQ} \end{array}\right]^{-1}\left[\frac{\partial \mathbf{m}_n(\boldsymbol{\theta}_n)}{\partial \boldsymbol{\theta}_n}\right]\right]^{-1} = o_p(1),
$$

where

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

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

Observe that for $g \neq h$,
$$
\frac{\partial \mathbf{m}_{g,n}}{\partial \boldsymbol{\theta}_{h,n}} = 0,
$$
and thus
$$
\frac{\partial \mathbf{m}_n}{\partial \boldsymbol{\theta}_n} = \mathbf{G}_n + o_p(1)
$$
with

$$
\begin{aligned}
\mathbf{G}_n &= \left[\begin{array}{cc} diag_{g=1}^G[\mathbf{G}_{g,n}^{LL}] & diag_{g=1}^G[\mathbf{G}_{g,n}^{LQ}] \\ diag_{g=1}^G[\mathbf{G}_{g,n}^{QL}] & diag_{g=1}^G[\mathbf{G}_{g,n}^{QQ}] \end{array}\right] \\
&= \left[\begin{array}{cc} -diag_{g=1}^G[\mathbf{Q}_{HZ,g*}(\boldsymbol{\rho}_{g,n})] & \mathbf{0} \\ diag_{g=1}^G[\boldsymbol{\alpha}_{g,n}'] & -diag_{g=1}^G[\mathbf{J}_{g,n}] \end{array}\right].
\end{aligned}
$$

14

Furthermore

$$\mathbf{S}_n = \mathbf{G}'_n \left[ \begin{array}{cc} (\mathbf{\Psi}_n^{LL})^{-1} & \mathbf{0} \\ \mathbf{0} & (\mathbf{\Psi}_n^{QQ})^{-1} \end{array} \right] \mathbf{G}_n = \left[ \begin{array}{cc} \mathbf{S}_n^{LL} & \mathbf{S}_n^{LQ} \\ \mathbf{S}_n^{QL} & \mathbf{S}_n^{QQ} \end{array} \right]$$

where

$$
\begin{aligned}
\mathbf{S}_n^{LL} &= \mathbf{G}_n^{LL\prime}(\mathbf{\Psi}_n^{LL})^{-1}\mathbf{G}_n^{LL} + \mathbf{G}_n^{QL\prime}(\mathbf{\Psi}_n^{QQ})^{-1}\mathbf{G}_n^{QL} \\
&= diag_{g=1}^{G}[\mathbf{Q}_{HZ,g*}(\boldsymbol{\rho}_{g,n})'][\mathbf{\Sigma}_{SQ,n}^{-1} \otimes n^{-1}\mathbf{H}'_n\mathbf{H}_n]^{-1}diag_{g=1}^{G}[\mathbf{Q}_{HZ,g*}(\boldsymbol{\rho}_{g,n})] \\
&\quad + diag_{g=1}^{G}[\boldsymbol{\alpha}_{g,n}][\mathbf{\Sigma}_{SQ,n}^{-1} \otimes (\mathbf{K}_n^{QQ})^{-1}]diag_{g=1}^{G}[\boldsymbol{\alpha}'_{g,n}], \\
\mathbf{S}_n^{LQ} &= \mathbf{S}_n^{LQ\prime} = \mathbf{G}_n^{QL\prime}(\mathbf{\Psi}_n^{QQ})^{-1}\mathbf{G}_n^{QQ} \\
&= -diag_{g=1}^{G}[\boldsymbol{\alpha}_{g,n}][\mathbf{\Sigma}_{SQ,n}^{-1} \otimes (\mathbf{K}_n^{QQ})^{-1}]diag_{g=1}^{G}[\mathbf{J}_{g,n}], \\
\mathbf{S}_n^{QQ} &= \mathbf{G}_n^{QQ\prime}(\mathbf{\Psi}_n^{QQ})^{-1}\mathbf{G}_n^{QQ} \\
&= diag_{g=1}^{G}[\mathbf{J}'_{g,n}][\mathbf{\Sigma}_{SQ,n}^{-1} \otimes (\mathbf{K}_n^{QQ})^{-1}]diag_{g=1}^{G}[\mathbf{J}_{g,n}].
\end{aligned}
$$

Let $\widetilde{\sigma}_{gh,n}$ denote some consistent estimator for $\sigma_{gh,n}$, and let $\widetilde{\mathbf{\Sigma}}_n = (\widetilde{\sigma}_{gh,n})$ and $\widetilde{\mathbf{\Sigma}}_{SQ,n} = (\widetilde{\sigma}_{gh,n}^2)$. The submatrices of $\mathbf{S}_n$ can be estimated consistently by

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

where $\widehat{\widehat{\boldsymbol{\alpha}}}_{g,n}^{o\prime} = [\widehat{\widehat{\boldsymbol{\alpha}}}_{g,1,n}^{o}, \ldots, \widehat{\widehat{\boldsymbol{\alpha}}}_{g,S,n}^{o}]'$ with

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

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

# F  Supplemental Remarks on Identification Condition

In the following we provide a number of observations on the identification condition maintained by Assumption 6. As discussed in the text, in general, this high-level condition is a necessary condition for the identification of the regression parameters, based on linear moment conditions only. For ease of presentation we focus the discussion on moment conditions corresponding to the untransformed model (8), and we drop subscripts $n$. For clarity we denote the true regression parameters of the $g$-th equation as $\boldsymbol{\delta}_g^o$, and thus $\mathbf{u}_g = \mathbf{y}_g - \mathbf{Z}_g \boldsymbol{\delta}_g^o$.

The 2SLS estimator exploits the population moment condition $En^{-1}\mathbf{H}'\mathbf{u}_g = 0$, or asymptotically $\operatorname{plim}_{n\to\infty} n^{-1}\mathbf{H}'_n\mathbf{u}_g = 0$. The corresponding sample moment vector satisfies

$$
\begin{aligned}
n^{-1}\mathbf{H}'(\mathbf{y}_g - \mathbf{Z}_g\boldsymbol{\delta}_g) &= n^{-1}\mathbf{H}'\mathbf{u}_g + n^{-1}\mathbf{H}'\mathbf{Z}_g(\boldsymbol{\delta}_g^o - \boldsymbol{\delta}_g) \\
&= \left[\operatorname{plim}_{n\to\infty} n^{-1}\mathbf{H}'\mathbf{Z}_g\right](\boldsymbol{\delta}_g^o - \boldsymbol{\delta}_g) + o_p(1).
\end{aligned}
$$

Under the maintained assumptions $\mathbf{Q}_{HZ,g} = \operatorname{plim}_{n\to\infty} n^{-1}\mathbf{H}'\mathbf{Z}_g = \lim_{n\to\infty} n^{-1}\mathbf{H}'E\mathbf{Z}_g$. Assumption 6 maintains that $\mathbf{Q}_{HZ,g}$ has full column rank. This ensures identification through the instruments $\mathbf{H}$ in that then $\boldsymbol{\delta}_g = \boldsymbol{\delta}_g^o$ is seen to be the unique solution of the $\mathbf{Q}_{HZ,g}(\boldsymbol{\delta}_g^o - \delta_g) = 0$. For ease of presentation we will proceed by discussing the finite sample analogue of the above assumption, i.e., that $\mathbf{H}'E\mathbf{Z}_g$ has full column rank. We will first relate this assumption to the rank condition in a classical simultaneous equation system, and then discuss some situations where the condition does not hold.

## F.1  Interpretation from the Perspective of a Classical System

Our simultaneous equation model with network interactions contains the classical simultaneous equation model without network interactions as a special case. For the classical simultaneous equation model it is well known that a necessary and sufficient condition for the identification of the structural parameters of an equation is that the matrix of reduced form parameters corresponding to the exogenous regressors, which do not appear in that equation, has full column rank; see, e.g. Dhrymes (1978, p. 283). In the following we show that in this special case the identification condition maintained in Assumption 6 is equiva-

lent to the classical assumption. WLOG we consider the first equation of (8), which under the classical setting reduces to

$$\mathbf{y}_1 = \mathbf{Z}_1 \boldsymbol{\delta}_1 + \mathbf{u}_1,$$

where $\mathbf{Z}_1 = [\mathbf{Y}_1, \mathbf{X}_1]$ and $\boldsymbol{\delta}_1 = [\boldsymbol{\beta}'_1, \boldsymbol{\gamma}'_1]'$. For ease of presentation we assume furthermore that the exogenous regressors are arranged such that $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}^*_1]$, where $\mathbf{X}^*_1$ represents the matrix of observation on the exogenous variables that do not appear in the first equation. Then $E\mathbf{Y}_1 = \mathbf{X}_1 \Pi_1 + \mathbf{X}^*_1 \Pi^*_1$, where $\Pi_1$ and $\Pi^*_1$ are the $K_1 \times G_1$ and $(K - K_1) \times G_1$ matrices of reduced form parameters corresponding to $\mathbf{X}_1$ and $\mathbf{X}^*_1$ respectively. For the classical simultaneous equation model we have $\mathbf{H} = \mathbf{X}$. Thus, observing that

$$E\mathbf{Z}_1 = [E\mathbf{Y}_1, \mathbf{X}_1] = \mathbf{X}\mathbf{F}_1, \qquad \mathbf{F}_1 = \begin{bmatrix} \Pi_1 & I_{K_1} \\ \Pi^*_1 & 0 \end{bmatrix},$$

we have $\mathbf{H}'E\mathbf{Z}_1 = (\mathbf{X}'\mathbf{X})\mathbf{F}_1$. Provided the $\mathbf{X}$ are not perfectly multicollinear it follows that $\mathbf{H}'E\mathbf{Z}_1$ has full column rank iff $\mathbf{F}_1$ has full column rank. However, $\mathbf{F}_1$ has full column rank iff $\Pi^*_1$ has full column rank. This proves that for a classical system the assumption that $\mathbf{H}'E\mathbf{Z}_1$ is equivalent to the classical rank condition. Of course, the order condition $K - K_1 \geq G_1$ is necessary for the rank condition to hold.

## F.2 Interpretations from a Stylized Two-Equation Model with Network Interactions

In the following we consider the simple stylized two-equation model:

$$\begin{aligned} \mathbf{y}_1 &= \mathbf{y}_2 \beta_1 + \mathbf{X}_1 \boldsymbol{\gamma}_1 + \lambda_1 \mathbf{W}\mathbf{y}_1 + \mathbf{u}_1 = \mathbf{Z}_1 \boldsymbol{\delta}_1 + \mathbf{u}_1, \\ \mathbf{y}_2 &= \mathbf{X}_2 \boldsymbol{\gamma}_2 + \mathbf{u}_2, \end{aligned}$$

where $\mathbf{Z}_1 = [\mathbf{y}_2, \mathbf{X}_1, \mathbf{W}\mathbf{y}_1]$ and $\boldsymbol{\delta}_1 = [\beta_1, \boldsymbol{\gamma}'_1, \lambda_1]'$. While the setup is simple it permits us to highlight several important scenarios where the identification condition for the parameters of the first equation fails.

We first compute the best instruments $E\mathbf{Z}_1$ for $\mathbf{Z}_1$. Observe that

$$\begin{aligned} \mathbf{y}_1 &= (\mathbf{X_2}\boldsymbol{\gamma_2} + \mathbf{u_2})\beta_1 + \mathbf{X}_1 \boldsymbol{\gamma}_1 + \lambda_1 \mathbf{W}\mathbf{y}_1 + \mathbf{u}_1 \\ &= \mathbf{X}\pi_1 + \lambda_1 \mathbf{W}\mathbf{y}_1 + \mathbf{u}_1 + \mathbf{u_2}\beta_1 \end{aligned}$$

17

with $\boldsymbol{\pi}_1 = (\boldsymbol{\gamma}_1', \beta_1 \boldsymbol{\gamma}_2')'$, and the reduced form for $\mathbf{y}_1$ is given by

$$\mathbf{y}_1 = (\mathbf{I}_n - \lambda_1 \mathbf{W})^{-1}[\mathbf{X}\pi_1 + \mathbf{u}_1 + \mathbf{u}_2\beta_1].$$

Consequently the best instruments $E\mathbf{Z}_1$ for $\mathbf{Z}_1$ is given by

$$
\begin{aligned}
E\mathbf{Z}_1 = [E\mathbf{y}_2, \mathbf{X}_1, \mathbf{W}E\mathbf{y}_1] &= [\mathbf{X}_2\boldsymbol{\gamma}_2, \mathbf{X}_1, \mathbf{W}(\mathbf{I}_n - \lambda_1\mathbf{W})^{-1}\mathbf{X}\pi_1] \\
&= [\mathbf{X}_2\boldsymbol{\gamma}_2, \mathbf{X}_1, \sum_{s=0}^{\infty} \lambda_1^s \mathbf{W}^{s+1}\mathbf{X}\pi_1].
\end{aligned}
$$

The best IV estimator for $\boldsymbol{\delta}_1$ is then given by

$$\widehat{\boldsymbol{\delta}}_1^B = [(E\mathbf{Z}_1)'\mathbf{Z}_1]^{-1}(E\mathbf{Z}_1)'\mathbf{y}_1.$$

From the above we see that the moment condition $E(E\mathbf{Z}_1)'\mathbf{u}_1 = 0$ corresponding to the best instruments is a weighted average of the basic underlying moment conditions

$$E(\mathbf{W}^s\mathbf{X})'\mathbf{u}_1 = 0, \ldots, s = 0, 1, \ldots, \infty. \tag{F.1}$$

The best instruments $E\mathbf{Z}_1$ depend on the inverse of the $n \times n$ matrix $\mathbf{I}_n - \lambda_1\mathbf{W}$, which may be computationally challenging in large samples. In light of (F.1), adapting Kelejian and Prucha (1998), we can define $\mathbf{H}$ to be composed of the linearly independent columns of $\mathbf{W}\mathbf{X}, \ldots, \mathbf{W}^S\mathbf{X}$ for some $S \geq 1$, and work with the moment condition

$$E\mathbf{H}'\mathbf{u}_1 = 0.$$

The corresponding optimal GMM estimator is the S2SLS estimator

$$\widehat{\boldsymbol{\delta}}_1 = [(\widehat{\mathbf{Z}}_1)'\mathbf{Z}_1]^{-1}(\widehat{\mathbf{Z}}_1)'\mathbf{y}_1$$

with $\widehat{\mathbf{Z}}_1 = \mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\mathbf{Z}_1$, and where $\widehat{\mathbf{Z}}_1$ can be viewed to represent an approximation of the ideal instruments.

Kelejian and Prucha (1998) discuss identification based on linear moment restrictions for a single equation cross sectional spatial model. In line with their discussion, clearly for $\mathbf{H}'E\mathbf{Z}_1$ to have full column rank we need $E\mathbf{Z}_1$ to have full column rank. To provide guidance on where this condition may fail, we next discuss several such scenarios.

■ Scenario 1: Suppose the true model does not contain any exogenous variables,

18

i.e., $\boldsymbol{\gamma}_1 = 0$ and $\gamma_2 = 0$. Obviously in this case $\boldsymbol{\pi}_1 = 0$ and $\mathbf{W}E\mathbf{y}_1 = 0$, and thus $E\mathbf{Z}_1$ does not have full column rank. Apart from a complete failure of identification by linear moments under this scenario, we expect the instruments $\mathbf{X}$, $\mathbf{W}\mathbf{X}$, $\mathbf{W}^2\mathbf{X}$, ... to be weak, and estimators based only on linear moment conditions to perform poorly, when the parameters of the exogenous variables are "small". Since the values of $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ depend on the chosen units of measurement of the exogenous variables, it seems intuitive that "small" is best interpreted as to correspond to a small ratio of the variance/signal stemming from the exogenous variables to the variance/noise of the disturbances.

■ Scenario 2: Suppose $\mathbf{W}$ is such that $\mathbf{W}^2 = c_1 \mathbf{I}_n + c_2 \mathbf{W}$, which implies that $(\mathbf{I}_n - \lambda_1 \mathbf{W})^{-1} = d_1 \mathbf{I}_n + d_2 \mathbf{W}$ where the constants $d_1$ and $d_2$ depend on $\lambda_1, c_1, c_2$.[1] Suppose further that $\mathbf{X}_1 = [\underline{\mathbf{X}}_1, \mathbf{W}\underline{\mathbf{X}}_1]$, and correspondingly $\boldsymbol{\gamma}_1 = [\boldsymbol{\gamma}_1^{x\prime}, \boldsymbol{\gamma}_1^{w\prime}]'$, and that $\beta_1 = 0$. In this case we have

$$
\begin{aligned}
\mathbf{W}E\mathbf{y}_1 &= \mathbf{W}(\mathbf{I}_n - \lambda_1 \mathbf{W})^{-1}\mathbf{X}\boldsymbol{\pi}_1 = (d_1 \mathbf{W} + d_2 \mathbf{W}^2)\mathbf{X}\boldsymbol{\pi}_1 \\
&= [c_1 d_2 \mathbf{I}_n + (d_1 + d_2 c_2)\mathbf{W}][\underline{\mathbf{X}}_1 \boldsymbol{\gamma}_1^{x} + \mathbf{W}\underline{\mathbf{X}}_1 \boldsymbol{\gamma}_1^{w}] \\
&= c_1 d_2 \underline{\mathbf{X}}_1 \boldsymbol{\gamma}_1^{x} + (d_1 + d_2 c_2)\mathbf{W}\underline{\mathbf{X}}_1 \boldsymbol{\gamma}_1^{x} \\
&\quad + c_1 d_2 \mathbf{W}\underline{\mathbf{X}}_1 \boldsymbol{\gamma}_1^{w} + (d_1 + d_2 c_2)[c_1 \mathbf{I}_n + c_2 \mathbf{W}]\underline{\mathbf{X}}_1 \boldsymbol{\gamma}_1^{w} \\
&= c_1 d_2 \underline{\mathbf{X}}_1 \boldsymbol{\gamma}_1^{x} + (d_1 + d_2 c_2)c_1 \underline{\mathbf{X}}_1 \boldsymbol{\gamma}_1^{w} \\
&\quad + (d_1 + d_2 c_2)\mathbf{W}\underline{\mathbf{X}}_1 \boldsymbol{\gamma}_1^{x} + c_1 d_2 \mathbf{W}\underline{\mathbf{X}}_1 \boldsymbol{\gamma}_1^{w} + (d_1 + d_2 c_2)c_2 \mathbf{W}\underline{\mathbf{X}}_1 \boldsymbol{\gamma}_1^{w}
\end{aligned}
$$

which is clearly collinear with the columns of $\mathbf{X}_1$. We note that the result is specific. It would generally not extend to the case where $\mathbf{X}_1$ only contains a subset of spatial lags of the exogenous variables. It would generally also not extend to the case where $\beta_1 \neq 0$, i.e., to the case where additional endogenous variables are present that would depend on additional exogenous variables that can serve as instruments.

A leading example where $\mathbf{W}^2 = c_1 \mathbf{I}_n + c_2 \mathbf{W}$, and instrumentation in terms of neighbor's characteristics fails arises if there are $R$ groups of size $m_g$, $g = 1, \ldots, R$, and social interactions take place only within groups, and all members of a group are friends of equal importance. If the calculation of group means includes all members we have $\mathbf{W} = diag_{g=1}^{R}(\mathbf{W}_{m_g})$ with $\mathbf{W}_{m_g} = e_{m_g}e'_{m_g}/m_g$, where $e_{m_g}$ denotes an $m_g \times 1$ vector of ones. If the calculation of group means affecting the $i$-th member excludes the $i$-th member we have $\mathbf{W} = diag_{g=1}^{R}(\mathbf{W}_{m_g})$

---

[1] This is readily verified by observing that $\mathbf{I}_n = (\mathbf{I}_n - \lambda_1 \mathbf{W})(d_1 \mathbf{I}_n + d_2 \mathbf{W})$ and utilizing the expression for $\mathbf{W}^2$.

with $\mathbf{W}_{m_g} = (e_{m_g} e'_{m_g} - I_{m_g})/(m_g - 1)$. Both in the first case and, provided that all groups are of the same size, also in the second case we have $\mathbf{W}^2 = c_1 \mathbf{I}_n + c_2 \mathbf{W}$ and identification via instruments fails. However, in the latter case identification is achievable if there is variation in the group size. For a further discussion of these cases for cross sectional data see Bramoulle, Djebbari and Fortin (2009) and Paula (2017), and Kelejian and Prucha (2002) and Kelejian et al. (2006) for an early discussion of identification in case of equal weights.

■ Scenario 3: Consider the model

$$\mathbf{y}_1 = \underline{\mathbf{X_1}} \boldsymbol{\gamma}_1^x + \mathbf{W} \underline{\mathbf{X}}_1 \boldsymbol{\gamma}_1^w + \lambda_1 \mathbf{W} \mathbf{y}_1 + \mathbf{u}_1 - \lambda_1 \mathbf{W} \mathbf{u}_1.$$

Then $E\mathbf{Z}_1 = [\underline{\mathbf{X_1}}, \boldsymbol{W} \underline{\mathbf{X}}_1, \boldsymbol{W} E \mathbf{y}_1]$ does not have full column rank for parameter constellations where $\boldsymbol{\gamma}_1^w = -\lambda_1 \boldsymbol{\gamma}_1^x$. To see this observe that for those parameter constellations $E\mathbf{y}_1 = \underline{\mathbf{X_1}} \boldsymbol{\gamma}_1^x$. For interpretation, we note that the above model is observationally equivalent to $\mathbf{y}_1 = \underline{\mathbf{X_1}} \boldsymbol{\gamma}_1^x + \mathbf{u}_1$, since pre-multiplication of this model with $\mathbf{I}_n - \lambda_1 \mathbf{W}$ yields the above model.

## F.3 Identification from Linear and Quadratic Moment Conditions

A standard assumption in the literature on GMM estimation is that the probability limit of the matrix of first order derivatives of the moment vector w.r.t. to the parameters has full column rank when evaluated at the true parameter (or a small sample analogue thereof). Let $\mathbf{m}_g^{\delta,u}(\boldsymbol{\delta}_g) = n^{-1} \mathbf{H}' \mathbf{u}_g(\boldsymbol{\delta}_g)$ with $\mathbf{u}_g(\boldsymbol{\delta}_g) = \mathbf{y}_g - \mathbf{Z}_g \boldsymbol{\delta}_g$. Then $\text{plim}_{n \to \infty} \partial \mathbf{m}_g^{\delta,u}(\boldsymbol{\delta}_g^o)/\partial \boldsymbol{\delta}_g = \mathbf{Q}_{HZ,g}$. Furthermore, let $\mathbf{m}_g^{\delta}(\boldsymbol{\rho}_g, \boldsymbol{\delta}_g)$ and $\mathbf{m}_g^{\rho}(\boldsymbol{\rho}_g, \boldsymbol{\delta}_g)$ be defined as in (12) and recall the equivalent definition of the latter in (14). Then, $\text{plim}_{n \to \infty} \partial \mathbf{m}_g^{\delta}(\boldsymbol{\rho}_g^o, \boldsymbol{\delta}_g^o)/\partial \boldsymbol{\delta}_g = \mathbf{Q}_{HZ,g*}(\boldsymbol{\rho}_g)$ and $\text{plim}_{n \to \infty} \partial \mathbf{m}_g^{\rho}(\boldsymbol{\rho}_g, \boldsymbol{\delta}_g)/\partial \boldsymbol{\rho}_g = \boldsymbol{\Gamma}_g[\partial \mathbf{r}_g/\partial \boldsymbol{\rho}_g]$. From this we see that the nature of Assumptions 6 and 7 is in line with assumptions maintained by the classical GMM literature; observe that $\partial \mathbf{r}_g/\partial \boldsymbol{\rho}_g$ has full column rank.

As remarked in the text, the conditions postulated in Assumption 6 are sufficient conditions to ensure identification of the regression parameters $\boldsymbol{\delta}_g$ from the linear moments only. Given $\boldsymbol{\delta}_g$ is identified, the conditions postulated in Assumption 7 are sufficient to ensure the identification of the autoregressive parameters $\boldsymbol{\rho}_g$. Assumptions 6 and 7 are geared towards two-step estimation.

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. In this case a sufficient condition for the parameters $\boldsymbol{\theta}_g = [\boldsymbol{\rho}_g', \boldsymbol{\delta}_g']'$ to be identified is that $\text{plim}_{n \to \infty} \partial \mathbf{m}_g(\boldsymbol{\theta}_g^o)/\partial \boldsymbol{\theta}_g$ has full column rank, where $\mathbf{m}_g(\boldsymbol{\theta}_g)$ denotes the stacked vector of linear and quadratic moment conditions. For contributions on identification with the help of quadratic moment conditions that ensure that $\text{plim}_{n \to \infty} \partial \mathbf{m}_g(\boldsymbol{\theta}_g^o)/\partial \boldsymbol{\theta}_g$ has full column rank see, e.g., Lee (2007a) and Kuersteiner and Prucha (2020) within a single equation framework, and see, e.g., Liu (2014, 2019, 2020), Liu and Saraiva (2019), and Yang and Lee (2017,2019) for contributions within a systems framework. These contributions focus on the case where the disturbances are uncorrelated in the cross section, i.e. where $\boldsymbol{\theta}_g = \boldsymbol{\delta}_g$.

# G  Monte Carlo Study

To analyze the small sample properties of our estimators we have conducted a Monte Carlo study. The study considers both weights matrices motivated by social interactions and spatial interactions. The study explores situations where the parameters are identified by the linear moment conditions alone. It also considers situations where the parameters are identified from utilizing both linear and quadratic moment conditions jointly, but where identification from the linear moment conditions alone is weak.

In Appendix G.1 we describe the Monte Carlo design and provide highlights of the study for two-step estimators for the case where the parameters are identified by linear moment conditions. The simulation results in Appendix G.1 are based on weights matrices corresponding to an underlying social interactions structure. In Appendices G.2 and G.3 we report on additional Monte Carlo simulations for the identified case for weights matrices motivated by social interactions as well as by spatial interactions. We also report on results for the three scenarios where identification is weak discussed in the Appendix F included in this Online Supplementary Appendix.

## G.1  Monte Carlo Design and Main Results

For the Monte Carlo results below we considered the following two equation system as a special case of (2):

$$
\begin{aligned}
\mathbf{y}_1 &= b_{21}\mathbf{y}_2 + [\lambda_{11,1}\mathbf{M}_1 + \lambda_{11,2}\mathbf{M}_2]\mathbf{y}_1 + \sum_{k=1}^{3} c_{k1}\mathbf{x}_k + \mathbf{u}_1, \\
\mathbf{y}_2 &= b_{12}\mathbf{y}_1 + [\lambda_{22,1}\mathbf{M}_1 + \lambda_{22,2}\mathbf{M}_2]\mathbf{y}_2 + \sum_{k=4}^{6} c_{k2}\mathbf{x}_k + \mathbf{u}_2, \\
\mathbf{u}_g &= [\rho_{g1}\mathbf{M}_1 + \rho_{g2}\mathbf{M}_2]\mathbf{u}_g + \boldsymbol{\epsilon}_g, \qquad g = 1,2.
\end{aligned} \tag{G.1}
$$

The stylized social-network design employes a group structure, which can be viewed as emulating groups of friends in a classroom setting. More specifically, suppose there are $P$ schools, and each school has three classrooms of size $m_1$, $m_2$, and $m_3$. Now consider the matrix

$$
E = (e_{ij}) = I_P \otimes \begin{bmatrix} E_{m_1} & 0 & 0 \\ 0 & E_{m_2} & 0 \\ 0 & 0 & E_{m_3} \end{bmatrix}
$$

where the $E_{m_q}$, $q = 1, 2, 3$, are $m_q \times m_q$ matrices with zeros on the diagonal and with ones off the diagonal. Then the elements of $E$ can be viewed as indicator

variables that are equal to one if two students belong to the same classroom, and zero otherwise. The matrix $E$ is of dimension $n \times n$, where the sample size is $n = mP$ with $m = m_1 + m_2 + m_3$. We used the following values in our Monte Carlo simulations:

$$m_1 = 10,\ m_2 = 15,\ m_3 = 25,\ P = 10 \text{ or } 20,$$

which implies a sample size of $n = 500$ or $1,000$.

To generate the weights matrices $\mathbf{M}_1$ and $\mathbf{M}_2$, let $\xi_{Gi}$ and $\xi_{Ii}$ be, respectively, an i.i.d. binary random variable taking values 0 and 1 with equal probability, and an i.i.d. discrete random variable taking values $1, 2, \ldots, 10$ with equal probability, and let $\mu_i$ be i.i.d. $N(0,1)$. Furthermore, $(\xi_{Gi})$, $(\xi_{Ii})$ and $(\mu_i)$ are generated independently. Now define

$$d_{ij} = [c_G(\xi_{Gi} - \xi_{Gj})/\sigma_G + c_I(\xi_{Ii} - \xi_{Ij})/\sigma_I + c_\mu(\mu_i - \mu_j)]/\sigma_\mu$$

where $\sigma_G^2 = 1/4$, $\sigma_I^2 = (100-1)/12$, $\sigma_\mu^2 = 1$ denote the variances of $\xi_{Gi}$, $\xi_{Ii}$, and $\mu_i$, respectively. For an exemplary interpretation, $\xi_{Gi}$ could be an indicator for the gender of an individual, $\xi_{Ii}$ could represent the family income decile of an individual, the $\mu_i$ are unobserved characteristics, and $d_{ij}$ could then be interpreted as a measure of similarity between two individuals. With this interpretation we define

$$
\begin{aligned}
m_{1,ij}^* &= \mathbf{1}(|d_{ij}| < d_*)e_{ij}, \\
m_{2,ij}^* &= \mathbf{1}(d_* \leq |d_{ij}| < d_{**})e_{ij},
\end{aligned}
$$

where $m_{1,ij}^*$ and $m_{2,ij}^*$ are now indicators that equal one, respectively, if two individuals are best friends or just friends, and zero otherwise.

Specific values that generate on average about 29% best friends and 40% friends are:

$$c_G = .4,\ c_I = .4,\ c_\mu = .2,\ d_* = .3,\ d_{**} = .8.$$

The weights matrices $\mathbf{M}_1 = [m_{1,ij}]$ and $\mathbf{M}_2 = [m_{2,ij}]$ are then obtained by applying the following normalization ($s = 1, 2$): $m_{s,ij} = m_{s,ij}^* / \sum_{j=1}^n m_{s,ij}^*$ if $\sum_{j=1}^n m_{s,ij}^* > 0$ and $m_{s,ij} = m_{s,ij}^*$ if $\sum_{j=1}^n m_{s,ij}^* = 0$. Note that the design allows for situations where an individual has no close friends or only close friends. That is, we allow for a row of $\mathbf{M}_1$ or $\mathbf{M}_2$ to only contain zeros. If a row contains

23

nonzero elements, then that row is normalized so that the row sum is one.

We consider three sets of parameters of model (G.1). Set I corresponds to positive spillovers, Set II to negative spillovers and Set III to zero spillovers. In particular, we consider the parameter values as given in Table 1:

Table 1: Configuration of Autoregressive Parameters in Set I-III

| | Auroregressive Parameters | | | | | | | |
| | Equation 1 | | | | Equation 2 | | | |
| | $\lambda_{11,1}$ | $\lambda_{11,2}$ | $\rho_{11}$ | $\rho_{12}$ | $\lambda_{22,1}$ | $\lambda_{22,2}$ | $\rho_{21}$ | $\rho_{22}$ |
|---|---|---|---|---|---|---|---|---|
| Set I | 0.30 | 0.20 | 0.20 | 0.10 | 0.30 | 0.15 | 0.10 | 0 |
| Set II | -0.30 | -0.20 | -0.20 | -0.10 | -0.30 | -0.15 | -0.10 | 0 |
| Set III | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

The remaining parameters are selected as $b_{12} = 0.3$, $b_{21} = 0.15$, $c_{1k} = 1$ for $k = 1, 2, 3$, and $c_{2k} = 1$ for $k = 4, 5, 6$. The observations on the exogenous regressors $\mathbf{x}_1, \ldots, \mathbf{x}_6$ are kept fixed for all Monte Carlo iterations, and are generated as independent of each other and as cross sectionally $i.i.d.$ N(1,3). The disturbances $\boldsymbol{\epsilon}_1$, $\boldsymbol{\epsilon}_2$ are generated as cross sectionally i.i.d. normal with mean 0, variance 1 and covariance .5.

In Table 2, given at the end of this subsection, we report on the bias and root mean squared error (RMSE) of the maximum-likelihood estimator (ML), the GS2SLS, and the GS3SLS estimator for parameter Set I based on $1,000$ Monte Carlo repetitions.[2] More specifically, to simplify the presentation, in Table 2 we only report on a subset of the parameters of the first equation of Model (G.1), corresponding to $\mathbf{y}_2$ as well as on the autoregressive parameters. For all estimators the biases are fairly small, indicating that the linear moments alone are able to identify the regression parameters. As expected, the ML estimator has the smallest RMSE. In general, in terms of RMSE, the ML only dominates GS3SLS slightly, and GS3SLS dominates GS2SLS. The differences in RMSE are the most pronounced for the estimates of the autoregressive parameters in the

---

[2]Our measure of bias is defined as the difference between the median and the true parameter value. Our measure corresponding to the RMSE is defined as $[bias^2 + (IQ/1.35)^2]^{1/2}$ where $IQ$ is the inter-quantile range. That is, $IQ = c_1 - c_2$ where $c_1$ is the .75 quantile and $c_2$ is the .25 quantile. If the distribution is normal, $IQ/1.35$ is (apart from rounding errors) equal to the standard deviation. In the following we will refer to our measures simply as bias and RMSE.

disturbance process. As expected, biases and RMSE decline with the sample size.

The results in Table 2 represent a subset of the results reported in Tables 4-6 in Appendix G.2. In Tables 4-6 we report on the estimators of all the parameters of both equations for parameter Sets I-III. In addition to considering the above described scenarios with social network interactions, we also report in Tables 7-9 in Appendix G.2 on results from scenarios with spatial network interactions, using a spatial rook design. In general, the results are in line with those reported above. In the Appendix G.3 we also cover several scenarios where identification by linear moment conditions alone is weak, and where in consequence GS2SLS and GS3SLS can be substantially biased. Our extended Monte Carlo results also report on the performance of LQ-GS2SLS and LQ-GS3SLS. Under weak identification LQ-GS2SLS and LQ-GS3SLS can greatly outperform GS2SLS and GS3SLS. However, for the well identified scenarios underlying the results in Tables 4-9 the benefit of combining linear and quadratic moment conditions seems limited.

A leading hypothesis of interest is the absence of spillovers. Focusing on equation 1 we can test the hypothesis $H_0 : \lambda_{11,1}^0 = \lambda_{11,2}^0 = \rho_{11}^0 = \rho_{12}^0 = 0$. In Table 3, given at the end of this subsection, we report on the power of the Wald test for this hypothesis based on GS2SLS and GS3SLS estimates. More specifically, we explore the power of the test for parameter values $(\lambda_{11,1}^0, \lambda_{11,2}^0, \rho_{11}^0, \rho_{12}^0) = \kappa(\lambda_{11,1}^I, \lambda_{11,2}^I, \rho_{11}^I, \rho_{12}^I)$ where $(\lambda_{11,1}^I, \lambda_{11,2}^I, \rho_{11}^I, \rho_{12}^I) = (.30, .30, .20, .10)$ is equal to the values considered by parameter Set I, and where the factor $\kappa = 0, 0.05, .., 0.30$. All other parameter values are kept as in parameter Set I. The significance level of the tests are close to the nominal 5 percent, especially for sample size $n = 1,000$, and the tests seem to have good power.

Table 2: Bias and RMSE of MLE, GS2SLS and GS3SLS Parameters of Equation 1, Social Interaction Weights Matrices

| | | Soc. Interact. Matrices; $n=500$ | | | | | | Soc. Interact. Matrices; $n=1000$ | | | | | |
| | | MLE | | GS2SLS | | GS3SLS | | MLE | | GS2SLS | | GS3SLS | |
| Parameter | True | Bias | RMSE | Bias | RMSE | Bias | RMSE | Bias | RMSE | Bias | RMSE | Bias | RMSE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Eq. 1 | Set I | | | | | | | | | | | | |
| $b_{21}$ | 0.15 | 0.00047 | 0.01372 | 0.00304 | 0.01358 | 0.00074 | 0.01357 | -0.00004 | 0.00973 | 0.00175 | 0.01011 | 0.00031 | 0.00978 |
| $\lambda_{11,1}$ | 0.3 | 0.00030 | 0.01556 | 0.00156 | 0.01708 | 0.00281 | 0.01551 | -0.00089 | 0.01129 | -0.00084 | 0.01276 | 0.00016 | 0.01147 |
| $\lambda_{11,2}$ | 0.2 | -0.00028 | 0.01651 | -0.00366 | 0.01849 | -0.00209 | 0.01800 | 0.00072 | 0.01093 | -0.00029 | 0.01202 | -0.00007 | 0.01121 |
| $\rho_{11}$ | 0.2 | -0.00536 | 0.05467 | -0.00261 | 0.05724 | -0.00271 | 0.05807 | -0.00157 | 0.03793 | 0.00074 | 0.04028 | 0.00021 | 0.04068 |
| $\rho_{12}$ | 0.1 | -0.00113 | 0.06185 | 0.00162 | 0.07189 | 0.00137 | 0.06941 | -0.00338 | 0.04305 | -0.00043 | 0.04806 | -0.00046 | 0.04829 |
| Eq. 1 | Set II | | | | | | | | | | | | |
| $b_{21}$ | 0.15 | 0.00051 | 0.01319 | 0.00451 | 0.01360 | 0.00208 | 0.01333 | -0.00005 | 0.00963 | 0.00225 | 0.01002 | 0.00095 | 0.00952 |
| $\lambda_{11,1}$ | -0.3 | 0.00057 | 0.02187 | -0.00054 | 0.02364 | -0.00017 | 0.02262 | -0.00124 | 0.01310 | -0.00185 | 0.01475 | -0.00152 | 0.01313 |
| $\lambda_{11,2}$ | -0.2 | 0.00013 | 0.02256 | -0.00204 | 0.02572 | -0.00096 | 0.02267 | 0.00160 | 0.01311 | 0.00039 | 0.01462 | 0.00053 | 0.01359 |
| $\rho_{11}$ | -0.2 | -0.00583 | 0.06664 | 0.00098 | 0.06941 | -0.00040 | 0.07170 | -0.00098 | 0.04289 | 0.00198 | 0.04644 | 0.00138 | 0.04717 |
| $\rho_{12}$ | -0.1 | -0.00048 | 0.07936 | 0.00398 | 0.08989 | 0.00403 | 0.09009 | -0.00279 | 0.05461 | -0.00069 | 0.06176 | -0.00095 | 0.05952 |
| Eq. 1 | Set III | | | | | | | | | | | | |
| $b_{21}$ | 0.15 | 0.00051 | 0.01351 | 0.00411 | 0.01367 | 0.00153 | 0.01336 | -0.00019 | 0.00980 | 0.00216 | 0.01014 | 0.00073 | 0.00982 |
| $\lambda_{11,1}$ | 0 | 0.00064 | 0.02031 | 0.00157 | 0.02164 | 0.00258 | 0.01980 | -0.00137 | 0.01282 | -0.00125 | 0.01434 | -0.00053 | 0.01274 |
| $\lambda_{11,2}$ | 0 | -0.00035 | 0.02179 | -0.00331 | 0.02369 | -0.00211 | 0.02164 | 0.00117 | 0.01311 | 0.00020 | 0.01429 | 0.00034 | 0.01296 |
| $\rho_{11}$ | 0 | -0.00538 | 0.06453 | -0.00145 | 0.06674 | -0.00351 | 0.06660 | -0.00084 | 0.04160 | 0.00055 | 0.04401 | -0.00010 | 0.04533 |
| $\rho_{12}$ | 0 | -0.00066 | 0.07220 | 0.00343 | 0.08182 | 0.00414 | 0.08272 | -0.00404 | 0.05041 | -0.00108 | 0.05518 | -0.00073 | 0.05425 |

Table 3: Power Function of Joint Wald Tests Corresponding to GS2SLS and GS3SLS for $H_0 : \lambda_{11,1}^0 = \lambda_{11,2}^0 = \rho_{11}^0 = \rho_2^0 = 0$. The True Autoregressive Parameters Equal Those of Set I Scaled by $\kappa$. The Significance Level is Displayed at $\kappa = 0$.

| | Social Interactions Matrices | | | |
| | $n = 500$ | | $n = 1000$ | |
| $\kappa$ | GS2SLS | GS3SLS | GS2SLS | GS3SLS |
|---|---|---|---|---|
| 0 | 0.068 | 0.072 | 0.058 | 0.054 |
| 0.05 | 0.129 | 0.170 | 0.226 | 0.272 |
| 0.1 | 0.423 | 0.514 | 0.773 | 0.840 |
| 0.15 | 0.832 | 0.889 | 0.993 | 0.999 |
| 0.2 | 0.987 | 0.994 | 1 | 1 |
| 0.25 | 1 | 1 | 1 | 1 |
| 0.3 | 1 | 1 | 1 | 1 |

## G.2 Additional Monte Carlo Simulations for the Identified Case

Tables 4-6, given at the end of this subsection, are an expansion of Table 2. In Tables 4-6 we report additionally on the performance of the one-step LQ-GS2SLS and LQ-GS3SLS estimators, as well as on a larger set of parameters. More specifically, in Tables 4-6 we provide information on the bias and RMSE for various estimators of the parameters in equations 1 and 2 of Model (G.1) for parameter Sets I-III and sample sizes $n = 500$ and $n = 1,000$ each. Tables 4-6 are based on the same design of the social network weights matrices as those underlying Table 2.

The results in Tables 4-6 are in line with the subset of results reported in Table 2. For all estimators the biases are fairly small, indicating that the linear moments alone are able to identify the regression parameters. As expected, the ML estimator has the smallest RMSE. In general, in terms of RMSE, the ML only dominates GS3SLS slightly, and GS3SLS dominates G2SLS. The differences in RMSE are the most pronounced for the estimates of the autoregressive parameters in the disturbance process. The differences in RMSE for the parameters of the exogenous parameters are especially small. As expected, biases and RMSE decline with the sample size. In Tables 4-6 we also report on the performance of the LQ-GS2SLS and LQ-GS3SLS. While LQ-GS2SLS and LQ-GS3SLS have the potential to greatly outperform GS2SLS and GS3SLS under weak identification, for the well identified scenarios underlying the results in Tables 4-6 the benefit of combining linear and quadratic moment conditions seems limited.

In Tables 7-9, given at the end of this subsection, we provide information on the bias and RMSE for the same estimators for the parameters of Model (G.1) as in Tables 4-6, but for an alternative set of weights matrices. We explore the same parameter specifications, but now generate the data based on weights matrices "inspired" by a spatial network. More specifically, we derive the weights matrices from a classical rook design. In more detail, to define locations and neighbors, consider a square grid with both the $x$ and $y$ coordinates only taking on the values $1, 2, ..., m$. Next, define the Euclidean distance between any pair of units, $i_1$ and $i_2$ with coordinates $(x_1, y_1)$ and $(x_2, y_2)$, respectively, as $d(i_1, i_2) = \left[ (x_1 - x_2)^2 + (y_1 - y_2)^2 \right]^{1/2}$. Moreover, define the cutoff distances $d_* = 1$ and $d_{**} = 2$ to determine inner- and outer-ring neighbors, respectively, around any spatial unit on the lattice. Now define the $(i, j)$-th element of a

27

row-normalized weights matrix $\mathbf{M}_1$ and $\mathbf{M}_2$ as

$$m_{r,ij} = m_{r,ij}^*/\sum_{j=1}^{n} m_{r,ij}^*, \quad r \in 1,2, \tag{G.2}$$

$$m_{1,ij}^* = \begin{cases} 1 & \text{if} \quad 0 < d(i_1,i_2) \le d_* \\ 0 & \text{else} \end{cases}, \tag{G.3}$$

$$m_{2,ij}^* = \begin{cases} 1 & \text{if} \quad d_* < d(i_1,i_2) \le d_{**} \\ 0 & \text{else} \end{cases}. \tag{G.4}$$

We consider two configurations. The two configurations correspond to $m = 22$ and $m = 31$, which implies a sample size of $n = 484$ and $n = 961$, respectively. The finding in Tables 7-9 are similar to those in Tables 4-6.

Table 4: Bias and RMSE of Various Estimators for Parameters of Equation 1
and 2, Parameter Set I with Social Interaction Weights Matrices

| | | MLE | | GS2SLS | | GS3SLS | | LQ-GS2SLS | | LQ-GS3SLS | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Soc. Interact. Weights Matrices; $n = 500$ | | | | | | | | | |
| Parameter | True | Bias | RMSE | Bias | RMSE | Bias | RMSE | Bias | RMSE | Bias | RMSE |
| Eq. 1 | Set I | | | | | | | | | | |
| $b_{21}$ | 0.15 | 0.00047 | 0.01372 | 0.00304 | 0.01358 | 0.00074 | 0.01357 | 0.00538 | 0.01459 | 0.00256 | 0.01356 |
| $\lambda_{11,1}$ | 0.3 | 0.00030 | 0.01556 | 0.00156 | 0.01708 | 0.00281 | 0.01551 | -0.00178 | 0.01656 | -0.00076 | 0.01570 |
| $\lambda_{11,2}$ | 0.2 | -0.00028 | 0.01651 | -0.00366 | 0.01849 | -0.00209 | 0.01800 | -0.00294 | 0.01911 | -0.00146 | 0.01784 |
| $\rho_{11}$ | 0.2 | -0.00536 | 0.05467 | -0.00261 | 0.05724 | -0.00271 | 0.05807 | 0.03205 | 0.07140 | 0.03635 | 0.07269 |
| $\rho_{12}$ | 0.1 | -0.00113 | 0.06185 | 0.00162 | 0.07189 | 0.00137 | 0.06941 | 0.01227 | 0.07803 | 0.01232 | 0.07579 |
| $c_{11}$ | 1 | -0.00072 | 0.02415 | -0.00124 | 0.02677 | -0.00131 | 0.02407 | -0.00090 | 0.02739 | -0.00065 | 0.02393 |
| $c_{12}$ | 1 | -0.00089 | 0.02580 | -0.00385 | 0.02596 | -0.00189 | 0.02559 | -0.00289 | 0.02587 | -0.00218 | 0.02561 |
| $c_{13}$ | 1 | 0.00034 | 0.02428 | -0.00288 | 0.02531 | -0.00040 | 0.02425 | -0.00336 | 0.02558 | -0.00037 | 0.02454 |
| Eq. 2 | Set I | | | | | | | | | | |
| $b_{12}$ | 0.3 | -0.00068 | 0.01335 | 0.00219 | 0.01356 | -0.00121 | 0.01331 | 0.00388 | 0.01374 | 0.00031 | 0.01320 |
| $\lambda_{22,1}$ | 0.3 | 0.00061 | 0.01451 | 0.00194 | 0.01540 | 0.00271 | 0.01496 | -0.00041 | 0.01579 | 0.00030 | 0.01509 |
| $\lambda_{22,2}$ | 0.15 | -0.00052 | 0.01457 | -0.00321 | 0.01693 | -0.00130 | 0.01547 | -0.00199 | 0.01685 | -0.00093 | 0.01584 |
| $\rho_{21}$ | 0.1 | -0.01171 | 0.05897 | -0.00924 | 0.06391 | -0.00992 | 0.06379 | 0.01850 | 0.07065 | 0.02952 | 0.07403 |
| $\rho_{22}$ | 0 | -0.00412 | 0.06792 | -0.00091 | 0.07142 | -0.00213 | 0.07009 | 0.00954 | 0.07986 | 0.01287 | 0.07974 |
| $c_{24}$ | 1 | -0.00118 | 0.02173 | -0.00160 | 0.02191 | -0.00157 | 0.02139 | -0.00112 | 0.02288 | -0.00011 | 0.02193 |
| $c_{25}$ | 1 | 0.00175 | 0.02423 | -0.00100 | 0.02605 | 0.00190 | 0.02437 | -0.00068 | 0.02658 | 0.00222 | 0.02438 |
| $c_{26}$ | 1 | 0.00192 | 0.02268 | 0.00005 | 0.02285 | 0.00128 | 0.02266 | 0.00008 | 0.02326 | 0.00207 | 0.02313 |
| | | $n = 1000$ | | | | | | | | | |
| Eq. 1 | Set I | | | | | | | | | | |
| $b_{21}$ | 0.15 | -0.00004 | 0.00973 | 0.00175 | 0.01011 | 0.00031 | 0.00978 | 0.00264 | 0.01025 | 0.00138 | 0.00968 |
| $\lambda_{11,1}$ | 0.3 | -0.00089 | 0.01129 | -0.00084 | 0.01276 | 0.00016 | 0.01147 | -0.00221 | 0.01308 | -0.00132 | 0.01161 |
| $\lambda_{11,2}$ | 0.2 | 0.00072 | 0.01093 | -0.00029 | 0.01202 | -0.00007 | 0.01121 | 0.00038 | 0.01202 | 0.00039 | 0.01140 |
| $\rho_{11}$ | 0.2 | -0.00157 | 0.03793 | 0.00074 | 0.04028 | 0.00021 | 0.04068 | 0.01550 | 0.04566 | 0.01838 | 0.04524 |
| $\rho_{12}$ | 0.1 | -0.00338 | 0.04305 | -0.00043 | 0.04806 | -0.00046 | 0.04829 | 0.00305 | 0.04854 | 0.00448 | 0.04896 |
| $c_{11}$ | 1 | -0.00001 | 0.01585 | -0.00108 | 0.01704 | -0.00060 | 0.01586 | -0.00132 | 0.01711 | -0.00011 | 0.01604 |
| $c_{12}$ | 1 | -0.00026 | 0.01632 | -0.00130 | 0.01798 | -0.00079 | 0.01634 | -0.00146 | 0.01786 | -0.00016 | 0.01630 |
| $c_{13}$ | 1 | -0.00025 | 0.01661 | -0.00107 | 0.01858 | -0.00050 | 0.01661 | -0.00136 | 0.01911 | -0.00033 | 0.01636 |
| Eq. 2 | Set I | | | | | | | | | | |
| $b_{12}$ | 0.3 | -0.00001 | 0.00912 | 0.00140 | 0.00949 | 0.00003 | 0.00941 | 0.00198 | 0.00953 | 0.00080 | 0.00931 |
| $\lambda_{22,1}$ | 0.3 | 0.00034 | 0.00957 | 0.00106 | 0.01111 | 0.00152 | 0.01020 | -0.00025 | 0.01136 | -0.00002 | 0.01017 |
| $\lambda_{22,2}$ | 0.15 | 0.00012 | 0.01052 | -0.00231 | 0.01154 | -0.00107 | 0.01058 | -0.00158 | 0.01154 | -0.00041 | 0.01074 |
| $\rho_{21}$ | 0.1 | -0.00593 | 0.04269 | -0.00494 | 0.04359 | -0.00404 | 0.04287 | 0.00970 | 0.04726 | 0.01557 | 0.04763 |
| $\rho_{22}$ | 0 | -0.00424 | 0.04485 | -0.00132 | 0.04997 | -0.00365 | 0.04912 | -0.00089 | 0.05335 | 0.00104 | 0.05281 |
| $c_{24}$ | 1 | 0.00051 | 0.01672 | -0.00047 | 0.01858 | 0.00056 | 0.01685 | -0.00075 | 0.01876 | 0.00089 | 0.01671 |
| $c_{25}$ | 1 | 0.00004 | 0.01698 | -0.00065 | 0.01838 | 0.00003 | 0.01667 | -0.00056 | 0.01795 | 0.00019 | 0.01634 |
| $c_{26}$ | 1 | -0.00046 | 0.01745 | -0.00087 | 0.01894 | -0.00037 | 0.01749 | -0.00067 | 0.01843 | -0.00005 | 0.01771 |

Table 5: Bias and RMSE of Various Estimators for Parameters of Equation 1 and 2, Parameter Set II with Social Interaction Weights Matrices

| | | Soc. Interact. Weights Matrices; $n = 500$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Parameter | True | MLE | | GS2SLS | | GS3SLS | | LQ-GS2SLS | | LQ-GS3SLS | |
| Eq. 1 | Set II | Bias | RMSE | Bias | RMSE | Bias | RMSE | Bias | RMSE | Bias | RMSE |
| $b_{21}$ | 0.15 | 0.00051 | 0.01319 | 0.00451 | 0.01360 | 0.00208 | 0.01333 | 0.00522 | 0.01390 | 0.00320 | 0.01369 |
| $\lambda_{11,1}$ | -0.3 | 0.00057 | 0.02187 | -0.00054 | 0.02364 | -0.00017 | 0.02262 | -0.00193 | 0.02467 | -0.00248 | 0.02260 |
| $\lambda_{11,2}$ | -0.2 | 0.00013 | 0.02256 | -0.00204 | 0.02572 | -0.00096 | 0.02267 | -0.00259 | 0.02628 | -0.00122 | 0.02354 |
| $\rho_{11}$ | -0.2 | -0.00583 | 0.06664 | 0.00098 | 0.06941 | -0.00040 | 0.07170 | 0.01529 | 0.08047 | 0.01857 | 0.08446 |
| $\rho_{12}$ | -0.1 | -0.00048 | 0.07936 | 0.00398 | 0.08989 | 0.00403 | 0.09009 | 0.01167 | 0.09889 | 0.01737 | 0.09786 |
| $c_{11}$ | 1 | 0.00081 | 0.02336 | -0.00170 | 0.02603 | 0.00043 | 0.02369 | -0.00181 | 0.02624 | 0.00060 | 0.02415 |
| $c_{12}$ | 1 | -0.00103 | 0.02473 | -0.00270 | 0.02627 | -0.00073 | 0.02483 | -0.00270 | 0.02603 | -0.00027 | 0.02502 |
| $c_{13}$ | 1 | -0.00064 | 0.02382 | -0.00328 | 0.02435 | -0.00088 | 0.02387 | -0.00385 | 0.02484 | -0.00063 | 0.02430 |
| Eq. 2 | Set II | | | | | | | | | | |
| $b_{12}$ | 0.3 | -0.00051 | 0.01284 | 0.00368 | 0.01333 | 0.00063 | 0.01272 | 0.00460 | 0.01373 | 0.00183 | 0.01314 |
| $\lambda_{22,1}$ | -0.3 | -0.00001 | 0.02002 | -0.00048 | 0.02121 | -0.00035 | 0.02044 | -0.00236 | 0.02247 | -0.00223 | 0.02137 |
| $\lambda_{22,2}$ | -0.15 | -0.00011 | 0.02255 | -0.00181 | 0.02561 | -0.00109 | 0.02320 | -0.00095 | 0.02537 | -0.00107 | 0.02332 |
| $\rho_{2,1}$ | -0.1 | -0.01172 | 0.06573 | -0.00280 | 0.07240 | -0.00337 | 0.07011 | 0.01500 | 0.07719 | 0.02104 | 0.07884 |
| $\rho_{2,2}$ | 0 | -0.00309 | 0.07527 | 0.00096 | 0.08191 | 0.00067 | 0.08050 | 0.01465 | 0.09191 | 0.01799 | 0.08863 |
| $c_{24}$ | 1 | -0.00137 | 0.02099 | -0.00218 | 0.02283 | -0.00098 | 0.02083 | -0.00211 | 0.02312 | 0.00028 | 0.02141 |
| $c_{25}$ | 1 | 0.00188 | 0.02360 | -0.00094 | 0.02559 | 0.00193 | 0.02378 | -0.00080 | 0.02598 | 0.00219 | 0.02341 |
| $c_{26}$ | 1 | 0.00154 | 0.02132 | 0.00044 | 0.02186 | 0.00186 | 0.02115 | -0.00006 | 0.02244 | 0.00263 | 0.02185 |
| | | $n = 1000$ | | | | | | | | | |
| Eq. 1 | Set II | | | | | | | | | | |
| $b_{21}$ | 0.15 | -0.00005 | 0.00963 | 0.00225 | 0.01002 | 0.00095 | 0.00952 | 0.00276 | 0.01010 | 0.00164 | 0.00944 |
| $\lambda_{11,1}$ | -0.3 | -0.00124 | 0.01310 | -0.00185 | 0.01475 | -0.00152 | 0.01313 | -0.00248 | 0.01496 | -0.00200 | 0.01315 |
| $\lambda_{11,2}$ | -0.2 | 0.00160 | 0.01311 | 0.00039 | 0.01462 | 0.00053 | 0.01359 | 0.00064 | 0.01485 | 0.00041 | 0.01369 |
| $\rho_{11}$ | -0.2 | -0.00098 | 0.04289 | 0.00198 | 0.04644 | 0.00138 | 0.04717 | 0.00584 | 0.05098 | 0.00817 | 0.04907 |
| $\rho_{12}$ | -0.1 | -0.00279 | 0.05461 | -0.00069 | 0.06176 | -0.00095 | 0.05952 | 0.00522 | 0.06380 | 0.00545 | 0.06143 |
| $c_{11}$ | 1 | -0.00005 | 0.01546 | -0.00106 | 0.01656 | -0.00043 | 0.01574 | -0.00120 | 0.01689 | -0.00002 | 0.01566 |
| $c_{12}$ | 1 | -0.00084 | 0.01666 | -0.00189 | 0.01768 | -0.00079 | 0.01661 | -0.00186 | 0.01789 | -0.00063 | 0.01655 |
| $c_{13}$ | 1 | -0.00031 | 0.01654 | -0.00131 | 0.01828 | -0.00033 | 0.01661 | -0.00148 | 0.01816 | 0.00014 | 0.01655 |
| Eq. 2 | Set II | | | | | | | | | | |
| $b_{12}$ | 0.3 | -0.00024 | 0.00878 | 0.00135 | 0.00868 | 0.00021 | 0.00876 | 0.00209 | 0.00885 | 0.00079 | 0.00874 |
| $\lambda_{22,1}$ | -0.3 | 0.00054 | 0.01225 | 0.00063 | 0.01402 | 0.00008 | 0.01245 | -0.00016 | 0.01392 | -0.00099 | 0.01262 |
| $\lambda_{22,2}$ | -0.15 | 0.00016 | 0.01450 | -0.00128 | 0.01557 | -0.00078 | 0.01466 | -0.00122 | 0.01616 | -0.00086 | 0.01457 |
| $\rho_{21}$ | -0.1 | -0.00490 | 0.04463 | -0.00314 | 0.04672 | -0.00352 | 0.04549 | 0.00840 | 0.04993 | 0.00925 | 0.04872 |
| $\rho_{22}$ | 0 | -0.00564 | 0.05227 | -0.00058 | 0.05685 | -0.00180 | 0.05542 | 0.00351 | 0.06174 | 0.00394 | 0.06008 |
| $c_{24}$ | 1 | 0.00065 | 0.01635 | -0.00079 | 0.01796 | 0.00075 | 0.01611 | -0.00106 | 0.01797 | 0.00128 | 0.01631 |
| $c_{25}$ | 1 | -0.00005 | 0.01689 | -0.00066 | 0.01835 | 0.00042 | 0.01672 | -0.00066 | 0.01816 | 0.00035 | 0.01691 |
| $c_{26}$ | 1 | -0.00020 | 0.01705 | -0.00054 | 0.01800 | 0.00040 | 0.01720 | -0.00040 | 0.01794 | 0.00008 | 0.01758 |

Table 6: Bias and RMSE of Various Estimators for Parameters of Equation 1
and 2, Parameter Set III with Social Interaction Weights Matrices

| | | Soc. Interact. Weights Matrices; $n = 500$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Parameter | True | MLE | | GS2SLS | | GS3SLS | | LQ-GS2SLS | | LQ-GS3SLS | |
| Eq. 1 | Set III | Bias | RMSE | Bias | RMSE | Bias | RMSE | Bias | RMSE | Bias | RMSE |
| $b_{21}$ | 0.15 | 0.00051 | 0.01351 | 0.00411 | 0.01367 | 0.00153 | 0.01336 | 0.00532 | 0.01429 | 0.00296 | 0.01359 |
| $\lambda_{11,1}$ | 0 | 0.00064 | 0.02031 | 0.00157 | 0.02164 | 0.00258 | 0.01980 | -0.00137 | 0.02182 | -0.00065 | 0.02038 |
| $\lambda_{11,2}$ | 0 | -0.00035 | 0.02179 | -0.00331 | 0.02369 | -0.00211 | 0.02164 | -0.00256 | 0.02472 | -0.00116 | 0.02210 |
| $\rho_{11}$ | 0 | -0.00538 | 0.06453 | -0.00145 | 0.06674 | -0.00351 | 0.06660 | 0.02438 | 0.07803 | 0.03037 | 0.08000 |
| $\rho_{12}$ | 0 | -0.00066 | 0.07220 | 0.00343 | 0.08182 | 0.00414 | 0.08272 | 0.01411 | 0.09080 | 0.01733 | 0.08996 |
| $c_{11}$ | 1 | -0.00008 | 0.02398 | -0.00109 | 0.02643 | -0.00098 | 0.02389 | -0.00140 | 0.02655 | 0.00016 | 0.02400 |
| $c_{12}$ | 1 | -0.00064 | 0.02561 | -0.00349 | 0.02604 | -0.00152 | 0.02548 | -0.00307 | 0.02642 | -0.00157 | 0.02532 |
| $c_{13}$ | 1 | 0.00004 | 0.02425 | -0.00380 | 0.02496 | -0.00081 | 0.02396 | -0.00412 | 0.02459 | 0.00003 | 0.02428 |
| Eq. 2 | Set III | | | | | | | | | | |
| $b_{12}$ | 0.3 | -0.00058 | 0.01339 | 0.00296 | 0.01351 | -0.00019 | 0.01331 | 0.00444 | 0.01403 | 0.00156 | 0.01321 |
| $\lambda_{22,1}$ | 0 | 0.00022 | 0.01861 | 0.00168 | 0.02032 | 0.00204 | 0.01930 | -0.00131 | 0.02129 | -0.00100 | 0.01936 |
| $\lambda_{22,2}$ | 0 | -0.00003 | 0.02049 | -0.00322 | 0.02294 | -0.00204 | 0.02120 | -0.00174 | 0.02329 | -0.00087 | 0.02135 |
| $\rho_{21}$ | 0 | -0.01218 | 0.06197 | -0.00768 | 0.06843 | -0.00745 | 0.06707 | 0.01536 | 0.07472 | 0.02470 | 0.07542 |
| $\rho_{22}$ | 0 | -0.00391 | 0.07103 | 0.00116 | 0.07894 | 0.00047 | 0.07619 | 0.01238 | 0.08824 | 0.01798 | 0.08705 |
| $c_{24}$ | 1 | -0.00168 | 0.02193 | -0.00191 | 0.02265 | -0.00146 | 0.02148 | -0.00153 | 0.02323 | -0.00036 | 0.02207 |
| $c_{25}$ | 1 | 0.00202 | 0.02401 | -0.00112 | 0.02574 | 0.00178 | 0.02385 | -0.00108 | 0.02594 | 0.00186 | 0.02404 |
| $c_{26}$ | 1 | 0.00165 | 0.02223 | 0.00001 | 0.02288 | 0.00163 | 0.02225 | 0.00002 | 0.02297 | 0.00251 | 0.02240 |
| | | $n = 1000$ | | | | | | | | | |
| Eq. 1 | Set III | | | | | | | | | | |
| $b_{21}$ | 0.15 | -0.00019 | 0.00980 | 0.00216 | 0.01014 | 0.00073 | 0.00982 | 0.00270 | 0.01020 | 0.00148 | 0.00983 |
| $\lambda_{11,1}$ | 0 | -0.00137 | 0.01282 | -0.00125 | 0.01434 | -0.00053 | 0.01274 | -0.00243 | 0.01458 | -0.00166 | 0.01262 |
| $\lambda_{11,2}$ | 0 | 0.00117 | 0.01311 | 0.00020 | 0.01429 | 0.00034 | 0.01296 | 0.00049 | 0.01457 | 0.00066 | 0.01320 |
| $\rho_{11}$ | 0 | -0.00084 | 0.04160 | 0.00055 | 0.04401 | -0.00010 | 0.04533 | 0.01129 | 0.04862 | 0.01391 | 0.04871 |
| $\rho_{12}$ | 0 | -0.00404 | 0.05041 | -0.00108 | 0.05518 | -0.00073 | 0.05425 | 0.00526 | 0.05768 | 0.00658 | 0.05572 |
| $c_{11}$ | 1 | -0.00017 | 0.01602 | -0.00130 | 0.01694 | -0.00051 | 0.01604 | -0.00150 | 0.01737 | -0.00018 | 0.01606 |
| $c_{12}$ | 1 | -0.00070 | 0.01680 | -0.00204 | 0.01831 | -0.00084 | 0.01676 | -0.00202 | 0.01841 | -0.00055 | 0.01661 |
| $c_{13}$ | 1 | -0.00007 | 0.01704 | -0.00142 | 0.01850 | -0.00001 | 0.01701 | -0.00155 | 0.01874 | -0.00004 | 0.01669 |
| Eq. 2 | Set III | | | | | | | | | | |
| $b_{12}$ | 0.3 | -0.00006 | 0.00933 | 0.00141 | 0.00931 | 0.00051 | 0.00958 | 0.00216 | 0.00943 | 0.00102 | 0.00936 |
| $\lambda_{22,1}$ | 0 | 0.00036 | 0.01268 | 0.00117 | 0.01364 | 0.00112 | 0.01266 | 0.00000 | 0.01386 | -0.00014 | 0.01250 |
| $\lambda_{22,2}$ | 0 | -0.00027 | 0.01352 | -0.00190 | 0.01480 | -0.00093 | 0.01373 | -0.00146 | 0.01491 | -0.00062 | 0.01395 |
| $\rho_{21}$ | 0 | -0.00587 | 0.04305 | -0.00522 | 0.04489 | -0.00467 | 0.04487 | 0.00870 | 0.04900 | 0.01195 | 0.04819 |
| $\rho_{22}$ | 0 | -0.00412 | 0.04906 | -0.00045 | 0.05481 | -0.00189 | 0.05234 | 0.00245 | 0.05758 | 0.00334 | 0.05741 |
| $c_{24}$ | 1 | 0.00060 | 0.01659 | -0.00079 | 0.01826 | 0.00057 | 0.01656 | -0.00093 | 0.01864 | 0.00076 | 0.01653 |
| $c_{25}$ | 1 | -0.00011 | 0.01692 | -0.00058 | 0.01836 | 0.00026 | 0.01674 | -0.00060 | 0.01811 | 0.00036 | 0.01664 |
| $c_{26}$ | 1 | -0.00019 | 0.01755 | -0.00100 | 0.01838 | -0.00011 | 0.01741 | -0.00057 | 0.01812 | 0.00001 | 0.01753 |

Table 7: Bias and RMSE of Various Estimators for Parameters of Equation 1 and 2, Parameter Set I with Spatial Rook-type Weights Matrices

| | | Rook-type Weights Matrices; $n = 484$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Parameter | True | MLE | | GS2SLS | | GS3SLS | | LQ-GS2SLS | | LQ-GS3SLS | |
| Eq. 1 | Set I | Bias | RMSE | Bias | RMSE | Bias | RMSE | Bias | RMSE | Bias | RMSE |
| $b_{21}$ | 0.15 | 0.00030 | 0.01395 | 0.00435 | 0.01465 | 0.00117 | 0.01405 | 0.00594 | 0.01540 | 0.00283 | 0.01471 |
| $\lambda_{11,1}$ | 0.3 | 0.00011 | 0.01822 | -0.00208 | 0.02199 | -0.00112 | 0.01904 | -0.00171 | 0.02134 | -0.00096 | 0.01863 |
| $\lambda_{11,2}$ | 0.2 | -0.00027 | 0.01813 | -0.00202 | 0.02055 | 0.00016 | 0.01813 | -0.00390 | 0.02091 | -0.00242 | 0.01926 |
| $\rho_{11}$ | 0.2 | -0.00284 | 0.05874 | -0.00119 | 0.06197 | -0.00082 | 0.06418 | 0.00668 | 0.06494 | 0.00552 | 0.06381 |
| $\rho_{12}$ | 0.1 | -0.00731 | 0.05813 | -0.00547 | 0.06625 | -0.00582 | 0.06349 | 0.01667 | 0.07192 | 0.02253 | 0.06910 |
| $c_{11}$ | 1 | 0.00017 | 0.02225 | -0.00171 | 0.02517 | -0.00029 | 0.02214 | -0.00215 | 0.02463 | -0.00042 | 0.02203 |
| $c_{12}$ | 1 | -0.00092 | 0.02288 | -0.00353 | 0.02509 | -0.00174 | 0.02278 | -0.00361 | 0.02606 | -0.00216 | 0.02249 |
| $c_{13}$ | 1 | 0.00039 | 0.02429 | -0.00124 | 0.02531 | -0.00008 | 0.02425 | -0.00181 | 0.02564 | -0.00018 | 0.02391 |
| Eq. 2 | Set I | | | | | | | | | | |
| $b_{12}$ | 0.3 | 0.00042 | 0.01300 | 0.00345 | 0.01401 | 0.00044 | 0.01340 | 0.00460 | 0.01420 | 0.00151 | 0.01346 |
| $\lambda_{22,1}$ | 0.3 | -0.00097 | 0.01852 | -0.00265 | 0.01951 | -0.00244 | 0.01884 | -0.00188 | 0.01960 | -0.00111 | 0.01854 |
| $\lambda_{22,2}$ | 0.15 | 0.00026 | 0.01878 | 0.00029 | 0.02081 | 0.00137 | 0.01903 | -0.00228 | 0.02195 | -0.00142 | 0.01938 |
| $\rho_{21}$ | 0.1 | 0.00073 | 0.06013 | 0.00307 | 0.06813 | 0.00118 | 0.06898 | 0.00316 | 0.07124 | 0.00520 | 0.07404 |
| $\rho_{22}$ | 0 | -0.00393 | 0.05632 | -0.00543 | 0.06503 | -0.00569 | 0.06153 | 0.01548 | 0.07350 | 0.01855 | 0.07269 |
| $c_{24}$ | 1 | -0.00136 | 0.02601 | -0.00336 | 0.02745 | -0.00211 | 0.02571 | -0.00275 | 0.02854 | -0.00005 | 0.02615 |
| $c_{25}$ | 1 | 0.00135 | 0.02378 | 0.00018 | 0.02529 | 0.00165 | 0.02374 | 0.00107 | 0.02505 | 0.00284 | 0.02382 |
| $c_{26}$ | 1 | 0.00064 | 0.02272 | -0.00100 | 0.02444 | 0.00089 | 0.02278 | -0.00088 | 0.02485 | 0.00135 | 0.02316 |
| | | $n = 961$ | | | | | | | | | |
| Eq. 1 | Set I | | | | | | | | | | |
| $b_{21}$ | 0.15 | -0.00020 | 0.00994 | 0.00213 | 0.00997 | 0.00057 | 0.00985 | 0.00291 | 0.01039 | 0.00137 | 0.01003 |
| $\lambda_{11,1}$ | 0.3 | 0.00011 | 0.01352 | -0.00205 | 0.01516 | -0.00111 | 0.01333 | -0.00200 | 0.01520 | -0.00122 | 0.01369 |
| $\lambda_{11,2}$ | 0.2 | -0.00049 | 0.01325 | 0.00018 | 0.01531 | 0.00015 | 0.01369 | -0.00088 | 0.01560 | -0.00093 | 0.01364 |
| $\rho_{11}$ | 0.2 | -0.00232 | 0.04129 | 0.00099 | 0.04607 | 0.00009 | 0.04424 | 0.00234 | 0.04798 | 0.00135 | 0.04576 |
| $\rho_{12}$ | 0.1 | -0.00231 | 0.04338 | -0.00043 | 0.04829 | -0.00213 | 0.04669 | 0.01064 | 0.05075 | 0.01222 | 0.04986 |
| $c_{11}$ | 1 | 0.00047 | 0.01766 | -0.00118 | 0.02092 | -0.00007 | 0.01789 | -0.00133 | 0.02058 | 0.00003 | 0.01759 |
| $c_{12}$ | 1 | -0.00015 | 0.01756 | -0.00107 | 0.01857 | -0.00093 | 0.01723 | -0.00121 | 0.01807 | -0.00072 | 0.01741 |
| $c_{13}$ | 1 | 0.00018 | 0.01677 | -0.00100 | 0.01879 | -0.00053 | 0.01696 | -0.00112 | 0.01907 | -0.00025 | 0.01729 |
| Eq. 2 | Set I | | | | | | | | | | |
| $b_{12}$ | 0.3 | -0.00005 | 0.00988 | 0.00176 | 0.00988 | 0.00010 | 0.00984 | 0.00223 | 0.00991 | 0.00048 | 0.00970 |
| $\lambda_{22,1}$ | 0.3 | 0.00083 | 0.01260 | -0.00072 | 0.01500 | 0.00027 | 0.01288 | -0.00035 | 0.01545 | 0.00047 | 0.01291 |
| $\lambda_{22,2}$ | 0.15 | -0.00085 | 0.01385 | -0.00089 | 0.01509 | -0.00003 | 0.01399 | -0.00129 | 0.01495 | -0.00133 | 0.01355 |
| $\rho_{21}$ | 0.1 | -0.00295 | 0.04679 | -0.00214 | 0.05208 | -0.00179 | 0.05039 | 0.00034 | 0.05322 | 0.00105 | 0.05215 |
| $\rho_{22}$ | 0 | -0.00502 | 0.04425 | -0.00498 | 0.04669 | -0.00673 | 0.04552 | 0.00361 | 0.04921 | 0.00553 | 0.04831 |
| $c_{24}$ | 1 | -0.00025 | 0.01627 | -0.00113 | 0.01750 | -0.00014 | 0.01633 | -0.00111 | 0.01757 | 0.00046 | 0.01627 |
| $c_{25}$ | 1 | 0.00193 | 0.01854 | 0.00104 | 0.02031 | 0.00199 | 0.01834 | 0.00108 | 0.02028 | 0.00236 | 0.01872 |
| $c_{26}$ | 1 | -0.00002 | 0.01765 | -0.00024 | 0.01862 | 0.00012 | 0.01764 | -0.00041 | 0.01872 | 0.00056 | 0.01791 |

Table 8: Bias and RMSE of Various Estimators for Parameters of Equation 1 and 2, Parameter Set II with Spatial Rook-type Weights Matrices

| | | Rook-type Weights Matrices; $n = 484$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Parameter | True | MLE | | GS2SLS | | GS3SLS | | LQ-GS2SLS | | LQ-GS3SLS | |
| Eq. 1 | Set II | Bias | RMSE | Bias | RMSE | Bias | RMSE | Bias | RMSE | Bias | RMSE |
| $b_{21}$ | 0.15 | -0.00006 | 0.01373 | 0.00475 | 0.01461 | 0.00228 | 0.01399 | 0.00584 | 0.01461 | 0.00308 | 0.01421 |
| $\lambda_{11,1}$ | -0.3 | 0.00045 | 0.01847 | -0.00196 | 0.02171 | -0.00230 | 0.01901 | 0.00049 | 0.02111 | -0.00015 | 0.01959 |
| $\lambda_{11,2}$ | -0.2 | -0.00055 | 0.01779 | -0.00040 | 0.02000 | -0.00030 | 0.01798 | -0.00360 | 0.02016 | -0.00253 | 0.01860 |
| $\rho_{11}$ | -0.2 | -0.00300 | 0.06501 | 0.00157 | 0.06740 | 0.00245 | 0.06735 | -0.00580 | 0.07495 | -0.00761 | 0.07299 |
| $\rho_{12}$ | -0.1 | -0.00452 | 0.05882 | -0.00295 | 0.06462 | -0.00408 | 0.06426 | 0.01089 | 0.07022 | 0.01641 | 0.06848 |
| $c_{11}$ | 1 | 0.00038 | 0.02213 | -0.00269 | 0.02446 | -0.00012 | 0.02218 | -0.00282 | 0.02527 | 0.00036 | 0.02140 |
| $c_{12}$ | 1 | -0.00033 | 0.02182 | -0.00202 | 0.02491 | -0.00088 | 0.02176 | -0.00321 | 0.02532 | -0.00127 | 0.02231 |
| $c_{13}$ | 1 | -0.00001 | 0.02382 | -0.00076 | 0.02489 | -0.00008 | 0.02342 | -0.00137 | 0.02492 | 0.00038 | 0.02388 |
| Eq. 2 | Set II | | | | | | | | | | |
| $b_{12}$ | 0.3 | 0.00038 | 0.01337 | 0.00382 | 0.01357 | 0.00132 | 0.01304 | 0.00486 | 0.01361 | 0.00237 | 0.01324 |
| $\lambda_{22,1}$ | -0.3 | -0.00028 | 0.01615 | -0.00384 | 0.01746 | -0.00352 | 0.01687 | -0.00144 | 0.01812 | -0.00132 | 0.01735 |
| $\lambda_{22,2}$ | -0.15 | 0.00040 | 0.01871 | 0.00151 | 0.02009 | 0.00093 | 0.01886 | -0.00270 | 0.02092 | -0.00204 | 0.01890 |
| $\rho_{21}$ | -0.1 | 0.00231 | 0.05824 | 0.00559 | 0.06581 | 0.00496 | 0.06723 | -0.00082 | 0.07060 | -0.00308 | 0.06936 |
| $\rho_{22}$ | 0 | -0.00225 | 0.05764 | -0.00465 | 0.06666 | -0.00518 | 0.06369 | 0.01972 | 0.07576 | 0.02030 | 0.07460 |
| $c_{24}$ | 1 | -0.00167 | 0.02478 | -0.00270 | 0.02770 | -0.00118 | 0.02492 | -0.00279 | 0.02755 | -0.00089 | 0.02507 |
| $c_{25}$ | 1 | 0.00167 | 0.02311 | 0.00043 | 0.02466 | 0.00247 | 0.02307 | 0.00132 | 0.02549 | 0.00372 | 0.02373 |
| $c_{26}$ | 1 | 0.00043 | 0.02266 | -0.00114 | 0.02475 | 0.00133 | 0.02278 | -0.00050 | 0.02506 | 0.00165 | 0.02258 |
| | | $n = 961$ | | | | | | | | | |
| Eq. 1 | Set II | | | | | | | | | | |
| $b_{21}$ | 0.15 | -0.00011 | 0.00966 | 0.00263 | 0.01011 | 0.00133 | 0.00982 | 0.00290 | 0.01024 | 0.00169 | 0.01006 |
| $\lambda_{11,1}$ | -0.3 | -0.00040 | 0.01361 | -0.00183 | 0.01531 | -0.00231 | 0.01383 | -0.00110 | 0.01557 | -0.00106 | 0.01378 |
| $\lambda_{11,2}$ | -0.2 | 0.00017 | 0.01322 | 0.00040 | 0.01476 | 0.00078 | 0.01342 | -0.00052 | 0.01499 | -0.00063 | 0.01366 |
| $\rho_{11}$ | -0.2 | -0.00160 | 0.04336 | 0.00207 | 0.04970 | 0.00114 | 0.04733 | -0.00482 | 0.05256 | -0.00571 | 0.05156 |
| $\rho_{12}$ | -0.1 | -0.00155 | 0.04317 | -0.00096 | 0.04821 | -0.00084 | 0.04759 | 0.00781 | 0.05042 | 0.00993 | 0.05103 |
| $c_{11}$ | 1 | 0.00062 | 0.01648 | -0.00116 | 0.01963 | 0.00056 | 0.01658 | -0.00114 | 0.01942 | 0.00073 | 0.01659 |
| $c_{12}$ | 1 | -0.00061 | 0.01827 | -0.00117 | 0.01840 | -0.00084 | 0.01825 | -0.00165 | 0.01857 | -0.00110 | 0.01856 |
| $c_{13}$ | 1 | -0.00018 | 0.01717 | -0.00169 | 0.01856 | -0.00018 | 0.01727 | -0.00194 | 0.01857 | -0.00053 | 0.01708 |
| Eq. 2 | Set II | | | | | | | | | | |
| $b_{12}$ | 0.3 | -0.00007 | 0.01019 | 0.00205 | 0.01025 | 0.00066 | 0.00994 | 0.00230 | 0.01002 | 0.00109 | 0.00978 |
| $\lambda_{22,1}$ | -0.3 | 0.00016 | 0.01393 | -0.00141 | 0.01486 | -0.00138 | 0.01383 | -0.00070 | 0.01518 | -0.00085 | 0.01406 |
| $\lambda_{22,2}$ | -0.15 | -0.00072 | 0.01295 | 0.00010 | 0.01378 | 0.00011 | 0.01288 | -0.00118 | 0.01409 | -0.00128 | 0.01283 |
| $\rho_{21}$ | -0.1 | -0.00120 | 0.04742 | 0.00031 | 0.05182 | -0.00029 | 0.04901 | -0.00144 | 0.05227 | -0.00320 | 0.05098 |
| $\rho_{22}$ | 0 | -0.00526 | 0.04545 | -0.00502 | 0.04915 | -0.00615 | 0.04796 | 0.00688 | 0.05104 | 0.00824 | 0.05135 |
| $c_{24}$ | 1 | -0.00038 | 0.01584 | -0.00090 | 0.01691 | 0.00004 | 0.01600 | -0.00119 | 0.01723 | 0.00035 | 0.01593 |
| $c_{25}$ | 1 | 0.00232 | 0.01822 | 0.00116 | 0.01978 | 0.00255 | 0.01816 | 0.00112 | 0.02001 | 0.00274 | 0.01842 |
| $c_{26}$ | 1 | 0.00011 | 0.01678 | -0.00064 | 0.01853 | 0.00034 | 0.01653 | -0.00092 | 0.01872 | 0.00069 | 0.01707 |

Table 9: Bias and RMSE of Various Estimators for Parameters of Equation 1 and 2, Parameter Set III with Spatial Rook-type Weights Matrices

| Parameter | True | MLE | | GS2SLS | | GS3SLS | | LQ-GS2SLS | | LQ-GS3SLS | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Rook-type Weights Matrices; $n = 484$ | | | | | | | | | |
| Eq. 1 | Set III | Bias | RMSE | Bias | RMSE | Bias | RMSE | Bias | RMSE | Bias | RMSE |
| $b_{21}$ | 0.15 | 0.00006 | 0.01409 | 0.00481 | 0.01473 | 0.00164 | 0.01413 | 0.00604 | 0.01520 | 0.00287 | 0.01447 |
| $\lambda_{11,1}$ | 0 | 0.00049 | 0.01856 | -0.00249 | 0.02100 | -0.00185 | 0.01905 | -0.00088 | 0.02125 | -0.00081 | 0.01923 |
| $\lambda_{11,2}$ | 0 | -0.00052 | 0.01833 | -0.00103 | 0.02026 | -0.00008 | 0.01808 | -0.00350 | 0.02026 | -0.00270 | 0.01909 |
| $\rho_{11}$ | 0 | -0.00328 | 0.06282 | 0.00100 | 0.06496 | -0.00031 | 0.06585 | 0.00211 | 0.06942 | 0.00081 | 0.06853 |
| $\rho_{12}$ | 0 | -0.00586 | 0.05788 | -0.00460 | 0.06559 | -0.00365 | 0.06381 | 0.01654 | 0.07267 | 0.02004 | 0.07024 |
| $c_{11}$ | 1 | 0.00052 | 0.02259 | -0.00258 | 0.02475 | -0.00010 | 0.02211 | -0.00284 | 0.02491 | -0.00009 | 0.02209 |
| $c_{12}$ | 1 | -0.00062 | 0.02244 | -0.00276 | 0.02536 | -0.00166 | 0.02224 | -0.00340 | 0.02553 | -0.00159 | 0.02303 |
| $c_{13}$ | 1 | 0.00020 | 0.02385 | -0.00107 | 0.02594 | -0.00019 | 0.02377 | -0.00170 | 0.02576 | 0.00023 | 0.02408 |
| Eq. 2 | Set III | | | | | | | | | | |
| $b_{12}$ | 0.3 | 0.00042 | 0.01381 | 0.00390 | 0.01438 | 0.00114 | 0.01395 | 0.00500 | 0.01457 | 0.00209 | 0.01349 |
| $\lambda_{22,1}$ | 0 | -0.00073 | 0.01763 | -0.00285 | 0.01822 | -0.00266 | 0.01756 | -0.00146 | 0.01870 | -0.00087 | 0.01763 |
| $\lambda_{22,2}$ | 0 | 0.00010 | 0.01939 | 0.00088 | 0.02074 | 0.00149 | 0.01963 | -0.00256 | 0.02204 | -0.00164 | 0.01976 |
| $\rho_{21}$ | 0 | 0.00189 | 0.05972 | 0.00508 | 0.06761 | 0.00385 | 0.06899 | 0.00152 | 0.07181 | 0.00129 | 0.07172 |
| $\rho_{22}$ | 0 | -0.00346 | 0.05684 | -0.00583 | 0.06572 | -0.00563 | 0.06269 | 0.01748 | 0.07473 | 0.01955 | 0.07454 |
| $c_{24}$ | 1 | -0.00182 | 0.02561 | -0.00292 | 0.02810 | -0.00190 | 0.02554 | -0.00312 | 0.02827 | -0.00044 | 0.02600 |
| $c_{25}$ | 1 | 0.00155 | 0.02290 | 0.00019 | 0.02515 | 0.00217 | 0.02307 | 0.00117 | 0.02570 | 0.00333 | 0.02380 |
| $c_{26}$ | 1 | 0.00077 | 0.02280 | -0.00111 | 0.02489 | 0.00104 | 0.02282 | -0.00070 | 0.02444 | 0.00169 | 0.02290 |
| | | $n = 961$ | | | | | | | | | |
| Eq. 1 | Set III | | | | | | | | | | |
| $b_{21}$ | 0.15 | -0.00014 | 0.00983 | 0.00254 | 0.01020 | 0.00094 | 0.00975 | 0.00298 | 0.01036 | 0.00156 | 0.01015 |
| $\lambda_{11,1}$ | 0 | 0.00002 | 0.01279 | -0.00200 | 0.01454 | -0.00135 | 0.01310 | -0.00156 | 0.01482 | -0.00097 | 0.01308 |
| $\lambda_{11,2}$ | 0 | -0.00005 | 0.01349 | 0.00046 | 0.01501 | 0.00091 | 0.01360 | -0.00090 | 0.01537 | -0.00037 | 0.01352 |
| $\rho_{11}$ | 0 | -0.00247 | 0.04337 | 0.00111 | 0.04784 | 0.00067 | 0.04729 | -0.00025 | 0.05003 | -0.00013 | 0.04842 |
| $\rho_{12}$ | 0 | -0.00140 | 0.04384 | -0.00116 | 0.04852 | -0.00188 | 0.04792 | 0.00981 | 0.05217 | 0.01188 | 0.05151 |
| $c_{11}$ | 1 | 0.00044 | 0.01753 | -0.00116 | 0.02016 | 0.00018 | 0.01746 | -0.00118 | 0.02024 | 0.00041 | 0.01719 |
| $c_{12}$ | 1 | -0.00037 | 0.01807 | -0.00074 | 0.01888 | -0.00071 | 0.01797 | -0.00151 | 0.01862 | -0.00108 | 0.01811 |
| $c_{13}$ | 1 | 0.00015 | 0.01709 | -0.00125 | 0.01883 | -0.00018 | 0.01698 | -0.00128 | 0.01906 | -0.00040 | 0.01712 |
| Eq. 2 | Set III | | | | | | | | | | |
| $b_{12}$ | 0.3 | 0.00017 | 0.01017 | 0.00186 | 0.01019 | 0.00059 | 0.00994 | 0.00238 | 0.01023 | 0.00086 | 0.00994 |
| $\lambda_{22,1}$ | 0 | 0.00010 | 0.01292 | -0.00102 | 0.01478 | -0.00069 | 0.01295 | -0.00045 | 0.01513 | -0.00065 | 0.01288 |
| $\lambda_{22,2}$ | 0 | -0.00076 | 0.01333 | -0.00014 | 0.01479 | 0.00005 | 0.01361 | -0.00137 | 0.01452 | -0.00120 | 0.01347 |
| $\rho_{21}$ | 0 | -0.00222 | 0.04786 | -0.00085 | 0.05181 | -0.00219 | 0.04922 | -0.00087 | 0.05351 | -0.00023 | 0.05171 |
| $\rho_{22}$ | 0 | -0.00493 | 0.04463 | -0.00465 | 0.04706 | -0.00625 | 0.04630 | 0.00571 | 0.04953 | 0.00730 | 0.05053 |
| $c_{24}$ | 1 | -0.00011 | 0.01627 | -0.00134 | 0.01729 | 0.00019 | 0.01625 | -0.00125 | 0.01746 | 0.00067 | 0.01606 |
| $c_{25}$ | 1 | 0.00231 | 0.01865 | 0.00101 | 0.02021 | 0.00271 | 0.01862 | 0.00107 | 0.02007 | 0.00250 | 0.01910 |
| $c_{26}$ | 1 | 0.00030 | 0.01714 | -0.00041 | 0.01869 | 0.00029 | 0.01736 | -0.00069 | 0.01880 | 0.00062 | 0.01717 |

## G.3 Monte Carlo Simulations for the Weakly Identified Case

In Appendix F included in this Online Supplementary Appendix we discussed three scenarios where identification from the linear moment conditions fails. In this case the two step GS2SLS and GS3SLS estimators, which rely on the linear moment conditions, will be inconsistent. In the following we illustratively report on the performance of ML, and one-step and two-step estimators for scenarios that are "close" to the non-identified scenarios. We refer to those experiments as Weak Identification Scenario 1, 2 and 3.

■ Weak Identification Scenario 1

To explore this scenario we consider a simplified variant of the first equation of the model discussed in Appendix G.1, based on parameter Set I, but with $b_{21} = \lambda_{11,2} = \rho_{11} = \rho_{12} = 0$. To explore the effect of the $X$ instruments being weak we consider a case where the parameters on the exogenous variables in equation 1 are $c_{1k} = 0.0001$, $k = 1, 2, 3$. The simulations employ the social interaction weights matrix $\mathbf{M}_1$ with sample size $n = 500$. The results reported in Table 10 at the end of this subsection show that in this case the GS2SLS estimator for autoregressive parameter $\lambda_{11,1}$ is substantially biased, in contrast to ML and LQ-GS2SLS.

■ Weak Identification for Scenario 2

To explore the effect when the weights matrix is close to the case where all weights are equal (except for the zero diagonal elements) we consider a simplified variant of the first equation of the model discussed in Appendix G.1, based on parameter Set I, but with $b_{21} = \lambda_{11,2} = \rho_{11} = \rho_{12} = 0$. For this simulation the social interaction weights matrix $\mathbf{M}_1$ with sample size $n = 500$ is generated as in Appendix G.1, however with $d_* = 1.5$ (instead of $d_* = .3$). This leaves most of the 500 considered individuals with more than 400 friends in the simulations presented in Table 11 at the end of this subsection. As expected, all estimators show bias for the autoregressive parameter $\lambda_{11,1}$. The absolute bias for the ML is close to 0.1. All other estimators exhibit even larger biases.

■ Weak Identification for Scenario 3

To explore this scenario we consider the single equation case with the data

generated from the following model:

$$\mathbf{y}_1 = \lambda_{11,1}\mathbf{M}_1\mathbf{y}_1 + \sum_{k=1}^{3} c_{k1}\mathbf{x}_k + \sum_{k=1}^{3} c_{k+3,1}\mathbf{M}_1\mathbf{x}_k + \mathbf{u}_1,$$
$$\mathbf{u}_1 = \rho_{11}\mathbf{M}_1\mathbf{u}_1 + \boldsymbol{\epsilon}_1.$$

The matrix $\mathbf{M}_1$, the exogenous covariates $\mathbf{x}_k$ and corresponding parameters $c_{k1}$, $k = 1, 2, 3$, and the innovations $\boldsymbol{\epsilon}_1$ are as described in Appendix G.1. For the case where $c_{k+3,1} = -\lambda_{11,1}c_{k1}$ and $\rho_{11} = -\lambda_{11,1}$ the above model is approximately equal to the following model without network interdependencies:

$$\mathbf{y}_1 = \sum_{k=1}^{3} c_{k1}\mathbf{x}_k + \boldsymbol{\epsilon}_1.$$

This is readily seen upon pre-multiplying the last equation with $\mathbf{I}-\lambda_{11,1}\mathbf{M}_1$ and exploiting the approximation $\mathbf{u}_1 = (\mathbf{I} - \rho_{11}\mathbf{M}_1)^{-1}\boldsymbol{\epsilon}_1 \doteq \boldsymbol{\epsilon}_1 + \rho_{11}\mathbf{M}_1\boldsymbol{\epsilon}_1$ for the disturbance process. Now let $c_{k+3,1} = -(\lambda_{11,1} + \nu)c_{k1}$ and $\rho_1 = -(\lambda_{11,1} + \nu)$, then we expect the parameters to only be weakly identified for $\nu$ small. (Of course, proximity to the non-identified set of parameters can be modeled in various ways. We adopted this approach for its simplicity.) In Table 12 at the end of this subsection we report on the small sample behavior of various estimators for $\lambda_{11,1} = .5$ and $\nu = .3$. The results in Table 12 indicate that for $v = .3$ the ML estimator, and to a lesser degree the LQ-GS2SLS estimator, are still able to estimate the parameters $\lambda_{11,1}$, $\rho_{11}$, $c_{41}$, $c_{51}$, $c_{61}$ reasonably well, while GS2SLS is already severely biased.

Table 10: Bias and RMSE under Scenario 1 of Various Estimators for Parameters of a Simplified Version of Equation 1. Parameter Set I, except that $b_{21} = \lambda_{11,2} = \rho_{11} = \rho_{12} = 0$ and $c_{1k} = 0.0001$ for all $k = \{1, 2, 3\}$.

| Parameter | | Soc. Interact. Matrix; $n = 500$ | | | | | |
| | | MLE | | GS2SLS | | LQ-GS2SLS | |
| One equ. | True | Bias | RMSE | Bias | RMSE | Bias | RMSE |
| --- | --- | --- | --- | --- | --- | --- | --- |
| $\lambda_{11,1}$ | 0.30 | -0.00134 | 0.04676 | 0.17807 | 0.50151 | 0.00179 | 0.04714 |
| $c_{11}$ | 0.0001 | 0.00013 | 0.02149 | 0.00182 | 0.02158 | 0.00034 | 0.02149 |
| $c_{12}$ | 0.0001 | -0.00050 | 0.02439 | -0.00090 | 0.02529 | -0.00064 | 0.02449 |
| $c_{13}$ | 0.0001 | 0.00015 | 0.02417 | 0.00058 | 0.02314 | 0.00049 | 0.02410 |

Table 11: Bias and RMSE under Scenario 2 of Various Estimators for Parameters of a Simplified Version of Equation 1. Parameter Set I, except that $b_{21} = \lambda_{11,2} = \rho_{11} = \rho_{12} = 0$. Social Interaction Weight Matrix with High Number of Friends

| Parameter | | Soc. Interact. Weights Matrices; $n = 500$ | | | | | | | | | |
| | | MLE | | GS2SLS | | GS3SLS | | LQ-GS2SLS | | LQ-GS3SLS | |
| Eq. 1 | True | Bias | RMSE | Bias | RMSE | Bias | RMSE | Bias | RMSE | Bias | RMSE |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| $b_{21}$ | 0.15 | -0.00009 | 0.01357 | 0.00273 | 0.01387 | 0.00143 | 0.01383 | 0.00316 | 0.01389 | 0.00198 | 0.01378 |
| $\lambda_{11,1}$ | 0.30 | -0.09312 | 0.44336 | 0.20473 | 0.36126 | 0.20307 | 0.36233 | 0.23556 | 0.43856 | 0.22588 | 0.47735 |
| $c_{11}$ | 1 | 0.00037 | 0.02424 | -0.00013 | 0.02642 | 0.00076 | 0.02461 | -0.00033 | 0.02644 | 0.00084 | 0.02434 |
| $c_{12}$ | 1 | -0.00113 | 0.02546 | -0.00142 | 0.02556 | -0.00064 | 0.02534 | -0.00125 | 0.02591 | -0.00004 | 0.02560 |
| $c_{13}$ | 1 | -0.00071 | 0.02450 | -0.00155 | 0.02584 | 0.00001 | 0.02476 | -0.00189 | 0.02559 | -0.00003 | 0.02482 |

Table 12: Bias and RMSE under Scenario 3 of Various Estimators in Neighborhood of Parameter Singularity (where Order of Spillovers is Overspecified)

| Parameter | | Soc. Interact. Matrix; $n = 500$ | | | | | |
| | | MLE | | GS2SLS | | LQ-GS2SLS | |
| One equ. | True | Bias | RMSE | Bias | RMSE | Bias | RMSE |
| --- | --- | --- | --- | --- | --- | --- | --- |
| $\lambda_{11,1}$ | 0.3 | -0.03320 | 0.17906 | -0.34951 | 0.69577 | 0.06243 | 0.20602 |
| $\rho_{11}$ | -0.3 | 0.03192 | 0.21385 | 0.36910 | 0.55993 | -0.08737 | 0.27132 |
| $c_{11}$ | 1 | -0.00103 | 0.03952 | -0.00992 | 0.04777 | 0.00161 | 0.03982 |
| $c_{12}$ | 1 | -0.00511 | 0.04634 | -0.01437 | 0.05319 | -0.00274 | 0.04586 |
| $c_{13}$ | 1 | -0.00280 | 0.04570 | -0.01507 | 0.05369 | 0.00051 | 0.04810 |
| $c_{14}$ | -0.4 | 0.03411 | 0.18339 | 0.30034 | 0.62913 | -0.04263 | 0.18271 |
| $c_{15}$ | -0.4 | 0.04620 | 0.18502 | 0.29730 | 0.64222 | -0.04365 | 0.18606 |
| $c_{16}$ | -0.4 | 0.02592 | 0.16655 | 0.28633 | 0.63019 | -0.05244 | 0.19939 |

# References

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

[2] Kelejian, H.H., and I.R. Prucha (2002) 2SLS and OLS in a spatial autoregressive model with equal spatial weights. *Regional Science and Urban Economics* 32, 691-707.

[3] Kelejian, H.H., I.R. Prucha, and Y. Yuzefovich (2006) Estimation problems in models with spatial weighting matrices which have blocks of equal elements. *Journal of Regional Science* 46, 507-517.

[4] Paula, Á. (2017) Econometrics of network models. In B. Honoré, A. Pakes, M. Piazzesi, & L. Samuelson (eds.), *Advances in Economics and Econometrics: Eleventh World Congress (Econometric Society Monographs*, pp. 268-323). Cambridge University Press.