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

Ingmar R. Prucha\(^1\), David M. Drukker\(^2\), and Peter H. Egger\(^3\)

December 26, 2016

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

\(^2\)StataCorp, 4905 Lakeway Drive, College Station, TX 77845, Tel.: +1-979-696-4600, e-mail: ddrukker@stata.com

\(^3\)KOF, ETH Zurich, Leonhardstrasse 21, 8092 Zurich, Switzerland; Tel.: +41-44-632-4108, e-mail: egger@kof.ethz.ch
Abstract

This paper develops an estimation methodology for network data generated from a system of simultaneous equations. Our specification 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 on the generalized method of moments (GMM) estimation approach introduced in Kelejian and Prucha (1998, 1999, 2004). The paper considers limited and full information estimators, and one- and two-step estimators, and establishes their asymptotic properties. In contrast to some of the earlier literature our asymptotic results facilitate joint tests for the absence of all forms of network spillovers.

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

JEL Classification: C21, C31
1 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 indexed by location, the developed methodology can be of interest to the estimation of a much wider class of networks, including social networks.

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

An important class of models for spatial networks originates from the seminal work by Whittle (1954) and Cliff and Ord (1971, 1983). 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.

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

---

1 We gratefully acknowledge financial support from the National Institute of Health through SBIR grants R43 AG027622 and R44 AG027622. We also thank the the CESifo for their hospitality.


3 See, e.g., Anselin (2010) for a review of the development of spatial econometric methods.
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 so far been limited. 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 estimators for such models. However, as discussed in more detail below, they do not provide a full asymptotic theory for all considered estimators and their setup only covers first-order spatial lags. Cohme-Cole, Liu, and Zenou (2014) and Liu (2014) employ the methodology of Kelejian and Prucha (2004) within the context of social interaction models, and provide further refinements.

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 (2015) consider the quasi maximum likelihood estimator for a multi-equation system with a first-order spatial lag in the dependent variable. In contrast to the current paper, those papers only consider first-order spatial lags in the dependent variable, and do not also consider spatial spillovers in the disturbance process. Those papers also differ in terms of the considered estimation methodology.

An important limitation of the estimation methodology developed in Kelejian and Prucha (2004) is that the paper only establishes the consistency, but not the asymptotic normality, of important spatial parameters. 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., though 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 relax the requirement regarding a priori knowledge of what weights should be assigned to different units in the construction of a spatial lag.

The paper is organized as follows: Section 2 specifies the considered simultaneous equation system with spatial/cross sectional network interactions. In Section 3 we 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 4, 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 5 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 Sections 6 and 7 we introduce specific limited and full information estimators, and provide specific expressions for consistent estimators of the associated asymptotic variance-covariance (VC) matrices of those estimators. The last section concludes with a summary of our findings and possible directions for future research. All technical derivations are given in appendices.

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

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

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

2 Model

In the following, we specify our simultaneous system of $G$ equations for $G$ endogenous variables observed for $n$ cross sectional units. This system allows 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 textbook simultaneous equation system. Second, simultaneity may stem from Cliff and Ord (1973, 1981) type higher-order cross-sectional network interactions, where spatial interactions represent a leading application.

---

4Our specification differs from, e.g., Anselin (1988) and Wang, Lee and Bao (2015) who consider systems for one variable.
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, we assume that the cross-sectional data are generated by the following system ($g = 1, \ldots, G$):

\[ y_{g,n} = \sum_{l=1}^{G} b_{l,g,n} y_{l,n} + \sum_{k=1}^{K} c_{k,g,n} x_{k,n} + \sum_{l=1}^{G} \left( \sum_{s=1}^{p} \lambda_{l|g,s,n} W_{s,n} \right) y_{l,n} + u_{g,n}, \quad (1) \]

\[ u_{g,n} = \left[ \sum_{r=1}^{q} \rho_{g,r,n} M_{r,n} \right] u_{g,n} + \varepsilon_{g,n}, \]

where $y_{g,n}$ is the $n \times 1$ vector of cross-sectional observations on the dependent variable in the $g$-th equation, $x_{k,n}$ is the $n \times 1$ vector of cross-sectional observations on the $k$-th exogenous variable, which is taken to be nonstochastic, $u_{g,n}$ is the $n \times 1$ disturbance vector in the $g$-th equation, $W_{s,n}$ and $M_{r,n}$ are $n \times n$ weights matrices, $\varepsilon_{g,n}$ is the $n \times 1$ vector of innovations entering the disturbance process for the $g$-th equation, and $n$ denotes the sample size. With $b_{l,g,n}$ and $c_{k,g,n}$ we denote the (scalar) parameters corresponding to the $l$-th endogenous and $k$-th exogenous variables, respectively.

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

\footnote{Of course, the structural model parameters are not identified without certain restrictions. Those restrictions will be introduced below.}

\footnote{In treating the exogenous variables as non-stochastic, the analysis may be viewed as conditional on the exogenous variables.}
normalizations of these matrices where the normalization factor(s) depend on the sample size.\(^7\) The \(n\)-th element of \(y_{l,s,n}\) is given by \(\psi_{l,s,n} = \sum_{j=1}^{n} w_{ij,s}y_{j,n}\). We note that even if the elements of the spatial weights matrices do not depend on sample size, the elements of the spatial lag \(y_{l,s,n}\) and, analogously, the elements of \(u_{g,r,n}\) will generally depend on sample size. This in turn implies that also the elements of \(y_{g,n}\) and \(u_{g,n}\) will generally depend on sample size. Form triangular arrays. In allowing the elements of \(x_{k,n}\) to depend on the sample size, we implicitly also allow for some of the exogenous variables to be spatial lags of exogenous variables, and 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 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

\[
\begin{align*}
Y_n &= Y_nB_n + X_nC_n + Y_n\Lambda_n + U_n, \\
U_n &= U_nR_n + E_n
\end{align*}
\]

with

\[
\begin{align*}
Y_n &= (y_{1,n}, \ldots, y_{G,n})_{n \times G}, \\
X_n &= (x_{1,n}, \ldots, x_{K,n})_{n \times K}, \\
U_n &= (u_{1,n}, \ldots, u_{G,n})_{n \times G}, \\
E_n &= (\varepsilon_{1,n}, \ldots, \varepsilon_{G,n})_{n \times G}, \\
Y_p &= (y_{1,1,n}, \ldots, y_{1,p,n}, \ldots, y_{G,1,n}, \ldots, y_{G,p,n})_{n \times pG}, \\
U_p &= (u_{1,1,n}, \ldots, u_{1,q,n}, \ldots, u_{G,1,n}, \ldots, u_{G,q,n})_{n \times qG},
\end{align*}
\]

and where the parameter matrices \(B_n = (b_{g,n})_{G \times G}\), \(C_n = (c_{g,n})_{K \times G}\), \(\Lambda_n = \)...

\(^7\)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.

\(^9\)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).
(λ_{g,s,n})_{pG×G}, and R_n = (ρ_{g,r,n})_{qG×G} are defined conformably.\textsuperscript{10}

Towards computing the reduced form of the above model, let

\[ y_n = \text{vec}(Y_n), \quad x_n = \text{vec}(X_n), \quad u_n = \text{vec}(U_n), \quad ε_n = \text{vec}(E_n), \]

and let

\[ W_n = [W_{1,n}', \ldots, W_{p,n}']', \quad M_n = [M_{1,n}', \ldots, M_{q,n}']'. \]

Observing that \( \text{vec}(Y_n) = \text{vec}(U_n) \) and \( \text{vec}(U_n) = \text{vec}(E_n) \), and that for any two conformable matrices \( A_1 \) and \( A_2 \), \( \text{vec}(A_1 A_2) = (\text{vec}(A_1)) \text{vec}(A_2) \) it is readily seen that the spatial simultaneous equation system \( (2) \) can be re-written in stacked notation as

\[ \begin{align*}
  y_n &= B_n^* y_n + C_n^* x_n + u_n, \\
  u_n &= R_n^* u_n + ε_n,
\end{align*} \tag{3} \]

where \( B_n^* = [(B_n' \otimes I_n) + (A_n' \otimes I_n) (I_G \otimes W_n)], \quad C_n^* = (C_n' \otimes I_n), \quad \text{and} \quad R_n^* = (R_n' \otimes I_n) (I_G \otimes M_n). \) Assuming invertability of \( I_{nG} - B_n^* \) and \( I_{nG} - R_n^* \), the reduced form of the system is now given by

\[ \begin{align*}
  y_n &= (I_{nG} - B_n^*)^{-1} [C_n^* x_n + u_n], \\
  u_n &= (I_{nG} - R_n^*)^{-1} ε_n.
\end{align*} \tag{4} \]

As remarked, the structural parameters of the spatial simultaneous equation system \( (1) \) and \( (2) \) are not identified unless we impose exclusion restrictions. Let \( β_{g,n}, γ_{g,n}, λ_{g,n}, \) and \( ρ_{g,n} \) denote the \( G_g \times 1, K_g \times 1, p_g \times 1 \) and \( q_g \times 1 \) vectors of non-zero elements of the \( g \)-th column of \( B_n, C_n, A_n, \) and \( R_n, \) respectively, and let \( Y_{g,n}, X_{g,n}, Y_{g,n}, \) and \( U_{g,n} \) 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 \( (g = 1, \ldots, G) \):

\[ \begin{align*}
  y_{g,n} &= Z_{g,n} δ_{g,n} + u_{g,n}, \\
  u_{g,n} &= U_{g,n} ρ_{g,n} + ε_{g,n},
\end{align*} \tag{5} \]

where \( Z_{g,n} = [Y_{g,n}, X_{g,n}, Y_{g,n}] \) and \( δ_{g,n} = [β'_{g,n}, γ'_{g,n}, λ'_{g,n}]'. \)

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 \( W_{s,n} \) and \( M_{r,n} \) are zero. (b) \( \|W_{s,n}\|_1 ≤ c, \|M_{r,n}\|_1 ≤ c \) for some finite constant \( c \) which does not depend on \( n \), and \( \|W_{s,n}\|_∞ = 1, \|M_{r,n}\|_∞ = 1. \)\textsuperscript{10}

\( \text{For clarity, we note that the} \ g \text{-th column of} \ A_n \text{ and} \ R_n \text{ are, respectively, given by} \ [λ_{1g,1,n}, \ldots, λ_{g,p,n}, \ldots, λ_{Gg,1,n}, \ldots, λ_{Gg,p,n}]' \text{ and} \ [0, \ldots, 0, ρ_{g,1,n}, \ldots, ρ_{g,q,n}, 0, \ldots, 0]' \).
Assumption 2  (a) The matrices $I_{nG} - B_n^*$ are nonsingular. (b) The spatial autoregressive parameters satisfy $\sup_n \sum_{r \in I_{g,r,n}} |\rho_{g,r,n}| < 1$ for $g = 1, \ldots, G$, where $I_{g,r,n} \subseteq \{1, \ldots, q\}$ denotes the set of indices associated with the elements of $\rho_{g,r,n}$. (c) The row and column sums of the matrices $[I_{nG} - B_n^*]^{-1}$ 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 $W_{s,n}$ and $M_{r,n}$ are uniformly bounded in absolute value. The assumption that $\|W_{s,n}\|_\infty = 1$ and $\|M_{r,n}\|_\infty = 1$ implies a normalization for the parameters. For interpretation, let $W_{s,n}$ be some spatial weights matrix with $\|W_{s,n}\|_\infty \neq 1$ and let $\lambda_{g,s,n}$ be the corresponding spatial autoregressive parameter on $W_{s,n}y_{i,n}$. Now define $W_{s,n} = W_{s,n}/\|W_{s,n}\|_\infty$ and $\lambda_{g,s,n} = \lambda_{g,s,n}^{\ast} \|W_{s,n}\|_\infty$, then $\|W_{s,n}\|_\infty = 1$ and $\lambda_{g,s,n}W_{s,n} = \lambda_{g,s,n}^{\ast}W_{s,n}$. Thus the normalizations $\|W_{s,n}\|_\infty = 1$ and $\|M_{r,n}\|_\infty = 1$ can always be achieved by appropriately re-scaling the elements of the spatial weights matrix, provided the corresponding spatial autoregressive parameter is correspondingly redefined; for further discussions see Kelejian and Prucha (2010).11

Assumption 2(a) ensures that the first equation of the expression for the reduced form in (4) is well defined. Next observe that $R_{g,n}^* = \text{diag}\{R_{g,n}^*\}$ with $R_{g,n}^* = R_{g,n}^*(\rho_{g,n}) = \sum_{r \in I_{g,r,n}} \rho_{g,r,n}^* M_{r,n}$. In light of this it follows from Assumptions 1(b) and 2(b) that $\|R_{g,n}^*\|_\infty \leq \max_g \sum_{r \in I_{g,r,n}} |\rho_{g,r,n}| < 1$, which in turn implies that $I_{nG} - R_{g,n}^*$ is nonsingular; see Horn and Johnson (1985, p. 301). Consequently, also the second equation of the expression for the reduced form in (4) is well defined, and thus $\gamma_n$ is uniquely defined by the model.

Assumptions 1(b) and 2(b) imply even that $\sup_n \|R_{g,n}^*\|_\infty < 1$, which implies that the row sums of the matrices $[I_n - R_{g,n}^*]^{-1}$ are uniformly bounded in absolute value. To see this observe that $\| [I_n - R_{g,n}^*]^{-1} \|_\infty \leq 1/\left[ 1 - \|R_{g,n}^*\|_\infty \right] \leq 1/\left[ 1 - \sup_n \|R_{g,n}^*\|_\infty \right] < \infty$; see Horn and Johnson (1985, p. 301).

Assumption 3  (a) The matrix of (nonstochastic) exogenous regressors $X_n$ in (2) has full column rank (for $n$ sufficiently large). Furthermore, the elements of $X_n$ are uniformly bounded in absolute value by some finite constant. (b) The elements of the parameter matrices $B_n$, $C_n$ and $\Lambda_n$ are uniformly bounded in absolute value.

An assumption such as Assumption 3(a) is common in the spatial literature. In treating $X_n$ as nonstochastic, our analysis should be viewed as conditional

---

11The 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.
on $X_n$. Note that in part (b) of Assumption 3 uniformity refers to $n$, and thus part (b) should not be interpreted to imply that the parameter space is assumed to be compact. 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. $\varepsilon_n$. In the following let $V_n = [v_{1.n}, \ldots, v_{G.n}]$ be an $n \times G$ matrix of basic innovations and let $v_n = \text{vec}(V_n)$.

**Assumption 4** The innovations $\varepsilon_n$ are generated as follows:

$$\varepsilon_n = (\Sigma_\varepsilon^* \otimes I_n)\nu_n,$$

where $\Sigma_\varepsilon^*$ is a nonsingular $G \times G$ matrix and the random variables $\{v_{i.g.n} : i = 1, \ldots, n, g = 1, \ldots, G\}$ 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 $\Sigma = \Sigma_\varepsilon^* \Sigma_\varepsilon$, then the diagonal elements of $\Sigma$ are bounded by some finite constant.

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

Given (4), we note that Assumption 4 implies furthermore that $Eu_n = 0$ and $Ey_n = (I_{nG} - B_n^*)^{-1}C_n^*x_n$, and that the VC matrices of $u_n$ and $y_n$ are given by, respectively,

$$\begin{align*}
\Omega_{u,n} &= (I_{nG} - R_n^*)^{-1}(\Sigma \otimes I_n)(I_{nG} - R_n^*)^{-1}, \\
\Omega_{y,n} &= (I_{nG} - B_n^*)^{-1}\Omega_{u,n}(I_{nG} - B_n^*)^{-1}.
\end{align*}$$

Assumptions 2 and 4 imply that the row and column sums of the VC matrix of $u_n$ (and similarly those of $y_n$) are uniformly bounded in absolute value, thus limiting the degree of correlation between, respectively, the elements of $u_n$ and of $y_n$.

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

$$n^{-1}Z_{g,n}'A_n u_{g,n} - n^{-1}EZ_{g,n}'A_n u_{g,n} = o_p(1).$$

for any $n \times n$ real matrix $A_n$ 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 (5) by $I_n - R_{g,n}^*(\rho_{g,n})$ yields

$$y_{g,n} = Z_{g,n}\delta_{g,n} + \varepsilon_{g,n},$$

with $y_{g,n} = y_{g,n}(\rho_{g,n}) = [I_n - R_{g,n}^*(\rho_{g,n})] y_{g,n}$ and $Z_{g,n} = Z_{g,n}(\rho_{g,n}) = [I_n - R_{g,n}^*(\rho_{g,n})] Z_{g,n}$. Stacking the transformed equations yields

$$y_{*n} = Z_{*n}\delta_{n} + \varepsilon_{n},$$

with $y_{*n} = [y_{*1,n}, \ldots, y_{*G,n}]'$, $Z_{*n} = diag_g^G [Z_{g,n}]$, $\delta_n = [\delta'_{1,n}, \ldots, \delta'_{G,n}]'$ and where $\varepsilon_{n}$ is as defined above.

### 3 Exemplary Applications

In the following we give two examples involving the use of higher-order spatial lags. 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).

#### 3.1 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. In case that those elements are specified incorrectly, the model is misspecified and estimates will generally be inconsistent. Allowing for higher-order spatial lags provides important flexibility and robustness in modeling network interactions.

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 $W = (w_{ij})$ be the basic spatial weights matrix, let $d_{ij}$ denote some distance measure between units $i$ and $j$, and let $d_{ij}^C$ be some contiguity measure taking on values of one or zero. Then, the researcher may specify the weights as the product of the contiguity measure and a polynomial in $d_{ij}$, treating the coefficients of the polynomial as unknown parameters:\(^{12}\)

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

Now, suppose the researcher models $y_g$ as a function of, say, $\lambda_{lg}W_{y_l}$, then clearly

$$\lambda_{lg}W_{y_l} = \left[\lambda_{lg} \sum_{s=1}^p \lambda_s W_s\right] y_l = \left[\sum_{s=1}^p \lambda_{lg,s} W_s\right] y_l$$

\(^{12}\)Alternatively the researcher could specify a polynomial in $1/d_{ij}$.
with \( \lambda_{ig,s} = \lambda_{ig} \lambda_s \), \( W_s = (w_{ij,s}) \), and \( w_{ij,s} = d_{ij}^s d_{ij}^* \). In allowing for a 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.\(^1\)

### 3.2 Social Interactions in Multiple Activities

The following example extends Cohen-Cole, Liu and Zenou (2014). We follow their basic setup apart, 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 \( \gamma_{ig,s} = \sum_{j=1}^n w_{ij,s} y_{ij} \) denote the average effort level for activity \( g \) by the \( s \)-th group of peers, and assume that the utility of individual \( i \) is of the following linear-quadratic form:

\[
\begin{align*}
\pi_{ig} &= \sum_{g=1}^G \pi_{ig} y_{ig} + \sum_{g=1}^G y_{ig} \sum_{l=1}^G \sum_{s=1}^p \lambda_{lg,s} \gamma_{il,s} - \frac{1}{2} \sum_{g=1}^G b_{lg} y_{ig}^2 - \sum_{g=1}^G \sum_{l=1,l\neq g}^G b_{lg} y_{il} y_{ig}.
\end{align*}
\]

The specification considered in Cohen-Cole, Liu and Zenou (2014) corresponds to \( p = 1 \). The first-order conditions for the maximum of \( u(y_{i1}, \ldots, y_{iG}) \) yield

\[
\begin{align*}
y_{ig} &= \pi_{ig} + \sum_{l=1,l\neq g} b_{lg} y_{il} + \sum_{l=1}^G \sum_{s=1}^p \lambda_{lg,s} \gamma_{il,s},
\end{align*}
\]

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

\[
\begin{align*}
y_g &= \pi_g + \sum_{l=1,l\neq g} b_{lg} y_l + \sum_{l=1}^G \sum_{s=1}^p \lambda_{lg,s} W_s y_l \quad (10)
\end{align*}
\]

with \( \pi_g = [\pi_{i1}, \ldots, \pi_{ig}]' \). Similar to Cohen-Cole, Liu and Zenou (2014) assume that \( \pi_g \) can be modeled as

\[
\begin{align*}
\pi_g &= \sum_{k=1}^K c_k x_k + u_g.
\end{align*}
\]

\(^1\)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.
Substituting (11) into (10) 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 \( \mathbf{W}_i \) 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 \( \mathbf{x}_i \) covariates may represent group indicator variables, and others may be spatial lags of some basic covariates.

4 Moment Conditions

Recall that from the Cochrane-Orcutt transformed form of the model (7) we have\(^\dagger\)

\[ \varepsilon_{g,n} = \varepsilon_{g,n}(\rho_{g,n}, \delta_{g,n}) = [I_n - R_{g,n}^*(\rho_{g,n})] [y_{g,n} - Z_{g,n} \delta_{g,n}] . \]

The estimators for the model parameters \( \rho_{g,n} \) and \( \delta_{g,n} \) considered in this paper will utilize a set of linear and quadratic moment conditions of the form \((g = 1, \ldots, G)\)

\[ E\mathbf{m}^\delta_{g,n}(\rho_{g,n}, \delta_{g,n}) = E n^{-1} \mathbf{H}_n \varepsilon_{g,n} = \mathbf{0}, \quad (12) \]

\[ E\mathbf{m}^\rho_{g,n}(\rho_{g,n}, \delta_{g,n}) = E \left[ \begin{array}{c} n^{-1} \varepsilon_{g,n}' A_{1,n} \varepsilon_{g,n} \\ \vdots \\ n^{-1} \varepsilon_{g,n}' A_{S,n} \varepsilon_{g,n} \end{array} \right] = \mathbf{0}, \quad (13) \]

where the \( n \times p_H \) instrument matrix \( \mathbf{H}_n \) in the linear form and the \( n \times n \) weighting matrices \( \mathbf{A}_{s,n} \) in the quadratic forms are non-stochastic. In the following we will also simply write \( \mathbf{m}^\delta_{g,n} \) and \( \mathbf{m}^\rho_{g,n} \) for the sample moment vector at the true parameter values.\(^\dagger\)

We maintain the following assumptions regarding the instruments \( \mathbf{H}_n \).

**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 \( \mathbf{H}_n = [\mathbf{X}_n, \mathbf{M}_{1,n} \mathbf{X}_n, \ldots, \mathbf{M}_{p,n} \mathbf{X}_n] \).

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

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

\(\dagger\) 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.

(a) \( Q_{HH} = \lim_{n \to \infty} n^{-1} \mathbf{H}_n' \mathbf{H}_n \) is finite and nonsingular.

---

\(\dagger\)
(b) $Q_{HZ,g} = \text{plim}_{n \to \infty} n^{-1} H_n^T Z_{g,n}$ and $Q_{HMZ,g} = \text{plim}_{n \to \infty} n^{-1} H_n^T M_{r,n} Z_{g,n}$ are finite and have full column rank.

c) Let $Q_{HZ,g}(\rho_{g,n}) = Q_{HZ,g} - \sum_{r \in \mathcal{I}_p} \rho_{g,r,n} Q_{HMZ,g,r,g}$, then

$$\lambda_{\min} \left[ Q_{HZ,g}(\rho_{g,n})^{-1} Q_{HZ,g}(\rho_{g,n}) \right] \geq c$$ 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. The ideal instruments for $Y_{g,n}$ and $\bar{Y}_{g,n}$ are given by their conditional means. Observe that in light of (3) we have

$$EY_n = (I_{nG} - B_n^*)^{-1} C_n x_n$$

with $B_n^* = [(B_n' \otimes I_n) + (A_n' \otimes I_n) (I_G \otimes W_n)]$. For large $n$ the accurate computation of the inverse of $I_{nG} - B_n^*$, 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 ideal instruments have so far only been obtained by restricting the parameter space to a compact interval in, say, $(-1, 1)$. To avoid these difficulties and limitations we employ an approximation of the ideal instruments, which is consistent in spirit with the approach adopted in the above cited literature.

Given $\|B_n^*\| < 1$ we have $(I_{nG} - B_n^*)^{-1} = \sum_{d=0}^{\infty} (B_n^*)^d$ and thus $EY_n = \sum_{d=0}^{\infty} (B_n^*)^d C_n x_n$. In light of the structure of $B_n^*$ it is not difficult to see that the blocks of $(I_{nG} - B_n^*)^{-1}$ can be expressed as infinite weighted sums of the matrices $I_n, \{W_{j_1,n} W_{j_2,n} \}^{P}_{j_1,j_2=1}, \{W_{j_1,n} W_{j_2,n} W_{j_3,n} \}^{P}_{j_1,j_2,j_3=1}, \ldots$ Adopting the notation $(A_j)^m := [A_1, \ldots, A_m]$ for any set of conformable matrices $A_1, \ldots, A_m$, define

$$X_{1,n} = [W_{j_1,n} x_n]^P_{j_1=1}, \ldots, X_{R,n} = [W_{j_1,n} W_{j_2,n} \ldots W_{j_{R-1},n} x_n]^P_{j_1,j_2,\ldots,j_{R-1}=1}$$

and let $H_{R,n} = [X_{1,n} X_{1,n} \ldots X_{R,n}]$. Now suppose $W_{s_1}, \ldots, W_{s_y}$ are the spatial weights matrices appearing in $Y_{g,n}$, then by including in $H_n$ the linearly independent columns of $[H_{R,n}, W_{s_1} H_{R,n}, \ldots, W_{s_y} H_{R,n}]$, we may view the fitted values of $Z_{g,n}$ as computationally simple approximations of the ideal instruments $EZ_{g,n}$. Suppose further that $M_{r_1}, \ldots, M_{r_y}$ are the spatial weights matrices in the disturbance process of the $g$-th equation, then by including in $H_n$ also the linearly independent columns of $M_{r_1} [H_{R,n}, W_{s_1} H_{R,n}, \ldots, W_{s_y} H_{R,n}]$, $M_{r_2} [H_{R,n}, W_{s_1} H_{R,n}, \ldots, W_{s_y} H_{R,n}]$, \ldots, $M_{r_y} [H_{R,n}, W_{s_1} H_{R,n}, \ldots, W_{s_y} H_{R,n}]$ we may view the fitted values of $Z_{g,n}$ as computationally simple approximations of the ideal instruments $E Z_{g,n}$. Preliminary Monte Carlo results suggest that in many situations relatively low values of $R$ are sufficient for providing a good approximation.

Assumption 6(a) is standard. Assumption 6(b) ensures the identification of $\delta_{g,n}$, while Assumption 6(c) ensures the identification of $\rho_{g,n}$.

We will maintain the following assumptions regarding the matrices $A_{s,n}$.
**Assumption 7**: The row and column sums of the matrices $A_{s,n}$, $s = 1, \ldots, S$, are bounded uniformly in absolute value by some finite constant and, furthermore, all diagonal elements of $A_{s,n}$ are zero for any $s = 1, \ldots, S$.

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

$$M_{r,n}, M'_{r,n}M_{r,n} - \text{diag}(M'_{r,n}M_{r,n}),$$
$$W_{s,n}, W'_{s,n}W_{s,n} - \text{diag}(W'_{s,n}W_{s,n}), M'_{s,n}W_{s,n} - \text{diag}(M'_{s,n}W_{s,n}).$$

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

$$\gamma_{g,n} - \Gamma_{g,n}r_{g,n}(\rho_{g,n}) = 0,$$

where

$$\gamma_{g,n} = \begin{bmatrix} \gamma_{1,g,n} \\ \vdots \\ \gamma_{S,g,n} \end{bmatrix}_{S \times 1}, \Gamma_{g,n} = \begin{bmatrix} \Gamma_{11,g,n} & \Gamma_{12,g,n} & \Gamma_{13,g,n} \\ \vdots & \vdots & \vdots \\ \Gamma_{S1,g,n} & \Gamma_{S2,g,n} & \Gamma_{S3,g,n} \end{bmatrix}_{S \times q'_g}, r_{g,n}(\rho_{g,n}) = \begin{bmatrix} r_{1,g,n} \\ r_{3,g,n} \end{bmatrix}_{q'_g \times 1},$$

with

$$\gamma_{s,g,n} = n^{-1}Eu'_{g,n}\overline{A}_{s,n}u_{g,n},$$
$$\Gamma_{s1,g,n} = n^{-1}(2Eu'_{g,n}M'_{1,n}\overline{A}_{s,n}u_{g,n}, \ldots, 2Eu'_{g,n}M'_{q_g,n}\overline{A}_{s,n}u_{g,n}),$$
$$\Gamma_{s2,g,n} = -n^{-1}(Eu'_{g,n}M'_{1,n}\overline{A}_{s,n}M_{1,n}u_{g,n}, \ldots, Eu'_{g,n}M'_{q_g,n}\overline{A}_{s,n}M_{q_g,n}u_{g,n}),$$
$$\Gamma_{s3,g,n} = -n^{-1}(2Eu'_{g,n}M'_{1,n}\overline{A}_{s,n}M_{2,n}u_{g,n}, \ldots, 2Eu'_{g,n}M'_{q_g-1,n}\overline{A}_{s,n}M_{q_g,n}u_{g,n}),$$
$$r_{1,g,n} = (\rho_{g,1,n}, \ldots, \rho_{g,q_g,n})',$$
$$r_{2,g,n} = (\rho_{g,1,n}^2, \ldots, \rho_{g,2,n}^2)',$$
$$r_{3,g,n} = (\rho_{g,1,n}\rho_{g,2,n}, \ldots, \rho_{g,1,n}\rho_{g,q_g,n}, \ldots, \rho_{g,q_g-1,n}\rho_{g,q_g,n})',$$

and $q'_g = 2q_g + q_g(q_g - 1)/2$. Now let $\tilde{\delta}_{g,n}$ be some estimator for $\delta_{g,n}$, let $\tilde{u}_{g,n} = y_{g,n} - Z_{g,n}\tilde{\delta}_{g,n}$, and let $\tilde{\Gamma}_{g,n}$ and $\tilde{\gamma}_{g,n}$ denote the corresponding estimators of $\Gamma_{g,n}$ and $\gamma_{g,n}$, respectively, which are obtained by suppressing the expectations operator and replacing $u_{g,n}$ by $\tilde{u}_{g,n}$ in the above expressions. Then

$$m_{g,n}(\rho_{g}, \tilde{\delta}_{g,n}) = \tilde{\gamma}_{g,n} - \tilde{\Gamma}_{g,n}r_{g,n}(\rho_{g}).$$

**5 Generic Asymptotic Properties**

In this section we give a generic discussion of the asymptotic properties of GMM estimators for $\rho_{g,n}$ and $\delta_{g,n}$ corresponding to the moment conditions in (12) and
We focus 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, see, e.g. Das et al. (2003). A two-step procedure also provides some robustness for the estimation of $\delta_{g,n}$ against misspecification of the disturbance process. As will be seen below, the limiting distribution of the GMM estimators for $\rho_{g,n}$ will depend on the estimator of $\delta_{g,n}$ used in computing the estimated residual. That is, $\delta_{g,n}$ is not a nuisance parameter for $\rho_{g,n}$ (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 $\rho_n$ 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.

5.1 Consistency of GMM Estimator for $\rho$

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

$$\tilde{\rho}_{g,n} = \tilde{\rho}_{g,n}(\tilde{\Upsilon}_{g,n}) = \arg\min_{\bar{\Upsilon}_{g,n}(\rho_{g,n})} m_{g,n}(\bar{\Upsilon}_{g,n}, \tilde{\delta}_{g,n})' \tilde{\Upsilon}_{g,n} m_{g,n}(\bar{\Upsilon}_{g,n}, \tilde{\delta}_{g,n})$$

with $a_p \geq 1$. We note that the objective function for $\tilde{\rho}_{g,n}$ remains well defined even for values of $\bar{\Upsilon}_{g,n}$ for which $I_n - R_{g,n}(\rho_{g,n})$ is singular, which allows us to take as the optimization space a compact set containing the true parameter space.

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

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

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

Assumption 8 requires $\Gamma_{g,n} \Gamma_{g,n}$ to be nonsingular and in conjunction with Assumption 9 ensures that the smallest eigenvalue of $\Gamma_{g,n} \Upsilon_{g,n} \Gamma_{g,n}$ is uniformly
bounded away from zero. This will be sufficient to demonstrate that $\rho_{g,n}$ is identifiable unique w.r.t. the nonstochastic analogue of the objective function of the GMM estimator

$$E \text{m}_{g,n}(\bar{\sigma}_{g}, \delta_{g,n})'Y_{g,n}E \text{m}_{g,n}(\bar{\sigma}_{g}, \delta_{g,n})$$

$$= [\gamma_{g,n} - \Gamma_{g,n} r_{g,n}(\bar{\sigma}_{g,n})]'Y_{g,n} [\gamma_{g,n} - \Gamma_{g,n} r_{g,n}(\bar{\sigma}_{g,n})].$$

Of course, Assumption 9 regarding $\bar{Y}_{g,n}$ and $Y_{g,n}$ is satisfied for $\bar{Y}_{g,n} = Y_{g,n} = I_{S}$. In this case, the estimator defined by (16) reduces to a nonlinear least squares estimator. Choices of $\bar{Y}_{g,n}$ which result in efficient estimates of $\rho_{g,n}$ will be discussed below in conjunction with the asymptotic normality result.

Our basic consistency result for $\rho_{g,n}$ is given by next theorem.

**Theorem 1 (Consistency)** Let $\tilde{\rho}_{g,n} = \hat{\rho}_{g,n}(\bar{Y}_{g,n})$ denote the GMM estimator for $\rho_{g,n}$ defined by (16). Suppose Assumptions 1-9 hold, and suppose that $n^{-1}(\delta_{g,n} - \delta_{g,n}) = O_{p}(1)$, then,

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

**5.2 Asymptotic Distribution of GMM Estimator for $\rho$**

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

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

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

with $T_{gh,n} = F_{gh,n} P_{gh,n}$, where $F_{gh,n}$ and $P_{gh,n}$ are, respectively, $n \times p_{F}$ and $p_{F} \times p_{\delta_{g}}$ real non-stochastic matrices whose elements are uniformly bounded 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 $T_{gh,n}$ are again uniformly bounded in absolute value by a finite constant).

In the Appendix we show that our spatial 2SLS and 3SLS estimators satisfy this assumption. For 2SLS estimators of the parameters of the, say, first equation, $\sum_{h=1}^{G} T'_{1h,n} \varepsilon_{h,n}$ reduces to $T'_{11,n} \varepsilon_{1,n}$. Also note that under Assumptions 4 and 10 we have $n^{1/2}(\delta_{g,n} - \delta_{g,n}) = O_{p}(1)$, 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

$$J_{g,n} = -E \frac{\partial m_{g,n}}{\partial \rho_{g,n}} = \Gamma_{g,n} \frac{\partial r_{g,n}(\rho_{n,g})}{\partial \rho_{g,n}}, \quad (17)$$

and the $S \times S$ matrix \( \Psi_{gg,n}^{pp} = (\psi_{rs,gg,n}^{pp}) \) where

$$\psi_{rs,gg,n}^{pp} = \sigma_{gg}^2 (2n)^{-1} tr \left[ \left( A_{r,n} + A'_{r,n} \right) \left( A_{s,n} + A'_{s,n} \right) \right]$$

$$+ n^{-1} \alpha_{g,r,n} \left( \sum_{h=1}^{G} \sum_{i=1}^{G} \sigma_{hi} T'_{gh,n} T_{gl,n} \right) \alpha_{g,s,n}, \tag{18}$$

with

$$\alpha_{g,r,n} = -n^{-1} E \left[ Z_{g,n} (I_n - R_{g,n}(\rho_{g,n})) (A_{r,n} + A'_{r,n}) (I_n - R_{g,n}(\rho_{g,n})) u_{g,n} \right].$$

As shown in the proof of the subsequent theorem, \( \Psi_{gg,n}^{pp} \) is the asymptotic VC matrix of the sample moment vector \( m_{g,n}(\rho_g, \delta_{g,n}) \). The second term in (18) stems from the fact that the sample moment vector depends on estimated residuals. If the true residuals were observable, we could take \( \delta_{g,n} = \delta_{g,n} \) or \( T_{gh,n} = 0 \), in which case the second term in (18) is zero.

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

$$n^{1/2}(\tilde{\rho}_{g,n} - \rho_{g,n}) = \left[ J'_{g,n} Y_{g,n} J_{g,n} \right]^{-1} J'_{g,n} Y_{g,n} \xi_{g,n} + o_p(1), \tag{19}$$

where

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

and \( (\Psi_{gg,n}^{pp})^{-1/2} \xi_{g,n} \xrightarrow{d} N(0, I_S) \). Furthermore \( n^{1/2}(\tilde{\rho}_{g,n} - \rho_{g,n}) = O_p(1) \) and

$$\lambda_{\min}(\Psi_{gg,n}^{pp}(Y_{g,n})) \geq \text{const} > 0$$

for

$$\Omega_{gg,n}^{pp}(Y_{g,n}) = \left( J'_{g,n} Y_{g,n} J_{g,n} \right)^{-1} J'_{g,n} Y_{g,n} \Psi_{gg,n}^{pp} Y_{g,n} J_{g,n} (J'_{g,n} Y_{g,n} J_{g,n})^{-1}. \tag{21}$$

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

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

We next define a consistent estimator for $\Omega_{gg,n}^{pp}(\bar{Y}_{g,n})$. As a preliminary result we have the following lemma.

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

$$\bar{\sigma}_{gh,n} = n^{-1}\bar{\varepsilon}'_{g,n}\bar{\varepsilon}_{h,n}$$

with $\bar{\varepsilon}_{g,n} = y_{g,n}(\bar{\rho}_{g,n}) - Z_{g,n}(\bar{\rho}_{g,n})\bar{\delta}_{g,n}$ and assume $\bar{\delta}_{g,n} - \delta_{g,n} = o_p(1)$ and $\bar{\rho}_{g,n} - \rho_{g,n} = o_p(1)$, then $\bar{\sigma}_{gh,n} - \sigma_{gh} = o_p(1)$.

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

$$\bar{J}_{g,n} = \bar{\Gamma}_{g,n} \frac{\partial \tau_{g,n}(\bar{\rho}_{n,g})}{\partial \rho_{n,g}}.$$ (23)

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

$$\bar{\psi}_{rs,gg,n}^{pp} = \bar{\sigma}_{gg,n}(2n)^{-1}tr \left[ (A_{r,r,n} + A'_{r,n}) (A_{s,s,n} + A'_{s,n}) \right] + n^{-1}\bar{\alpha}'_{g,r,n} \sum_{h=1}^{G} \sum_{i=1}^{G} \bar{T}'_{gh,n} \bar{T}_{gh,n} \bar{\alpha}_{g,s,n}$$

with

$$\bar{\alpha}_{g,r,n} = -n^{-1} [Z_{g,n}(I_n - R_{g,n}(\bar{\rho}_{g,n}))(A_{r,n} + A'_{r,n})(I_n - R_{g,n}(\bar{\rho}_{g,n}))\bar{u}_{g,n}],$$

where $\bar{T}_{gh,n}$ is some estimator for $T_{gh,n}$, and $\bar{u}_{g,n} = y_{g,n} - Z_{g,n}\bar{\delta}_{g,n}$. Given estimators for $J_{g,n}$ and $\Psi_{gg,n}^{pp}$ we can now formulate the following estimator for $\Omega_{gg,n}^{pp}$:

$$\Omega_{gg,n}^{pp}(\bar{\Sigma}_{g,n}) = (J'_{g,n}\bar{\Sigma}_{g,n}\bar{J}_{g,n})^{-1}(J'_{g,n}\bar{\Sigma}_{g,n}\bar{J}_{g,n})^{-1}(J'_{g,n}\bar{Y}_{g,n}\bar{\Psi}_{gg,n}\bar{Y}_{g,n}\bar{\Sigma}_{g,n}\bar{J}_{g,n})^{-1},$$

(25)

The next theorem establishes the consistency of $\bar{\Psi}_{gg,n}^{pp}$ and $\Omega_{gg,n}^{pp}$.

---

This follows from Corollary F4 in Pötscher and Prucha (1997). Compare also the discussion on pp. 86-87 in that reference.
Theorem 3 (VC Matrix Estimation) Suppose all assumptions of Theorem 2 hold, except for Assumption 9, and suppose that \( n^{-1} T'_{gh,n} T_{gl,n} - n^{-1} T'_{gh,n} T_{gl,n} = o_p(1) \). Then, provided that \( \tilde{\rho}_{g,n} - \rho_{g,n} = o_p(1) \),

\[
\tilde{\Psi}_{pp}^{gg,n} - \Psi_{pp}^{gg,n} = o_p(1), \quad (\tilde{\Psi}_{pp}^{gg,n})^{-1} - (\Psi_{pp}^{gg,n})^{-1} = o_p(1),
\]

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

\[
\Omega_{pp}^{gg,n} - \Omega_{pp}^{gg,n} = o_p(1), \quad (\Omega_{pp}^{gg,n})^{-1} - (\Omega_{pp}^{gg,n})^{-1} = o_p(1),
\]

and \( \Omega_{pp}^{gg,n} = O(1), \quad (\Omega_{pp}^{gg,n})^{-1} = O(1) \).

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

5.3 Joint Asymptotic Distribution of Estimators for \( \rho \) and \( \delta \)

In the following, let \( \delta_n = [\delta_{1,n}', \ldots, \delta_{G,n}']' \) and \( \rho_n = [\rho_{1,n}', \ldots, \rho_{G,n}']' \), and let \( \tilde{\delta}_n = [\tilde{\delta}_{1,n}', \ldots, \tilde{\delta}_{G,n}']' \) and \( \tilde{\rho}_n = [\tilde{\rho}_{1,n}', \ldots, \tilde{\rho}_{G,n}']' \) denote the corresponding estimators as defined in the previous subsections. In this section, we derive the joint limiting distribution of \( \delta_n \) and \( \tilde{\rho}_n \). 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 \( [\eta_n', \xi_n'] \) with \( \eta_n = [\eta_{1,n}', \ldots, \eta_{G,n}']' \) and \( \xi_n = [\xi_{1,n}', \ldots, \xi_{G,n}']' \), where \( \eta_{g,n} = n^{-1/2} \sum_{h=1}^{G} T'_{gh,n} \xi_{h,n} \) and \( \xi_{g,n} \) are defined in (20). Let\(^{16}\)

\[
T_n = \begin{bmatrix}
T_{11,n} & \cdots & T_{G1,n} \\
\vdots & \ddots & \vdots \\
T_{1G,n} & \cdots & T_{GG,n}
\end{bmatrix},
\]

then the VC matrix of \( [\eta_n', \xi_n'] \) is given by

\[
\Psi_n = \begin{bmatrix}
\Psi_n^{\delta \delta} & \Psi_n^{\delta \rho} \\
\Psi_n^{\rho \delta} & \Psi_n^{\rho \rho}
\end{bmatrix}, \quad (26)
\]

where

\[
\Psi_n^{\delta \delta} = E\xi_n' \xi_n = \begin{bmatrix}
\Psi_{11,n}^{\delta \delta} & \cdots & \Psi_{G1,n}^{\delta \delta} \\
\vdots & \ddots & \vdots \\
\Psi_{1G,n}^{\delta \delta} & \cdots & \Psi_{GG,n}^{\delta \delta}
\end{bmatrix},
\]

\[
\Psi_n^{\rho \rho} = E\eta_n' \eta_n = n^{-1} T_n' (\Sigma \otimes I_n) T_n,
\]

\[
\Psi_n^{\delta \rho} = E\xi_n' \eta_n = n^{-1} T_n' (\Sigma \otimes I_n) T_n diag_g \left[ \alpha_{g,1,n}, \ldots, \alpha_{g,S,n} \right],
\]

\[\text{PPP Internal Note: The indices for the submatrices of } T_n \text{ are not as usual. We may want to change this throughout the paper. PPP}\]
and where the \((r,s)\)-th element of \(\Psi_{gh,n}^{pp}\) is given by
\[
\psi_{rs,gh,n}^{pp} = \sigma_{gh,n}^2 (2n)^{-1} \text{tr} \left[ (A_{r,n} + A'_{r,n}) \left( A_{s,n} + A'_{s,n} \right) \right] \\
+ n^{-1} \alpha_{g,r,n} \left( \sum_{u=1}^{G} \sum_{v=1}^{G} \sigma_{uv,n} T'_{gu,n} T_{hv,n} \right) \alpha_{h,s,n}.
\]

Analogous to (24) consider the following estimator \(\tilde{\Psi}_{gh,n}^{pp} = \left( \tilde{\psi}_{rs,gh,n}^{pp} \right)^{-1}\) for \(\Psi_{gh,n}^{pp}\), where
\[
\tilde{\psi}_{rs,gh,n}^{pp} = \tilde{\sigma}_{gh,n}^2 (2n)^{-1} \text{tr} \left[ (A_{r,n} + A'_{r,n}) \left( A_{s,n} + A'_{s,n} \right) \right] \\
+ n^{-1} \tilde{\alpha}_{g,r,n} \left( \sum_{u=1}^{G} \sum_{v=1}^{G} \tilde{\sigma}_{uv,n} \tilde{T}'_{gu,n} \tilde{T}_{hv,n} \right) \tilde{\alpha}_{h,s,n},
\]
with
\[
\tilde{\alpha}_{g,r,n} = -n^{-1} \left[ Z'_{g,n}(I_n - R_{g,n}^*(\tilde{\rho}_{g,n}))(A_{r,n} + A'_{r,n})(I_n - R_{g,n}^*(\tilde{\rho}_{g,n}))\tilde{u}_{g,n} \right],
\]
with \(\tilde{T}_{gh,n}\) being some estimator for \(T_{gh,n}\), and \(\tilde{u}_{g,n} = y_{g,n} - Z_{g,n}\tilde{\delta}_{g,n}\). Next, consider the estimators
\[
\tilde{\Psi}_{\tilde{\rho}} = n^{-1} \tilde{T}_n (\tilde{\Sigma}_n \otimes I_n) \tilde{T}_n, \\
\tilde{\Psi}_{\tilde{\delta}} = n^{-1} \tilde{T}_n (\tilde{\Sigma}_n \otimes I_n) \tilde{T}_n \text{diag}^G_{j=1} (\tilde{\alpha}_{g,1,n}, \ldots, \tilde{\alpha}_{g,S,n}),
\]
then our estimator for \(\Psi_n\) can be formulated as
\[
\tilde{\Psi}_n = \begin{bmatrix} \tilde{\Psi}_{\tilde{\rho}}^{\tilde{\delta}} \\ \tilde{\Psi}_n^{\tilde{\rho}} \tilde{\Psi}_n^{\tilde{\delta}} \end{bmatrix}.
\]

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

**Theorem 4 (Asymptotic Normality)** Let \(\tilde{\rho}_n = [\tilde{\rho}_{1,n}, \ldots, \tilde{\rho}_{G,n}]\) where \(\tilde{\rho}_{g,n} = \tilde{\rho}_{g,n}(\tilde{Y}_{g,n})\) denotes the GMM estimator for \(\rho_{g,n}\) defined by (16), and let \(\tilde{\delta}_n = [\tilde{\delta}_{1,n}, \ldots, \tilde{\delta}_{G,n}]\) be an estimator for \(\delta\). Then given Assumptions 1-10, \(n^{-1} \tilde{T}_{g,n}^\prime \tilde{T}_{g,n} - n^{-1} \tilde{T}_{g,n}^\prime \tilde{T}_{kl,n} = o_p(1)\), and given that \(\lambda_{\min}(\Psi_n) \geq \text{const} > 0\), we have
\[
n^{1/2} \begin{bmatrix} \tilde{\delta}_n - \delta_n \\ \tilde{\rho}_n - \rho_n \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & \text{diag}_G [J'_{g,n}Y_{g,n}J_{g,n}]^{-1} J'_{g,n} Y_{g,n} \end{bmatrix} \begin{bmatrix} \eta_n \\ \xi_n \end{bmatrix} + o_p(1),
\]
with
\[
\Psi_n^{-1/2} \begin{bmatrix} \eta_n \\ \xi_n \end{bmatrix} \xrightarrow{d} N(0, I_d),
\]
19
where \( d = \sum_{g=1}^{G} (G_g + K_g + p_g + q_g) \). Furthermore, let

\[
\Omega_n = \begin{bmatrix}
1 & 0 \\
0 & \text{diag} G_g = 1 \left( (J'_{g,n} Y_{g,n} J_{g,n})^{-1} \right) \end{bmatrix} \Psi_n
\]

(28)

\[
\bar{\Omega}_n = \begin{bmatrix}
1 & 0 \\
0 & \text{diag} G_g = 1 \left( (\bar{J}'_{g,n} \bar{Y}_{g,n} \bar{J}_{g,n}) \right) \end{bmatrix} \bar{\Psi}_n
\]

Then, \( \bar{\Psi}_n - \Psi_n = o_p(1), \bar{\Psi}_n^{-1} - \Psi_n^{-1} = o_p(1), \bar{\Psi}_n = O(1), \Psi_n = O(1) \) and \( \bar{\Omega}_n - \Omega_n = o_p(1), \bar{\Omega}_n^{-1} - \Omega_n^{-1} = o_p(1), \Omega_n = O(1), \bar{\Omega}_n^{-1} = O(1) \).

Theorem 4 implies that the difference between the joint cumulative distribution function of all model parameters \( n^{1/2}[(\delta_n - \delta'_n), (\rho_n - \rho'_n)]' \) and that of \( N(0, \Omega_n) \) converges pointwise to zero so that using the latter as an approximation of the former is justified.\(^\text{17}\) The theorem also states that \( \Omega_n \) is a consistent estimator of \( \Omega \). 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 \( n^{1/2}[(\delta_n - \delta'_n), (\rho_n - \rho'_n)]' \).

6 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 \( \rho_{g,n} \), which employ an initial estimator for \( \delta_{g,n} \) 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.

6.1 Definition of Limited Information Estimators

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

\(^{17}\)This follows from Corollary F4 in Pötscher and Prucha (1997). Compare also the discussion on pp. 86-87 in that reference.
Step 1a: 2SLS estimator of $\delta_{g,n}$
As a first step, we apply 2SLS to the $g$-th equation of the untransformed model (5) using the instrument matrix $H_n$ in Assumptions 5 and 6 to estimate $\delta_{g,n}$. The 2SLS estimator, say $\hat{\delta}_{g,n}$, is then defined as

$$\hat{\delta}_{g,n} = (\hat{Z}_{g,n}^T Z_{g,n})^{-1} \hat{Z}_{g,n}^T y_{g,n},$$

(29)

where $\hat{Z}_{g,n} = P_{H_n}Z_{g,n} = (X_{g,n}, \hat{Y}_{g,n}, \bar{y}_{g,n})$, $\hat{Y}_{g,n} = P_{H_n}Y_{g,n}$, $\bar{y}_{g,n} = P_{H_n}y_{g,n}$, and where $P_{H_n} = H_n(H_n'H_n)^{-1}H_n'$. \(^{18}\)

Step 1b: Initial GMM estimator of the vector $\rho_{g,n}$ based on 2SLS residuals
Let $\bar{u}_{g,n} = u_{g,n}(\hat{\delta}_{g,n}) = y_{g,n} - Z_{g,n}\hat{\delta}_{g,n}$ denote the 2SLS residuals of the $g$-th equation, and let $m_{g,n}(\bar{\rho}_{g,n}^*)$ denote the corresponding sample moment vector as defined in (15). Our initial GMM estimator for $\rho_{g,n}$ is now defined as

$$\bar{\rho}_{g,n} = \arg\min_{\rho_{g,n}} m_{g,n}(\bar{\rho}_{g,n}^*; \hat{\delta}_{g,n})'m_{g,n}(\bar{\rho}_{g,n}^*; \hat{\delta}_{g,n}).$$

(30)

with $a_p \geq 1$.

Step 2a: GS2SLS estimator of $\delta_{g,n}$
Analogous to Kelejian and Prucha (1998), we next compute a generalized spatial two-stage least squares (GS2SLS) estimator of $\delta_{g,n}$, $\hat{\delta}_{g,n}(\bar{\rho}_{g,n})$. This estimator is defined as the 2SLS estimator of the $g$-th equation of the spatially Cochrane-Orcutt transformed model (7) with $\rho_{g,n}$ replaced by $\bar{\rho}_{g,n}$, i.e.,

$$\hat{\delta}_{g,n} = \bar{\delta}_{g,n}(\bar{\rho}_{g,n}) = \left(\tilde{Z}_{*g,n}(\bar{\rho}_{g,n})'Z_{*g,n}(\bar{\rho}_{g,n})\right)^{-1}\tilde{Z}_{*g,n}(\bar{\rho}_{g,n})'y_{*g,n}(\bar{\rho}_{g,n}),$$

(31)

where $y_{*g,n}(\bar{\rho}_{g,n}) = [I_n - R_{*g,n}(\bar{\rho}_{g,n})]y_{g,n}$, $Z_{*g,n}(\bar{\rho}_{g,n}) = [I_n - R_{*g,n}(\bar{\rho}_{g,n})]Z_{g,n}$, and $\tilde{Z}_{*g,n}(\bar{\rho}_{g,n}) = P_{H_n}Z_{*g,n}(\bar{\rho}_{g,n})$. We shall also utilize the following estimator for the $(g,g)$-th block of $\Psi_n^{\delta\delta}$ (corresponding to $\hat{\delta}_{g,n}$):

$$\hat{\Psi}_{g,g,n}^{\delta\delta} = \hat{\delta}_{g,n}[n^{-1}\tilde{Z}_{*g,n}(\bar{\rho}_{g,n})'\tilde{Z}_{*g,n}(\bar{\rho}_{g,n})]^{-1},$$

where $\hat{\delta}_{g,g,n} = n^{-1}\tilde{\epsilon}_{g,n}^T\tilde{\epsilon}_{g,n}$ with $\tilde{\epsilon}_{g,n} = y_{*g,n}(\bar{\rho}_{g,n}) - Z_{*g,n}(\bar{\rho}_{g,n})\hat{\delta}_{g,n}$.

Step 2b: Efficient GMM estimator of $\rho_{g,n}$ based on GS2SLS residuals
Let $\bar{u}_{g,n} = y_{g,n} - Z_{g,n}\hat{\delta}_{g,n}$ denote the GS2SLS residuals of the $g$-th equation, and let $m_{g,n}(\bar{\rho}_{g,n}^*)$ denote the corresponding sample moment vector as defined in (15). Then, the corresponding efficient GMM estimator for $\rho_{g,n}$ based on GS2SLS residuals is given by

$$\hat{\rho}_{g,n} = \arg\min_{\rho_{g,n}} m_{g,n}(\bar{\rho}_{g,n}^*; \hat{\delta}_{g,n})'\left(\hat{\Psi}_{g,g,n}^{\delta\delta}\right)^{-1}m_{g,n}(\bar{\rho}_{g,n}^*; \hat{\delta}_{g,n}),$$

(32)

\(^{18}\)In the previous section we used tilde to denote generic estimators. In the following tilde is used to denote our initial 2SLS based estimators.
where \( \hat{\Psi}_{gg,n}^{pp} = (\hat{\psi}_{rs,gg,n}^{pp} ) \) is an estimator of the VC matrix of the limiting distribution of the normalized sample moments \( n^{1/2} m_{g,n}(\rho_{g,n}, \delta_{g,n}). \) Specifically, we have

\[
\hat{\psi}_{rs,gg,n}^{pp} = (2n)^{-1} \hat{a}_{gg,n}^{2} \left[ (A_{r,n} + A'_{s,n})(A_{s,n} + A'_{s,n}) \right] + \hat{\alpha}_{g,r,n}^{'\delta_{g}} \hat{\alpha}_{g,s,n}^{'},
\]

with

\[
\hat{\alpha}_{g,r,n} = -n^{-1} \left[ Z_{g,n}(\hat{\rho}_{g,n})(A_{r,n} + A'_{s,n})(I_{n} - R_{g,n}^{'\hat{\rho}_{g,n}}) \hat{u}_{g,n} \right].
\]

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

### 6.2 Asymptotic Properties of Limited Information Estimators

In this subsection, we derive results concerning the joint limiting distribution of the GS2SLS estimators \( \hat{\rho}_{g,n} \) and \( \hat{\delta}_{g,n} \) by applying the generic limit theory developed in Theorem 4. In preparation we first specialize the expressions for estimators of the \((g,g)\)-th blocks of \( \Psi_{gg,n}^{2}, \Psi_{pp}^{2}, \Omega_{gg}^{2}, \Omega_{pp}^{2}, \Omega_{pp}^{0} \) as implied by the specific structure of \( \hat{\delta}_{g,n} \) and \( \hat{\rho}_{g,n} \). More specifically, let

\[
\hat{\Omega}_{gg,n}^{2} = \hat{\Omega}_{gg,n}^{0}, \quad \hat{\omega}_{gg,n}^{2} = \hat{\Psi}_{gg,n}^{0}, \quad \hat{\Omega}_{gg,n}^{pp} = \left[ \hat{J}_{g,n} \left( \hat{\Psi}_{gg,n}^{pp} \right)^{-1} \hat{J}_{g,n} \right]^{-1},
\]

with

\[
\hat{\Psi}_{gg,n}^{0} = \hat{\Psi}_{gg,n}^{0} \left[ \hat{\alpha}_{g,1,n}, \ldots, \hat{\alpha}_{g,S,n} \right], \\
\hat{\omega}_{g,n} = \left( \hat{\Psi}_{gg,n}^{0} \right)^{-1} \hat{J}_{g,n} \left[ \hat{J}_{g,n} \left( \hat{\Psi}_{gg,n}^{0} \right)^{-1} \hat{J}_{g,n} \right]^{-1},
\]

and \( \hat{J}_{g,n} = J_{g,n}(\hat{\rho}_{g,n}) \). We now have the following result concerning the joint asymptotic distribution of \( \hat{\delta}_{g,n} \) and \( \hat{\rho}_{g,n} \).

**Theorem 5** (Joint Asymptotic Normality of \( \hat{\rho}_{g,n} \) and \( \hat{\delta}_{g,n} \)) Suppose Assumptions 1-8 hold, and that the smallest eigenvalues of \( \Psi_{g,n} \) are bounded away from zero.\(^{19}\) Then, \( \hat{\rho}_{g,n} \) is efficient among the class of GMM estimators based on GS2SLS residuals, and

\[
\left[ \begin{array}{c}
\hat{\delta}_{g,n} - \delta_{g,n} \\
\hat{\rho}_{g,n} - \rho_{g,n}
\end{array} \right] \sim AN \left[ n^{-1} \left( \begin{array}{c}
\hat{\Omega}_{gg,n}^{0} - \Omega_{gg,n}^{0} \\
\hat{\Omega}_{gg,n}^{pp} - \Omega_{gg,n}^{pp}
\end{array} \right) \right].
\]

\(^{19}\)Explicit expressions for the sub-matrices composing \( \Psi_{g,n} \) specialized to \( \hat{\rho}_{g,n} \) and \( \hat{\delta}_{g,n} \) are given in the proof of the theorem.
In the above theorem the estimators for the asymptotic VC matrix of $\hat{\rho}_{g,n}$ and $\hat{\delta}_{g,n}$ employ $\hat{\rho}_{g,n}$ as an estimator for $\rho_{g,n}$. The theorem continues to hold if $\hat{\rho}_{g,n}$ is replaced by $\tilde{\rho}_{g,n}$ (or any other consistent estimator).

### 6.3 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 (7). In the following, we consider full information estimation, where all parameters are estimated jointly from the stacked spatially Cochrane-Orcutt transformed model (8). In particular, we will consider a generalized spatial three-stage least squares (GS3SLS) estimator.

#### Step 3a: GS3SLS estimator of $\delta_n$

The GS3SLS estimator of $\delta_n$ based on the stacked spatially Cochrane-Orcutt transformed model (8) is given by

$$
\hat{\delta}_n = \left( \hat{Z}_n(\hat{\rho}_n) (\hat{\Sigma}_n^{-1} \otimes I_n) \hat{Z}_n(\hat{\rho}_n) \right)^{-1} \hat{Z}_n(\hat{\rho}_n) (\hat{\Sigma}_n^{-1} \otimes I_n)^{-1} \hat{y}_n(\hat{\rho}_n),
$$

(33)

where $\hat{y}_n(\hat{\rho}_n) = [y_{1,n}(\hat{\rho}_n), \ldots, y_{G,n}(\hat{\rho}_n)]'$, $\hat{Z}_n(\hat{\rho}_n) = \text{diag}(\hat{Z}_{1,n}(\hat{\rho}_n), \ldots, \hat{Z}_{G,n}(\hat{\rho}_n))$, $\hat{Z}_{g,n}(\hat{\rho}_n) = P_{\rho_n} Z_{g,n}(\hat{\rho}_n)$, and where $\hat{\Sigma}_n = (\hat{\sigma}_{gh,n})$ with $\hat{\delta}_{gh,n} = n^{-1} \hat{\delta}_{g,n} \hat{\delta}_{h,n}$ and $\hat{\delta}_{g,n} = y_{g,n}(\hat{\rho}_n) - Z_{g,n}(\hat{\rho}_n) \hat{\delta}_{g,n}$. Below, we shall also utilize the following estimator for $\hat{\Psi}_n^{\delta\delta}$ (corresponding to $\hat{\delta}_n$):

$$
\hat{\Psi}_n^{\delta\delta} = \left[ n^{-1} \hat{Z}_n(\hat{\rho}_n) (\hat{\Sigma}_n^{-1} \otimes I_n) \hat{Z}_n(\hat{\rho}_n) \right]^{-1}
$$

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

#### 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 GS2SLS residuals $\hat{u}_{g,n} = y_{g,n} - Z_{g,n} \hat{\delta}_{g,n}$, where $\hat{\delta}_{g,n}$ denotes the $g$-th component of $\hat{\delta}_n$. Let $m_{g,n}(\hat{\rho}_n, \hat{\delta}_n)$ 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

$$
\hat{\rho}_{g,n} = \arg\min_{\rho_n \in \Gamma} m_{g,n}(\hat{\rho}_n, \hat{\delta}_n)' \left( \hat{\Psi}_{gg,n}^{\delta\delta} \right)^{-1} m_{g,n}(\hat{\rho}_n, \hat{\delta}_n),
$$

(34)

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

$$
\hat{\Psi}_{rs, gh, n}^{pp} = (2n)^{-1}\hat{\sigma}_{gh, n}^{2} (A_{r, n} + A'_{r, n}) (A_{s, n} + A'_{s, n}) + \hat{\alpha}_{g, r, n}^{'} \hat{\Psi}_{gh, n}^{pp} \hat{\alpha}_{h, s, n},
$$

with $\hat{\alpha}_{g, r, n} = -n^{-1}\left[Z_{g, n}'(I_{n} - R_{g, n}')(\hat{\rho}_{g, n})A_{r, n}' + A_{r, n}'(I_{n} - R_{g, n}')(\hat{\rho}_{g, n})\hat{\alpha}_{g, n}\right]$.

### 6.4 Asymptotic Properties of Full Information Estimators

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

$$
\hat{\Omega}_{n}^{\delta\delta} = \hat{\Omega}_{n}, \quad \hat{\Omega}_{n}^{\delta\rho} = \hat{\Omega}_{n} diag_{g=1}^{G} (\hat{\alpha}_{g, n}), \quad \hat{\Omega}_{n}^{\rho\rho} = diag_{g=1}^{G} (\hat{\alpha}_{g, n}) \hat{\Omega}_{n} diag_{g=1}^{G} (\hat{\alpha}_{g, n})
$$

with

$$
\hat{\Psi}_{n}^{\delta\rho} = \hat{\Psi}_{n} diag_{g=1}^{G} (\hat{\alpha}_{g, 1, n}, \ldots, \hat{\alpha}_{g, S, n}),
$$

$$
\hat{\Omega}_{g, n} = \left(\hat{\Psi}_{gg, n}^{pp}\right)^{-1} \hat{J}_{g, n} \left(\hat{\Psi}_{gg, n}^{pp}\right)^{-1} \hat{J}_{g, n},
$$

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

**Theorem 6** (Joint Asymptotic Normality of $\hat{\rho}_{n}$ and $\hat{\delta}_{n}$) Suppose Assumptions 1-8 hold, and that the smallest eigenvalues of $\Psi_{n}$ are bounded away from zero. Then,

$$
\begin{bmatrix}
\hat{\delta}_{n} - \delta_{n} \\
\hat{\rho}_{n} - \rho_{n}
\end{bmatrix}
\sim AN\left[n^{-1} \begin{bmatrix}
\hat{\Omega}_{n}^{\delta\delta} & \hat{\Omega}_{n}^{\delta\rho} \\
\hat{\Omega}_{n}^{\rho\rho} & \hat{\Omega}_{n}^{\rho\rho}
\end{bmatrix}\right].
$$

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

---

20 Explicit expressions for the sub-matrices $\Psi_{n}$ are given in the proof of the theorem.
In the above theorem the estimators for the asymptotic VC matrix of \( \hat{\rho}_n \) and \( \hat{\delta}_n \) employ \( \hat{\rho}_g,n \) as an estimator for \( \rho_{g,n} \). The theorem continues to hold if \( \hat{\rho}_g,n \) is replaced by \( \hat{\rho}_g,n \) (or any other consistent estimator.)

7 Limited and Full Information One-Step Estimators

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

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

\[
m_{g,n}(\rho_{g,n}, \delta_{g,n}) = \begin{bmatrix} m_{g,n}^\delta(\rho_{g,n}, \delta_{g,n}) \\ m_{g,n}^\rho(\rho_{g,n}, \delta_{g,n}) \end{bmatrix},
\]

where \( m_{g,n}^\delta(\rho_{g,n}, \delta_{g,n}) \) and \( m_{g,n}^\rho(\rho_{g,n}, \delta_{g,n}) \) are the vectors of linear and quadratic sample moments for the \( g \)-th equation as defined in (12) and (13). Let

\[
\Phi_{g\delta,n}^\delta = \sigma_{g\delta,n} [n^{-1} H_n H_n'] \quad \text{and} \quad \Phi_{g\rho,n}^{\rho\rho} = (\phi_{r\delta,gg,n})
\]

with \( \phi_{r\delta,gg,n} = (2n)^{-1} \sigma_{gg,n}^2 \text{tr} \left[ (A_{r,n} + A_{s,n}')(A_{s,n} + A_{s,n}') \right] \).

Then, \( E m_{g,n}(\rho_{g,n}, \delta_{g,n}) = 0 \) and

\[
VC(n^{1/2} m_{g,n}(\rho_{g,n}, \delta_{g,n})) = \begin{bmatrix} \Phi_{g\delta,n}^\delta & 0 \\ 0 & \Phi_{g\rho,n}^{\rho\rho} \end{bmatrix}.
\]

Let \( \hat{\Phi}_{g\delta,n}^\delta \) and \( \hat{\Phi}_{g\rho,n}^{\rho\rho} \) denote the corresponding estimators, where \( \sigma_{g\delta,n} \) is replaced by some consistent estimator \( \tilde{\sigma}_{g\delta,n} \). Then the one-step limited information GMM estimator is defined as

\[
(\hat{\delta}_{g,n}, \hat{\rho}_{g,n}) = \arg \min_{\bar{\delta}_{g,n}, \bar{\rho}_{g,n}} m_{g,n}(\bar{\rho}_{g,n}, \bar{\delta}_{g,n})' \begin{bmatrix} \Phi_{g\delta,n}^\delta & 0 \\ 0 & \Phi_{g\rho,n}^{\rho\rho} \end{bmatrix}^{-1} m_{g,n}(\bar{\rho}_{g,n}, \bar{\delta}_{g,n}).
\]

A simple adaptation of the methodology used to derive the limiting distribution of two-step estimators yields

\[
\begin{bmatrix} \hat{\delta}_{g,n} - \delta_{g,n} \\ \hat{\rho}_{g,n} - \rho_{g,n} \end{bmatrix} \sim AN \left( n^{-1} \begin{bmatrix} \tilde{\Omega}_{g\delta,n}^{\delta\delta} \\ 0 \\ 0 \end{bmatrix} \right).
\]

with

\[
\tilde{\Omega}_{g\delta,n}^{\delta\delta} = \tilde{\sigma}_{gg}[n^{-1} \tilde{Z}_{g,n}(\hat{\rho}_{g,n})' \tilde{Z}_{g,n}(\hat{\rho}_{g,n})]^{-1},
\]

\[
\hat{\Omega}_{g\rho,n}^{\rho\rho} = \begin{bmatrix} J_{g,n}(\hat{\rho}_{g,n}) (\tilde{\Phi}_{g\rho,n}^{\rho\rho})^{-1} J_{g,n}(\hat{\rho}_{g,n}) \end{bmatrix}^{-1}.
\]
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 the difference stems solely from terms in the latter expression involving estimates for \( \alpha_{g,r,n} \). They reflect that the two-step GMM estimator for \( \rho_{g,n} \) is based on estimated disturbances. If the disturbances would be observed, we would have \( \alpha_{g,r,n} = 0 \), and both one- and two-step estimators would have the same limiting distribution.

Towards defining our one-step full information GMM estimator consider the stacked sample moment vector

\[
\mathbf{m}(\delta_n, \rho_n) = \begin{bmatrix}
\mathbf{m}_{1,n}(\rho_{1,1}, \delta_{1,1}) \\
\vdots \\
\mathbf{m}_{G,n}(\rho_{G,1}, \delta_{G,1}) \\
\vdots \\
\mathbf{m}_{G,n}(\rho_{G,n}, \delta_{G,n})
\end{bmatrix}.
\]

Let

\[
\Phi_{\delta\delta} = \Sigma \otimes [\mathbf{n}^{-1}\mathbf{H}_n^t \mathbf{H}_n]
\]

and

\[
\Phi_{\rho\rho} = \begin{bmatrix}
\Phi_{11,n}^{pp} & \cdots & \Phi_{1G,n}^{pp} \\
\vdots & \ddots & \vdots \\
\Phi_{G1,n}^{pp} & \cdots & \Phi_{GG,n}^{pp}
\end{bmatrix},
\]

where the \((r, s)\)-th element of \( \Phi_{gh,n}^{pp} \) is given by

\[
\phi_{r,s,gh,n}^{pp} = \sigma_{gh,n}^2 (2n)^{-1/2} \text{tr} \left[ (\mathbf{A}_{r,n} + \mathbf{A}_{s,n}) (\mathbf{A}_{s,n} + \mathbf{A}_{r,n}) \right].
\]

Then, \( \mathbb{E}(\mathbf{m}(\delta_n, \rho_n)) = 0 \) and

\[
\text{VC}(\mathbf{m}(\rho_n, \delta_n)) = n^{-1} \begin{bmatrix}
\Phi_{\delta\delta} & 0 \\
0 & \Phi_{\rho\rho}
\end{bmatrix}.
\]

Let \( \hat{\Phi}_{\delta\delta} \) and \( \hat{\Phi}_{\rho\rho} \) denoted the corresponding estimators where \( \sigma_{gh,n} \) is replaced by some consistent estimator \( \hat{\sigma}_{gh,n} \). Then the one-step full information GMM estimator is defined as

\[
(\hat{\delta}_n, \hat{\rho}_n, \hat{\sigma}_n, \hat{\alpha}) = \arg \min_{\delta_n, \rho_n, \sigma_n, \alpha_n} \mathbb{E}(\mathbf{m}(\rho_n, \delta_n)) \left[ \begin{bmatrix}
\hat{\Phi}_{\delta\delta} & 0 \\
0 & \hat{\Phi}_{\rho\rho}
\end{bmatrix}^{-1} \mathbf{m}(\rho_n, \delta_n) \right].
\]

Again, a simple adaptation of the methodology used to derive the limiting distribution of two-step estimators yields

\[
\begin{bmatrix}
\hat{\delta}_n - \delta_n \\
\hat{\rho}_n - \rho_n
\end{bmatrix} \sim \text{AN} \left[ n^{-1} \begin{bmatrix}
\hat{\Omega}_{n,\delta\delta} & 0 \\
0 & \hat{\Omega}_{n,\rho\rho}
\end{bmatrix} \right],
\]

26
where
\[
\Omega_n = \left[ n^{-1} \tilde{Z}_{s,n} (\tilde{\rho}_n) (\Sigma_n^{-1} \otimes I_n) \tilde{Z}_{s,n} (\tilde{\rho}_n) \right]^{-1},
\]
\[
\Omega_{n,pp} = \text{diag}_{j=1}^{G} \left( \tilde{3}_{g,n} (\tilde{\Phi}_{g,n}^{-1}) \text{diag}_{j=1}^{G} \tilde{3}_{g,n} \right),
\]
with
\[
\tilde{3}_{g,n} = \left( \Phi_{gg,n}^{-1} \right)^{-1} J_{g,n}(\hat{\alpha}_g) \left[ J'_{g,n}(\hat{\alpha}_g) \left( \Phi_{gg,n}^{-1} \right)^{-1} J_{g,n}(\hat{\alpha}_g) \right]^+. 
\]

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 6 reveals again that the difference stems solely from terms in the latter expression involving estimates for \(\alpha_{g,r,n} \).

### 8 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-step and one-step estimators. An important aim in specifying our estimators was that the estimators remain feasible even for very large data sets and general weight matrices.

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 interaction in different activities; e.g., the level of different physical activities among groups of friends connected via an activity tracker such as Fitbit. Future research includes an extension of the methodology to panel data.
A Appendix: Preliminary Results

In this appendix we collect some preliminary results. All proofs are relegated to an auxiliary appendix, which will be made available on our web sites.

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 $A = (\alpha, \beta)$ be some $m \times m$ real matrix where the row sums of the absolute elements are bounded uniformly in $m$ by some finite constant. Let $\mu = (\mu_1, \ldots, \mu_m)$ and $\eta = (\eta_1, \ldots, \eta_m)$ be some $m \times 1$ random vectors with $\sup_n \max_{i=1}^m E|\mu_i|^p < \infty$ and $\sup_n \max_{i=1}^m E|\eta_i|^p < \infty$ for some $p > 1$, and let $x_n = (x_1, \ldots, x_m)' = \mu + A\eta$. Then $\sup_n \max_{i=1}^m E|x_i|^p < \infty$.

Lemma A.2: Suppose Assumptions 1-4 hold. Let $Z_n = [Y_n, X_n, \tilde{Y}_n]$, then

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

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

Lemma A.3: Suppose Assumptions 1-4 hold. Let $Z_n = [Y_n, X_n, \tilde{Y}_n]$ and let $A_n = (a_{ij,n})$ 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 $n^{-1}u_{h,n}A_nu_{g,n} = O_p(1)$, $n^{-1}Z_nA_nu_{g,n} = O_p(1)$ and $n^{-1}Z_nA_nZ_n = O_p(1)$ and furthermore

$$n^{-1}Z_nA_nu_{g,n} - n^{-1}EZ_nA_nu_{g,n} = o_p(1).$$

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

$$\tilde{\delta}_{g,n} = (Z_{g,n}'Z_{g,n})^{-1}Z_{g,n}'y_{g,n},$$

where $Z_{g,n} = P_{H_n}Z_{g,n}$ and $P_{H_n} = H_n(H_n'G_n)^{-1}H_n$. Then,

(a) $n^{1/2}(\tilde{\delta}_{g,n} - \delta_{g,n}) = n^{-1/2}T_{g,n}^r\varepsilon_{g,n} + o_p(1)$ with $T_{g,n} = F_{g,n}P_{g,n}$ and

where

$$P_{g,n} = Q_{HH}Q_{Hg,n}Q_{g,n}^{-1}Q_{Hg,n}^{-1},$$

$$F_{g,n} = (I_n - R_{g,n}^r)^{-1}H_n.$$
(b) \( n^{-1/2}T'_{gg,n} \varepsilon_{g,n} = O_p(1) \).
(c) \( P_{gg} \) is a finite matrix and \( \tilde{P}_{gg,n} - P_{gg} = o_p(1) \) for
\[
\tilde{P}_{gg,n} = (n^{-1}H_n'Z_n^{-1}H_n n^{-1}H_n'Z_n)_{gg}^{-1}.
\]
(d) \( \lambda_{\min}(n^{-1}T'_{gg,n} T_{gg,n}) \geq c \) for some \( c > 0 \) for all large \( n \).

**Lemma A.5** : Suppose Assumptions 1-4, 5 and 6 hold. Consider the GS2SLS estimator
\[
\hat{\delta}_g = [\tilde{Z}_{gg,n}(\hat{\rho}_g,n)'Z_{gg,n}(\hat{\rho}_g,n)]_{gg}^{-1}\tilde{Z}_{gg,n}(\hat{\rho}_g,n)'y_{gg,n}(\hat{\rho}_g,n),
\]
where \( \tilde{Z}_{gg,n}(\hat{\rho}_g,n) = P_{gg,n}Z_{gg,n}(\hat{\rho}_g,n) \), where \( \hat{\rho}_g,n \) is any consistent estimator for \( \rho_{gg,n} \). Then,
(a) \( n^{1/2}[\delta_{g,n} - \hat{\delta}_{g,n}] = n^{1/2}T'_{gg,n} \varepsilon_{g,n} + o_p(1) \) with \( T_{gg,n} = P_{gg,n}P_{gg,n} \) and where
\[
P_{gg} = Q_H^{-1}Q_{Hz,g*}(\rho_{gg,n})Q_{Hz,g*}^{-1}(\rho_{gg,n})Q_{Hz,g*}^{-1}(\rho_{gg,n})Q_{Hz,g*}^{-1}(\rho_{gg,n}),
\]
(b) \( n^{-1/2}T_{gg,n} \varepsilon_{g,n} = O_p(1) \).
(c) \( P_{gg,n} = O_p(1) \) and \( \tilde{P}_{gg,n} - P_{gg,n} = o_p(1) \) for
\[
\tilde{P}_{gg,n} = (n^{-1}H_n'Z_n^{-1}H_n n^{-1}H_n'Z_n)_{gg}^{-1}.
\]
(d) \( \lambda_{\min}(n^{-1}T'_{gg,n} T_{gg,n}) \geq c \) for some \( c > 0 \) for large \( n \).

**Lemma A.6** : Suppose Assumptions 1-4, 5-6 hold. Consider the GS3SLS estimator
\[
\hat{\delta}_n = [\tilde{Z}_{sn,n}(\hat{\rho}_n)'(\Sigma^{-1} \otimes I_n)Z_{sn}(\hat{\rho}_n)]_{nn}^{-1}\tilde{Z}_{sn,n}(\hat{\rho}_n)'(\Sigma^{-1} \otimes I_n)y_n(\hat{\rho}_n),
\]
where \( \tilde{Z}_{sn,n}(\hat{\rho}_n) = \text{diag}^G [\hat{Z}_{sg,n}(\hat{\rho}_g,n)] \) with \( \hat{Z}_{sg,n}(\hat{\rho}_g,n) = P_{H_g,n}Z_{sg,n}(\hat{\rho}_g,n) \), and \( \hat{\rho}_n = [\hat{\rho}_{1,n}, \ldots, \hat{\rho}_{G,n}]' \) is any consistent estimator for \( \rho_n \) and \( \Sigma_n \) is any consistent estimator for \( \Sigma \). Then
(a) \( n^{1/2}[\delta_n - \hat{\delta}_n] = n^{-1/2}T_{n} \varepsilon_n + o_p(1) \) with \( T_{n} = F_nP_{n} \), and where
\[
P_{n} = \text{diag} \left[ \text{diag} \left[ Q_{Hz,g*}(\rho_{gg,n}) \right] \right]_{nn}^{-1}.
\]

29
and
\[ P_n^{**} = I_G \otimes H_n. \]

(b) \( n^{-1/2} T_n^{**} \xi_n = O_p(1). \)
(c) \( P_n^{**} = O_p(1) \) and \( P_n^{**} - P_n^{**} = o_p(1) \) for
\[
\tilde{P}_n^{**} = \left[ \tilde{\Sigma}_n^{-1} \otimes (n^{-1}H_n')^{-1} \right] \text{diag} \left[ n^{-1}H_n'Z_{\gamma,n}(\hat{\rho}_n) \right]
\times \left[ n^{-1}Z_n(\hat{\rho}_n)(\tilde{\Sigma}_n^{-1} \otimes I_n)Z_n(\hat{\rho}_n) \right]^{-1}.
\]
(d) \( \lambda_{\min}(n^{-1}T_n^{**}T_n^{**}) \geq c \) 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 \( \times 1 \) disturbance vector \( u_n \) and corresponding forms based on a predictor \( \tilde{u}_n. \)

**Assumption A.1** For \( n \geq 1 \) the \( n \times 1 \) disturbance vector \( u_n \) is generated by
\[
\mathcal{R}_n u_n = \varepsilon_n,
\]
where \( \mathcal{R}_n \) is a non-stochastic nonsingular \( n \times n \) matrix, and the row and column sums of the absolute elements of \( \mathcal{R}_n \) and \( \mathcal{R}_n^{-1} \) are bounded uniformly by some finite constant, and the innovations \( \varepsilon_n = (\varepsilon_{n,1}, \ldots, \varepsilon_{n,n})' \) have the following properties: For each \( n \geq 1 \) the random variables \( \varepsilon_{n,1}, \ldots, \varepsilon_{n,n} \) are totally independent with \( E\varepsilon_{n,i} = 0, \) \( E(\varepsilon_{n,i}^2) = \sigma^2 > 0, \) and \( \sup_{1 \leq i \leq n, n \geq 1} E|\varepsilon_{n,i}|^{4+v} < \infty \) for some \( v > 0. \)

**Assumption A.2 :** The predictor \( \tilde{u}_n \) for \( u_n \) satisfies that
\[
\tilde{u}_n - u_n = \mathcal{D}_n \Delta_n,
\]
where \( \mathcal{D}_n = (\mathcal{D}_{ij,n}) \) is an \( n \times p_{\Delta} \) random matrix and \( \Delta_n \) is a \( p_{\Delta} \times 1 \) random vector. Furthermore \( \sup_n \sup_{1 \leq i \leq n, 1 \leq j \leq p_{\Delta}} E|\mathcal{D}_{ij,n}|^{2+\delta} < \infty \) for some \( \delta > 0, \) and \( n^{1/2} \|\Delta_n\| = O_p(1). \)

**Assumption A.3** For any \( n \times n \) real matrix \( A_n^{**}, \) whose row and column sums are bounded uniformly in absolute value,
\[
n^{-1}\mathcal{D}_n' A_n^{**} u_n - n^{-1} E\mathcal{D}_n' A_n^{**} u_n = o_p(1).\]

**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 \( u_n \) corresponds to the disturbance term of the \( g \)-th equation of the model defined by (1)-(5). Then the quantities considered in Assumptions A.1-A.3 should be interpreted as \( u_n = u_{g,n}, D_n = -Z_{g,n}, \)
\( R_n = I - \sum_{r \in L, r \neq g} \theta_{g,r,n} M_{r,n}, \) \( \epsilon_n = \epsilon_{g,n}, \) and \( \Delta_n = \delta_{g,n} - \delta_{g,n} \), where \( \delta_{g,n} \) is some estimator for the parameter vector \( \delta_{g,n} \). Observe that under Assumptions 1-4, and given \( \delta_{g,n} \) 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 \( A^*_n \) 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) \( n^{-1}E|u_n'A^*_nu_n| = O(1), \ var(n^{-1}u_n'A^*_nu_n) = o(1) \) and
\[
n^{-1}u_n'A^*_nu_n - n^{-1}E\left[u_n'A^*_nu_n\right] = o_p(1).
\]

(b) \( n^{-1}E|D_n'A^*_nu_n| = O(1), \) and
\[
n^{-1}D_n'A^*_nu_n - n^{-1}E\left[D_n'A^*_nu_n\right] = o_p(1).
\]

(c) If furthermore Assumption A.3 holds, then
\[
n^{-1/2}u_n'A^*_nu_n - n^{-1/2}u_n'A^*_nu_n + \alpha^*_n n^{1/2} \Delta_n + o_p(1)
\]
with \( \alpha^*_n = n^{-1}E[D_n'A^*_nu_n + A^*_nu_n]. \) (Of course, in light of (a) and (b) we have \( \alpha^*_n = O(1) \) and \( n^{-1}E[D_n'A^*_nu_n - A^*_nu_n] = o_p(1). \))

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 \( \Delta_n \) to the disturbance term of the model defined by (1)-(5). Then the quantities considered in Assumptions A.1-A.3 should be interpreted as \( u_n = u_{g,n}, D_n = -Z_{g,n}, \)
\( R_n = I - \sum_{r \in L, r \neq g} \theta_{g,r,n} M_{r,n}, \) \( \epsilon_n = \epsilon_{g,n}, \) and \( \Delta_n = \delta_{g,n} - \delta_{g,n} \), where \( \delta_{g,n} \) is some estimator for the parameter vector \( \delta_{g,n} \). Observe that under Assumptions 1-4, and given \( \delta_{g,n} \) a \( n^{1/2} \)-consistent estimator, these quantities clearly satisfy Assumptions A.1-A.3 in light of Lemmata A.1 and A.2.

**Assumption A.4**: (a) The vector of innovations \( \epsilon_n = [\epsilon_{g,1}, \ldots, \epsilon_{g,n}] \) satisfies Assumption 4. (b) The estimator \( \Delta_n \) is asymptotically linear in the sense that
\[
n^{1/2} \Delta_n = n^{-1/2} \sum_{h=1}^{G} T_{h,n}^\epsilon \epsilon_{h,n} + o_p(1),
\]
with \( T_{h,n} = F_{h,n} P_{h,n} \) where the \( F_{h,n} = (f_{hs,n}) \) are \( n \times p_F \) dimensional real nonstochastic matrices with \( \sup_h n^{-1} \sum_{s=1}^{n} |f_{hs,n}|^\eta < \infty \) with \( \eta > 2 \) for \( h = 1, \ldots, G, s = 1, \ldots, p_F, \) and \( P_{h,n} = (p_{hskt,n}) \) 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 \( \varepsilon_n = \varepsilon_{g,n} \). Furthermore, let \( A_n \) be an \( n \times n \) real matrix whose row and column sums are bounded uniformly in absolute value.

(a) Then

\[
\begin{align*}
\left( \rho_{g,n} \right) = & \frac{1}{2} \sum_{h=1}^{G} \left( a_{h,n} \varepsilon_{h,n} + o_p(1) \right), \\
\end{align*}
\]

where \( a_{h,n} = (a_{h1,n}, \ldots, a_{hn,n})' = T_{h,n} \alpha_n \) with \( \alpha_n = n^{-1} E \Omega' \Omega (A_n + A'_n) \Omega u_n \). Furthermore, \( \sup_{n} n^{-1} \sum_{i=1}^{n} |a_{hi,n}|^\eta < \infty \) for \( \eta > 2 \).

(b) If furthermore the diagonal elements of \( A_n \) are zero, then

\[
\begin{align*}
E \left[ n^{-1/2} \varepsilon_{g,n}' A_n \varepsilon_{g,n} + n^{-1/2} \sum_{h=1}^{G} a_{h,n}' a_{h,n} \right] &= 0, \\
\text{var} \left[ n^{-1/2} \varepsilon_{g,n}' A_n \varepsilon_{g,n} + n^{-1/2} \sum_{h=1}^{G} a_{h,n}' a_{h,n} \right] &= n^{-1} \sigma_{gg}^2 \text{tr} [A_n (A_n + A'_n)] + n^{-1} \sum_{h=1}^{G} \sum_{k=1}^{G} \sigma_{hk} a_{h,n}' a_{k,n}. \\
\end{align*}
\]

**B Appendix: Proofs for Section 5**

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

\[
\begin{align*}
R_n(\omega, \rho_g) &= m_{g,n}(\rho_g, \delta_{g,n})' \tilde{Y}_g n m_{g,n}(\rho_g, \delta_{g,n}), \\
\overline{R}_n(\rho_g) &= \left[ \Gamma_{g,n} \rho_{g,n}(\rho_g) - \gamma_g \right]' \tilde{Y}_g n \left[ \Gamma_{g,n} \rho_{g,n}(\rho_g) - \gamma_g \right], \\
\end{align*}
\]

where the quantities appearing in the above equation are defined before the theorem in the text. To prove the consistency of \( \rho_{g,n} \) 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 \( \rho_{g,n} \) is an identifiably unique sequence of minimizers of
\(\overline{R}_n\). Observe that \(\overline{R}_n(\overline{p}_g) \geq 0\) and that \(\overline{R}_n(\rho_{g,n}) = 0\), since \(\gamma_n = \Gamma_{g,n}r_{g,n}(\rho_{g,n})\) by (14). Utilizing Assumptions 8 and 9 we get

\[
\overline{R}_n(\overline{p}_g) - \overline{R}_n(\rho_{g,n}) = \overline{R}_n(\overline{p}_g) = \left[\Gamma_{g,n}r_{g,n}(\overline{p}_g) - \Gamma_{g,n}r_{g,n}(\rho_{g,n})\right]'\overline{\gamma}_{u_n} \left[\Gamma_{g,n}r_{g,n}(\overline{p}_g) - \Gamma_{g,n}r_{g,n}(\rho_{g,n})\right] = \left[r_{g,n}(\overline{p}_g) - r_{g,n}(\rho_{g,n})\right]'\Gamma_{g,n}r_{g,n}(\overline{p}_g) - r_{g,n}(\rho_{g,n})] \geq \lambda_{\min}(\overline{\gamma}_{u_n})\lambda_{\min}(\Gamma_{g,n}r_{g,n}) \left[r_{g,n}(\overline{p}_g) - r_{g,n}(\rho_{g,n})\right]'\left[r_{g,n}(\overline{p}_g) - r_{g,n}(\rho_{g,n})\right] \geq \lambda_* \sum_{s=1}^q [\overline{p}_{g,s} - \rho_{g,s,n}]^2
\]

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

\[
\inf_{\overline{p}_g} \sum_{r \in I_{g,z}} \|r_{g,r} - \rho_{g,r,n}\| \geq \varepsilon \Rightarrow \inf_{\overline{p}_g} \sum_{r \in I_{g,z}} \|\overline{p}_{g,r} - \rho_{g,r,n}\| \geq \lambda_* \varepsilon^2 > 0,
\]

which proves that \(\rho_{g,n}\) is identifiably unique. Next let \(\Phi_n = [\Gamma_n, -\gamma_n]\) and \(\tilde{\Phi}_n = [\Gamma_n, -\tilde{\gamma}_n]\), then

\[
|\overline{R}_n(\omega, \overline{p}_g) - \overline{R}_n(\overline{p}_g)| = \left[\overline{r}_{g,n}(\overline{p}_g)'\right] \left[\overline{\Phi}_n'\overline{\gamma}_{u_n} \Phi_n - \Phi_n'\overline{\gamma}_{u_n} \Phi_n\right] \left[\overline{r}_{g,n}(\overline{p}_g)'\right]' \leq \left[\overline{\Phi}_n'\overline{\gamma}_{u_n} \Phi_n - \Phi_n'\overline{\gamma}_{u_n} \Phi_n\right] \left[\overline{r}_{g,n}(\overline{p}_g)'\right]' \leq \lambda_* \sum_{r \in I_{g,z}} [\overline{p}_{g,r} - \rho_{g,r,n}]^2 \leq \lambda_* \varepsilon^2 > 0.
\]

Next observe that the elements of \(\Phi_n\) and \(\tilde{\Phi}_n\) are all of the form \(n^{-1}Eu_{n}'A_nu_n\) and \(n^{-1}\tilde{u}_{n}'A_n\tilde{u}_n\), where the row and column sums of \(A_n\) are bounded uniformly in absolute value. Recall that \(u_{g,n} = \left[\sum_{r \in I_{g,z}} \rho_{g,r,n}M_{r,n}\right] u_{g,n} + \varepsilon_{g,n}\) and observe that \(\tilde{u}_{g,n} - u_{g,n} = -Z_{g,n}(\tilde{\delta}_{g,n} - \delta_{g,n})\). By Assumption 1 and 2 the row and column sums of the absolute elements of \(\sum_{r \in I_{g,z}} \rho_{g,r,n}M_{r,n}\) and \(\sum_{r \in I_{g,z}} \rho_{g,r,n}M_{r,n}^{-1}\) are bounded in absolute value by some finite constant. By Assumptions 4 the elements of \(\varepsilon_{g,n}\) are totally independent with zero mean, positive variance and finite \(4 + v\) absolute moments for some \(v > 0\). Furthermore, observe that by Lemma A.2 the fourth absolute moments of the elements of \(Z_{g,n}\) are uniformly bounded by a finite constant and \(n^{1/2}(\tilde{\delta}_{g,n} - \delta_{g,n}) = O_p(1)\). It now follows immediately from Lemma A.7(a) that \(\Phi_n - \Phi_n \overset{p}{\rightarrow} 0\), that the elements of \(\Phi_n\) are \(O(1)\) and, consequently, that the elements of \(\Phi_n\) are \(O_p(1)\). The elements of \(\tilde{\Phi}_n\) and \(\Phi_n\) have the analogous properties in light of condition (b) in the theorem. Given this it follows from the above inequality that
\( R_n(\omega, \overline{\rho}_g) - \overline{R}_n(\overline{\rho}_g) \) converges to zero uniformly over the optimization space, i.e.,

\[
\sup_{\overline{\rho}_g \sum_{r \in I_{\rho,g}} |\overline{\rho}_g| \in [-a, a]} \left| R_n(\omega, \overline{\rho}_g) - \overline{R}_n(\overline{\rho}_g) \right| \\
\leq \left\| \Phi_n \overline{\Phi}_n - \Phi_n \Phi_n \right\| \left\{ 1 + \frac{1 + q(q - 1)/2}{n^2} \right\} \to 0
\]
as \( n \to \infty \). The consistency of \( \tilde{\rho}_{g,n} \) 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 \( \tilde{\rho}_{g,n} \) defined in (16) is consistent. Apart on a set, whose probability tends to zero, the estimator satisfies the following first-order condition

\[
m_{g,n}(\tilde{\rho}_{g,n}, \overline{\delta}_{g,n}) \frac{\partial m_{g,n}(\tilde{\rho}_{g,n}, \overline{\delta}_{g,n})}{\partial \rho_g} = 0.
\]

Substituting the mean value theorem expression

\[
m_{g,n}(\tilde{\rho}_{g,n}, \overline{\delta}_{g,n}) = m_{g,n}(\rho_{g,n}, \overline{\delta}_{g,n}) + \frac{\partial m_{g,n}(\tilde{\rho}_{g,n}, \overline{\delta}_{g,n})}{\partial \rho_g} (\tilde{\rho}_{g,n} - \rho_{g,n}),
\]

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

\[
\frac{\partial m_{g,n}(\tilde{\rho}_{g,n}, \overline{\delta}_{g,n})}{\partial \rho_g} \tilde{\phi}_{g,n} = \frac{\partial m_{g,n}(\tilde{\rho}_{g,n}, \overline{\delta}_{g,n})}{\partial \rho_g} (\overline{\delta}_{g,n} - \rho_{g,n})
\]

(B.1)

Observe that

\[
\frac{\partial m_{g,n}(\rho_{g}, \overline{\delta}_{g,n})}{\partial \rho_g} = -\Gamma_{g,n} \frac{\partial \Gamma_{g,n}(\rho_{g})}{\partial \rho_g}
\]

(B.2)

and consider

\[
\Xi_{g,n} = \frac{\partial m_{g,n}(\tilde{\rho}_{g,n}, \overline{\delta}_{g,n})}{\partial \rho_g} \tilde{\phi}_{g,n} = \frac{\partial \Xi_{g,n}(\tilde{\rho}_{g,n}, \overline{\delta}_{g,n})}{\partial \rho_g}
\]

(B.3)

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

\[
\Xi_{g,n} - \Xi_{g,n} \overset{P}{\to} 0
\]

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

\[
\lambda_{\min}(\Xi_{g,n}) \geq \lambda_{\min}(Y_{g,n}) \lambda_{\min}(\Gamma_{g,n}'\Gamma_{g,n}) \lambda_{\min} \left\{ \frac{\partial r_{g,n}(\rho_{g,n})}{\partial \rho_g} \frac{\partial r_{g,n}(\rho_{g,n})}{\partial \rho_g^*} \right\} \geq \lambda_\Xi
\]

for some \( \lambda_\Xi > 0 \), observing that

\[
\lambda_{\min} \left\{ \frac{\partial r_{g,n}(\rho_{g,n})}{\partial \rho_g} \frac{\partial r_{g,n}(\rho_{g,n})}{\partial \rho_g^*} \right\} = \lambda_{\min} \left\{ I_{g} + \text{semi-positive matrix} \right\} \geq 1.
\]

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

\[
\Xi_{g,n}^{+} - \Xi_{g,n}^{-1} \overset{P}{\to} 0
\]

as \( n \to \infty \).

 Premultiplying (B.1) by \( \Xi_{g,n}^{+} \) and rearranging terms yields

\[
n^{1/2}(\bar{\rho}_{g,n} - \rho_{g,n}) = \left[ I - \Xi_{g,n}^{+} \Xi_{g,n}^{-1} \right] n^{1/2}(\bar{\rho}_{g,n} - \rho_{g,n})
\]

\[
- \Xi_{g,n}^{+} \frac{\partial m_{g,n}(\bar{\rho}_{g,n}, \bar{\delta}_{g,n})}{\partial \bar{\rho}_g} \bar{Y}_{g,n} n^{1/2} m_{g,n}(\rho_{g,n}, \bar{\delta}_{g,n}).
\]

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

\[
n^{1/2}(\bar{\rho}_{g,n} - \rho_{g,n}) = - \Xi_{g,n}^{+} \frac{\partial m_{g,n}(\bar{\rho}_{g,n}, \bar{\delta}_{g,n})}{\partial \bar{\rho}_g} \bar{Y}_{g,n} n^{1/2} m_{g,n}(\rho_{g,n}, \bar{\delta}_{g,n}) + o_p(1).
\]

(B.6)

Observe that

\[
\Xi_{g,n}^{+} \frac{\partial m_{g,n}(\bar{\rho}_{g,n}, \bar{\delta}_{g,n})}{\partial \bar{\rho}_g} \bar{Y}_{g,n} - \Xi_{g,n}^{-1} \frac{\partial r_{g,n}(\rho_{g,n})}{\partial \rho_g} \Gamma_{g,n}' Y_{g,n} = o_p(1).
\]

(B.7)

In light of (15) the elements of \( n^{1/2} m_{g,n}(\rho_{g,n}, \bar{\delta}_{g,n}) \) are of the form \( (s = 1, \ldots, S) \)

\[
n^{-1/2} \bar{u}_{g,n} \left[ I_n - R_{g,n}^s(\rho_{g,n}) \right] A_s, n \left[ I_n - R_{g,n}^s(\rho_{g,n}) \right] \bar{u}_{g,n}
\]

35
where $\alpha$

Recall that if we de

light of Lemmata A.1-A.3. Hence it follows from Lemma A.8 that Kelejian and Prucha (2010) that are zero and that the VC matrix of

forms on the r.h.s. of (B.8), then observing that the diagonal elements of $D$

We note that $\varepsilon$

the vectors of the linear forms, and — in light of Assumption 4 — the innovations $\xi$

the lemma implies that for some $\eta > 2$ the sample moments of the absolute elements $T_{h,n}\alpha_{g,s,n}$ are uniformly bounded. We now have

Let $\Psi_{g,g,n}^{\rho} = (\psi_{r,s,g,n}^{\rho})$ 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 $A_{s,n}$ are zero and that the VC matrix of $\varepsilon$ is $\Sigma \otimes I_n$, it follows from Lemma A.1 in Kelejian and Prucha (2010) that

We note that $\lambda_{\min}(\Psi_{g,g,n}^{\rho}) \geq \text{const} > 0$ by assumption. Since the matrices $A_{s,n}$, the vectors of the linear forms, and — in light of Assumption 4 — the innovations $\varepsilon$, 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

\[ \xi_{g,n} = -\Psi_{g,g,n}^{\rho} - 1/2 n^{-1/2} \begin{bmatrix} 1/2 \varepsilon_{g,n}(A_{1,n} + A_{1,n}') \varepsilon_{g,n} + \alpha_{g,1,n}' \sum_{h=1}^{G} T_{gh,n} \varepsilon_{h,n} \\ \vdots \\ 1/2 \varepsilon_{g,n}(A_{S,n} + A_{S,n}') \varepsilon_{g,n} + \alpha_{g,S,n}' \sum_{h=1}^{G} T_{gh,n} \varepsilon_{h,n} \end{bmatrix} \xrightarrow{d} N(0, I_S). \]
Next observe that \( \Psi_{g,n}^{gg} = O(1) \), and hence \( (\Psi_{g,n}^{gg})^{1/2} = O(1) \), since the row and column sums of the absolute elements of \( A_{r,n} \) are uniformly bounded by assumption, and since in light of the above assumptions the terms \( n^{-1} \alpha_{g,r,n} T_{r,n} T_{g,n} \alpha_{g,s,n} \) are uniformly bounded in absolute value. It now follows from (B.6), (B.7) and (B.10) that

\[
n^{1/2}(\bar{\rho}_{g,n} - \rho_{g,n}) = [J'_{g,n} Y_{g,n} J_{g,n}]^{-1} J'_{g,n} Y_{g,n} (\Psi_{g,n}^{gg})^{1/2} \xi_{g,n} + o_{p}(1), \tag{B.11}
\]

observing that \( \Xi_{g,n} = J'_{g,n} Y_{g,n} J_{g,n} \). This establishes (19). Since all of the nonstochastic terms on the r.h.s. of (B.11) are \( O(1) \), it follows that \( n^{1/2}(\bar{\rho}_{g,n} - \rho_{g,n}) = O_{p}(1) \). Next recall that \( 0 < \lambda_{\Xi}^{\ast} \leq \lambda_{\min}(\Xi_{g,n}) \leq \lambda_{\max}(\Xi_{g,n}) \leq \lambda_{\Xi}^{\ast} < \infty \) and observe that \( \lambda_{\min}(\Xi_{g,n}^{-1}) = 1/\lambda_{\max}(\Xi_{g,n}) \). Hence

\[
\lambda_{\min}(\Xi_{g,n}^{-1} J'_{g,n} Y_{g,n} \Psi_{g,n}^{gg} Y_{g,n} J_{g,n} \Xi_{g,n}^{-1}) \\
\geq \lambda_{\min}(\Psi_{g,n}^{gg}) \lambda_{\min}(Y_{g,n}) \lambda_{\min}(\Xi_{g,n}^{-1} J'_{g,n} Y_{g,n} J_{g,n} \Xi_{g,n}^{-1}) \\
\geq \lambda_{\min}(\Psi_{g,n}^{gg}) \lambda_{\min}(Y_{g,n}) / \lambda_{\Xi}^{\ast} \geq \text{const} > 0.
\]

This establishes the last claim of the theorem.

**Proof of Lemma 1:** Observe that \( \bar{u}_{g,n} = u_{g,n} - Z_{g,n} \Delta_{g,n} \) with \( \Delta_{g,n} = \tilde{\delta}_{g,n} - \delta_{g,n} \), and thus

\[
\tilde{\varepsilon}_{g,n} = \left[ I_{n} - R_{g,n}^{*}(\bar{\rho}_{g,n}) \right] \bar{u}_{g,n} \\
= \varepsilon_{g,n} - R_{g,n}^{*}(\bar{\rho}_{g,n} - \rho_{g,n}) u_{g,n} - \left[ I_{n} - R_{g,n}^{*}(\bar{\rho}_{g,n}) \right] Z_{g,n} \Delta_{g,n}.
\]

Consequently

\[
\bar{\sigma}_{gh,n} = n^{-1} \tilde{\varepsilon}_{g,n} \tilde{\varepsilon}_{h,n} = n^{-1} \varepsilon_{g,n} \varepsilon_{h,n} \\
+ n^{-1} u'_{g,n} \left[ R_{g,n}^{*}(\bar{\rho}_{g,n} - \rho_{g,n}) \right]' R_{h,n}^{*}(\bar{\rho}_{h,n} - \rho_{h,n}) u_{h,n} \\
+ n^{-1} \bar{\varepsilon}'_{g,n} \left[ I_{n} - R_{g,n}^{*}(\bar{\rho}_{g,n}) \right]' \left[ I_{n} - R_{h,n}^{*}(\bar{\rho}_{h,n}) \right] Z_{h,n} \Delta_{h,n} \\
- n^{-1} \varepsilon'_{g,n} R_{h,n}^{*}(\bar{\rho}_{h,n} - \rho_{h,n}) u_{h,n} \\
- n^{-1} \bar{\varepsilon}'_{g,n} R_{g,n}^{*}(\bar{\rho}_{g,n} - \rho_{g,n}) u_{g,n} \\
+ n^{-1} \bar{u}'_{g,n} [ R_{g,n}^{*}(\bar{\rho}_{g,n} - \rho_{g,n}) ]' [ I_{n} - R_{h,n}^{*}(\bar{\rho}_{h,n}) ] Z_{h,n} \Delta_{h,n} \\
+ n^{-1} u'_{g,n} [ R_{h,n}^{*}(\bar{\rho}_{h,n} - \rho_{h,n}) ]' [ I_{n} - R_{g,n}^{*}(\bar{\rho}_{g,n}) ] Z_{g,n} \Delta_{g,n}.
\]

By Assumption 4 we have \( \sigma_{gh} = \sum_{l=1}^{G} \sigma_{s_{g}^{l} s_{h}^{l}} \) and \( \varepsilon_{q} = \sum_{l=1}^{G} \sigma_{s_{g}^{l} v_{l}} \). Since the elements of \( v \) are i.i.d. \( (0,1) \), it follows that \( n^{-1} v'_{l} v_{l} = 1 + o_{p}(1) \) and \( n^{-1} v'_{k} v_{k} = o_{p}(1) \) for \( l \neq k \). Hence

\[
n^{-1} \varepsilon_{g,n}^{'} \varepsilon_{h,n} = \sum_{l=1}^{G} \sum_{k=1}^{G} \sigma_{s_{g}^{l} s_{h}^{k}} n^{-1} v'_{l} v_{k} = \sum_{l=1}^{G} \sigma_{s_{g}^{l} s_{h}^{l}} + o_{p}(1) = \sigma_{gh} + o_{p}(1).
\]

Next observe that, taking into account that \( \varepsilon_{h,n} = \left[ I_{n} - R_{g,n}^{*}(\rho_{g,n}) \right] u_{g,n} \), all the other terms consist of expressions of the form \( o_{p}(1)n^{-1} u'_{g,n} A_{n} u_{h,n} \),
o_p(1)n^{-1}g_n'A_nZ_{b,h,n}^2 and o_p(1)n^{-1}Z_{g,n}'A_nZ_{b,h,n}, where 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 \( \bar{\sigma}_{gh,n} - \sigma_{gh} = o_p(1) \) as claimed.

**Proof of Theorem 3:** We first demonstrate that \( \Psi_{gg,n}^{pp} - \Psi_{gg,n}^{pp} = o_p(1) \), where the elements of \( \Psi_{gg,n}^{pp} \) and \( \Psi_{gg,n}^{pp} \) are defined in (18) and (24). By Lemma 1 we have \( \bar{\sigma}_{gh,n} - \sigma_{gh} = o_p(1) \). Furthermore, by assumption \( n^{-1}T_{gh,n}'T_{gh,n} - n^{-1}T_{gh,n}'T_{gh,n} = o_p(1) \), where \( n^{-1}T_{gh,n}'T_{gh,n} = O(1) \) in light of Assumption 10. Next, observe that under the maintained assumption the row and column sums of the absolute elements of the matrices \( A_{r,n} \) and \( A_{s,n} \) are uniformly bounded, and thus clearly are those of the matrices \( A_n = (a_{ij,n}) \) with \( a_{ij,n} = (a_{ij,n} + a_{ji,n})/(a_{ij,n} + a_{ji,n}) \). It then follows directly from Lemma A.7 that \( \bar{\alpha}_{g,r,n} - \alpha_{g,r,n} = o_p(1) \), where \( \bar{\alpha}_{g,r,n} = O(1) \) and \( \alpha_{g,r,n} = O_p(1) \). Hence, clearly \( \psi_{rs,gg,n}^{pp} - \psi_{rs,gg,n}^{pp} = o_p(1) \), \( \psi_{rs,gg,n}^{pp} = O(1) \), and \( \psi_{rs,gg,n}^{pp} = O_p(1) \), as well as \( \Psi_{gg,n}^{pp} - \Psi_{gg,n}^{pp} = o_p(1) \), \( \Psi_{gg,n}^{pp} = O(1) \) and \( \Psi_{gg,n}^{pp} = O_p(1) \). Observing that \( \lambda_{\min}(\Psi_{gg,n}^{pp}) \geq c_\Psi > 0 \) it follows further that \( (\Psi_{gg,n}^{pp})^{-1} - (\Psi_{gg,n}^{pp})^{-1} = o_p(1) \), \( (\Psi_{gg,n}^{pp})^{-1} = O(1) \) and \( (\Psi_{gg,n}^{pp})^{-1} = O_p(1) \).

By Assumption 9 \( \Upsilon_{g,n} - \Upsilon_{g,n} = o_p(1) \), \( \Upsilon_{g,n} = O(1) \) and thus \( \bar{\Upsilon}_{g,n} = O_p(1) \). In proving Theorem 2 we have verified that \( \bar{J}_{g,n} - J_{g,n} \sim \mathcal{N}(0,1) \) and \( J_{g,n} = O_p(1) \), and furthermore that \( \bar{J}_{g,n} J_{g,n}^{-1} = O_p(1) \) and \( (\bar{J}_{g,n} J_{g,n}^{-1}) = O(1) \). The claim that \( \Omega_{gg,n}^{pp} - \Omega_{gg,n}^{pp} = o_p(1) \) and \( \Omega_{gg,n}^{pp} = O(1) \) is now obvious. Observing that by Theorem 2 \( \lambda_{\min}(\Omega_{gg,n}^{pp}(\Upsilon_{g,n})) \geq \text{const} > 0 \) it follows further that \( (\Omega_{gg,n}^{pp})^{-1} - (\Omega_{gg,n}^{pp})^{-1} = o_p(1) \), \( (\Omega_{gg,n}^{pp})^{-1} = O(1) \) and \( (\Omega_{gg,n}^{pp})^{-1} = O_p(1) \).

We shall utilize the following lemma.

**Lemma B.1:** Suppose the nG × 1 vector of innovations \( \epsilon_n \) is generated as postulated in Assumption 4. Let \( A = \text{diag}_{g=1}^G(A_{gg}) \), and \( B = \text{diag}_{g=1}^G(B_{gg}) \) be symmetric nG × nG matrices with zero diagonal elements, and let \( a = [a_1, \ldots, a_G]' \) and \( b = [b_1, \ldots, b_G]' \) be nG × 1 nonstochastic vectors. Then

\[
E(\epsilon' A \epsilon + a' \epsilon) = 0
\]

\[
cov(\epsilon' A \epsilon + a' \epsilon, \epsilon' B \epsilon + b' \epsilon) = 2tr[A(\Sigma \otimes I)B(\Sigma \otimes I)] + a'(\Sigma \otimes I)b
\]

\[
= 2 \sum_{h=1}^G \sum_{l=1}^G \sigma_{hl}^2 tr[A_{hh}B_{ll}] + \sum_{h=1}^G \sum_{l=1}^G \sigma_{hl}a_h'b_l.
\]

**Proof.** The result follows immediately from Lemma A1 in Kelejian and Prucha (2010), observing that the diagonal elements of \( A = (\Sigma \otimes I)A(\Sigma \otimes I) \) and \( B = (\Sigma \otimes I)B(\Sigma \otimes I) \) are zero.
Proof of Theorem 4: Observe that \( \eta_n = n^{-1/2} T'_n \varepsilon_n \) and thus clearly \( \Psi_n^{\delta \delta} = E \eta_n' \eta_n = n^{-1} T'_n E \varepsilon_n \varepsilon_n' T_n = n^{-1} T'_n (\Sigma \otimes I_n) T_n \) as claimed.

Next observe that \( \xi_{g,s,n} = n^{-1/2} [\varepsilon'_n B \varepsilon_n + b' \varepsilon_n] \) with \( B = \text{diag}(0, \ldots, B_{gg}, \ldots, 0) \), where \( B_{gg} = (A_{s,n} + A'_{s,n})/2 \) and \( b = T_{g,n} \alpha_{g,s,n} \) with \( T_{g,n} = [T'_{g1,n}, \ldots, T'_{gG,n}]' \).

It now follows from Lemma B.1 that

\[
\text{cov}(\eta_n, \xi_{g,s,n}) = \text{cov}(n^{-1} T'_n \varepsilon_n, n^{-1/2} [\varepsilon'_n B \varepsilon_n + b' \varepsilon_n]) = \text{cov}(n^{-1} \eta_n, n^{-1/2} b' \varepsilon_n)
\]

and thus

\[
\text{cov}(\eta_n, \xi_{g,s,n}) = n^{-1} T'_n (\Sigma \otimes I_n) T_{g,n} \alpha_{g,s,n}
\]

and

\[
\Psi_n^{\alpha \alpha} = \text{cov}(\eta_n, \xi_n) = n^{-1} T'_n (\Sigma \otimes I_n) T_n \text{diag}_{g=1} \alpha_{g,1,n}, \ldots, \alpha_{g,S,n}
\]

as claimed.

Define \( \xi_{g,r,n} = n^{-1/2} [\varepsilon'_n A \varepsilon_n + a' \varepsilon_n] \) with \( A = \text{diag}(0, \ldots, A_{gg}, \ldots, 0) \) where \( A_{gg} = (A_{r,n} + A'_{r,n})/2 \) and \( a = T_{g,n} \alpha_{g,r,n} \). Then applying Lemma B.1 we see that

\[
\text{cov}(\xi_{g,r,n}, \xi_{h,s,n}) = \sigma^2_{g,h,n}(2n)^{-1} \text{tr} \left( \left( A_{r,n} + A'_{r,n} \right) \left( A_{s,n} + A'_{s,n} \right) \right)
\]

\[
+ \alpha_{g,r,n} \sum_{u=1}^{G} \sum_{v=1}^{G} \sigma_{u,v,n} T'_{g,u,n} T_{h,v,n} \alpha_{h,s,n}
\]

which verifies the expressions for the elements of \( \Psi_n = \text{cov}(\xi_n, \xi_n) \).

In light of Assumption 10 and Theorem 2 we have

\[
n^{1/2} \left[ \begin{array}{c} \delta_n - \delta_n \\ \rho_n - \rho_n \end{array} \right] = \left[ \begin{array}{cc} 1 & 0 \\ 0 & \text{diag}_{g}\left[ \left( J'_{g,n} Y_{g,n} J_{g,n} \right)^{-1} J'_{g,n} Y_{g,n} \right] \end{array} \right] \left[ \begin{array}{c} \eta_n \\ \xi_n \end{array} \right] + o_p(1)
\]

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

\[
\Psi_n^{-1/2} \left[ \begin{array}{c} \eta_n \\ \xi_n \end{array} \right] \xrightarrow{d} N(0, I_d),
\]

which proves the first part of the theorem.

Note that \( \lambda_{\min}(\Psi_{gg,n}^{pp}) \geq \lambda_{\min}(\Psi_{gg,n}^{\delta \delta}) \geq \lambda_{\min}(\Psi_{n}^{\delta \delta}) \geq \text{const} > 0 \). In proving Theorem 3 we have shown that \( \Psi_{gg,n}^{pp} - \Psi_{gg,n}^{\delta \delta} = o_p(1), \Psi_{gg,n}^{pp} = O(1) \) and \( \Psi_{gg,n}^{\delta \delta} = O_p(1) \). The proof that \( \Psi_{gg,n}^{pp} - \Psi_{gg,n}^{\delta \delta} = o_p(1), \Psi_{gg,n}^{pp} = O(1) \) and \( \Psi_{gg,n}^{\delta \delta} = O_p(1) \) is analogous. Thus \( \Psi_{gg,n}^{pp} - \Psi_{gg,n}^{\delta \delta} = o_p(1), \Psi_{gg,n}^{pp} = O(1) \) and \( \Psi_{gg,n}^{\delta \delta} = O_p(1) \). Observing that \( \lambda_{\min}(\Psi_{n}^{pp}) \geq \text{const} > 0 \) it follows further that \( (\Psi_n^{pp})^{-1} \geq (\Psi_n^{\delta \delta})^{-1} = o_p(1), (\Psi_n^{pp})^{-1} = O(1) \) and \( (\Psi_n^{\delta \delta})^{-1} = O_p(1) \).
Next recall that in proving Theorem 3 we demonstrated that $\tilde{\alpha}_{g.r,n} - \alpha_{g,r,n} = o_p(1)$, $\alpha_{g,r,n} = O(1)$ and $\tilde{\alpha}_{g.r,n} = O_p(1)$, and furthermore that $\sigma_{gh,n} - \sigma_{gh} = o_p(1)$. Since $n^{-1}T_{gh,n}T_{kl,n} - n^{-1}T_{gh,n}T_{kl,n} = o_p(1)$ it follows that $\Psi_n^{\delta \delta} - \Psi_n^{\delta \delta} = o_p(1)$ and $\ddot{\Psi}_n - \Psi_n = o_p(1)$. Also observe that in light of Assumption 10 we have $\Psi_n^{\delta \delta} = O(1)$ and $\Psi_n^{\delta \delta} = O(1)$. This and the above results imply that $\ddot{\Psi}_n - \Psi_n = o_p(1)$, $\ddot{\Psi}_n = O(1)$ and $\ddot{\Psi}_n = O_p(1)$. Recalling that $\lambda_{\min}(\Psi_n) \geq const > 0$, it follows further that $\ddot{\Psi}_n^{-1} = \Psi_n^{-1} = o_p(1)$ and $\ddot{\Psi}_n^{-1} = O_p(1)$.

By Assumption 9, $\ddot{\Psi}_g,n - \Psi_g,n = o_p(1), \Psi_g,n = O(1)$ and thus $\ddot{\Psi}_g,n = O_p(1)$. Also recall from the proof of Theorem 3 that $J_g,n - J_g,n = o_p(1), J_g,n = O(1)$ and $J_g,n = O_p(1)$, and furthermore that $(J_g,n)^{-1} - (J_g,n)^{-1} = o_p(1)$, $(J_g,n)^{-1} = O_p(1)$ and $(J_g,n)^{-1} = O(1)$. The claim that $\Omega_n - \Omega_n = o_p(1)$ and $\Omega_n = O(1)$ is now obvious. Observing that

$$\lambda_{\min}(\Omega_n) \geq \lambda_{\min}(\Psi_n)\lambda_{\min}\{diag((J_g,n)^{-1})\} \geq const > 0$$

utilizing that $\lambda_{\min}(J_g,n) \geq const > 0$ as demonstrated in the proof of Theorem 2, it follows further that $\ddot{\Omega}_n^{-1} - \Omega_n^{-1} = o_p(1)$, $\ddot{\Omega}_n^{-1} = O(1)$ and $\ddot{\Omega}_n^{-1} = O_p(1)$.

\section{Appendix: Proofs for Section 6}

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

\textbf{Proof.} To prove the claim we verify the assumptions of Theorem 1. Assumptions 1-8 are maintained. Assumption 9 holds trivially with $\ddot{\Psi}_g,n = \Psi_g,n = I$. Observing furthermore that by Lemma A.4 we have $n^{1/2}(\ddot{\delta}_g,n - \delta_g,n) = O_p(1)$ completes the proof.

\textbf{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 $\ddot{\delta}_g,n = (\ddot{\delta}_{1,n}, \ldots, \ddot{\delta}_{G,n})^\prime$ and $\ddot{\rho}_n = (\ddot{\rho}_{1,n}, \ldots, \ddot{\rho}_{G,n})^\prime$. The limiting distribution of $\ddot{\delta}_g,n$ and $\ddot{\rho}_g,n$ is then obtained as a trivial specialization. For clarity we divide the remainder of the proof into several parts.

\textbf{Part 1:} (Verification of Assumption 10 for $\ddot{\delta}_g,n$) In light of Lemma A.5 we have $n^{1/2}[\ddot{\delta}_g,n - \delta_g,n] - n^{-1/2}T_{gg,n} + o_p(1)$ with $T_{gg,n} = H_nP_{gg,n}$ with $P_{gg,n} = Q_{HH}^{-1}Q_{Hz,g}^{*}(\rho_{g,n})Q_{Hz,g}^{*}(\rho_{g,n})Q_{HH}^{-1}Q_{Hz,g}^{*}(\rho_{g,n})^{-1}$.
and thus

\[ n^{-1}T'_{gg,n} T_{hh,n} = [Q_{HH}(\rho_{g,n})Q_{HH}^{-1}Q_{HH}(\rho_{g,n})]^{-1}Q_{HH}(\rho_{g,n})Q_{HH}^{-1} \times (n^{-1}H'_n H_n)^{-1}Q_{HH}^{-1}Q_{HH}(\rho_{h,n})[Q_{HH}(\rho_{h,n})Q_{HH}^{-1}Q_{HH}(\rho_{h,n})]^{-1}. \]

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

**Part 2:** (Specialized Expressions for \( \Psi_n \) and the Corresponding Estimator)

Next observe that the components of \( \Psi_n \) defined in (26) simplified in obvious ways in that \( T_n = \text{diag}(T_{gg,n}) \), and thus all terms involving a \( T_{gh,n} \) with \( g \neq h \) are zero. In particular, the \((g, h)\)-th blocks of \( \Psi_{g,h}^{\delta} \) and \( \Psi_{g,h}^{\rho} \) are given by

\[
\Psi_{g,h}^{\delta} = \sigma_{gh,n}^{-1}T'_{gg,n} T_{hh,n},
\]

\[
\Psi_{g,h}^{\rho} = \sigma_{gh,n}^{-1}T'_{gg,n} T_{hh,n} \alpha_{h,1,n}, \ldots, \alpha_{h,S,n},
\]

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

\[
\psi_{rs,gh,n} = \sigma_{gh,n}^{-1}T'_{gg,n} T_{hh,n} \alpha_{h,1,n}, \ldots, \alpha_{h,S,n}.
\]

Now consider the estimator

\[
\hat{T}_{gg,n} = H_n(n^{-1}H'_n H_n)^{-1}(n^{-1}H'_n Z_{s,n}(\tilde{\rho}_{g,n})) \times \]

\[
[ (n^{-1}Z'_{s,n}(\tilde{\rho}_{g,n}) H_n)(n^{-1}H'_n H_n)^{-1}(n^{-1}H'_n Z_{s,n}(\tilde{\rho}_{g,n}))]^{-1},
\]

where \( \tilde{\rho}_{g,n} \) denotes the first-stage estimator for \( \rho_{g,n} \), let

\[
\tilde{\alpha}_{g,r,n} = -n^{-1}(Z'_{g,n}(I_n - R'_{g,n}(\tilde{\rho}_{g,n}))(A_{r,n} + A'_{r,n})(I_n - R'_{g,n}(\tilde{\rho}_{g,n}))\tilde{u}_{g,n},
\]

with \( \tilde{u}_{g,n} = y_{g,n} - Z_{g,n}\tilde{g}_{g,n} \), and let \( \tilde{\sigma}_{g,n} = n^{-1}Z'_{g,n}(\tilde{\rho}_{g,n})\tilde{\alpha}_{g,n} \) then the components of the estimator for \( \Psi_n \) defined in (27) simplify to

\[
\hat{\Psi}_{g,h,n}^{\delta} = \tilde{\sigma}_{gh,n}^{-1}T'_{gg,n} T_{hh,n},
\]

\[
\hat{\Psi}_{g,h,n}^{\rho} = \tilde{\sigma}_{gh,n}^{-1}T'_{gg,n} T_{hh,n} \tilde{\alpha}_{h,1,n}, \ldots, \tilde{\alpha}_{h,S,n},
\]

and the elements of the estimator \( \hat{\Psi}_{g,h,n}^{\rho} \) of \( \Psi_{g,h}^{\rho} \) are given by

\[
\hat{\psi}_{rs,gh,n} = \tilde{\sigma}_{gh,n}^{-1}T'_{gg,n} T_{hh,n} \tilde{\alpha}_{h,1,n}, \ldots, \tilde{\alpha}_{h,S,n}.
\]

Note that

\[ n^{-1}T'_{gg,n} T_{hh,n} = [n^{-1}Z'_{s,n}(\tilde{\rho}_{g,n})\hat{Z}_{s,n}(\tilde{\rho}_{g,n})]^{-1} n^{-1}Z'_{s,n}(\tilde{\rho}_{g,n})\hat{Z}_{s,n}(\tilde{\rho}_{g,n}) \times \]

\[ n^{-1}Z'_{s,n}(\tilde{\rho}_{h,n})\hat{Z}_{s,n}(\tilde{\rho}_{h,n})]^{-1}, \]

41
and

\[ n^{-1} \hat{T}_{gg,n} \hat{T}_{gg,n} = \begin{bmatrix} n^{-1} \hat{Z}_{g,n}( \hat{\rho}_{g,n} ) \hat{Z}_{g,n}( \hat{\rho}_{g,n} ) \end{bmatrix}^{-1} = \begin{bmatrix} n^{-1} \hat{Z}_{g,n}( \hat{\rho}_{g,n} ) \hat{Z}_{g,n}( \hat{\rho}_{g,n} ) \end{bmatrix}^{-1}. \]

Also note that in light of Lemma A.5 we have \( n^{-1} \hat{T}_{gg,n} \hat{T}_{hh,n} = 0 \) for \( g \neq h \).

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

**Part 4:** (Limiting Distribution of \( \hat{\delta}_{g,n} \) and \( \hat{\rho}_{g,n} \)) 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 \( T_{gh,n} = 0 \) for \( g \neq h \) the expression for \( \Omega_n \) given in (28) simplify to:

\[ \Omega_n = \begin{bmatrix} \Omega_{\delta \delta}^n & \Omega_{\delta \rho}^n & \Omega_{\rho \delta}^n & \Omega_{\rho \rho}^n \\ \Omega_{\delta \rho}^n & \Omega_{\rho \rho}^n & \Omega_{\rho \rho}^n & \Omega_{\rho \rho}^n \end{bmatrix} \]

with

\[ \Omega_{\delta \delta}^n = \Psi_{\delta \delta}^n, \]

\[ \Omega_{\delta \rho}^n = \Psi_{\delta \rho}^n diag_{g=1}^G \left( (\Psi_{gg,n}^{\rho \rho})^{-1} J_{g,n} (\Psi_{gg,n}^{\rho \rho})^{-1} J_{g,n} \right)^{-1}, \]

\[ \Omega_{\rho \rho}^n = diag_{g=1}^G \left( (J_{g,n}(\Psi_{gg,n}^{\rho \rho})^{-1} J_{g,n}(\Psi_{gg,n}^{\rho \rho})^{-1} J_{g,n})^{-1} \right) \times \Psi_{\rho \rho}^{G,G} \]

By Theorem 4, \( n^{1/2} [\hat{\delta}_{g,n} - \delta_{g,n}, (\hat{\rho}_{g,n} - \rho_{g,n})]' \overset{d}{\rightarrow} N(0, \Omega_n) \), and as a specialization,

\[ n^{1/2} \begin{bmatrix} \hat{\delta}_{g,n} - \delta_{g,n} \\ \hat{\rho}_{g,n} - \rho_{g,n} \end{bmatrix} \overset{d}{\rightarrow} N \begin{bmatrix} \Omega_{\delta \delta}^{gg,n} & \Omega_{\delta \rho}^{gg,n} & \Omega_{\rho \delta}^{gg,n} & \Omega_{\rho \rho}^{gg,n} \\ \Omega_{\delta \rho}^{gg,n} & \Omega_{\rho \rho}^{gg,n} & \Omega_{\rho \rho}^{gg,n} & \Omega_{\rho \rho}^{gg,n} \end{bmatrix} \]

with

\[ \Omega_{\delta \delta}^{gg,n} = \Psi_{\delta \delta}^{gg,n} = \sigma_{gg,n} n^{-1} T_{gg,n} T_{gg,n}', \]

\[ \Omega_{\delta \rho}^{gg,n} = \Psi_{\delta \rho}^{gg,n} (\Psi_{gg,n}^{\rho \rho})^{-1} J_{g,n} (\Psi_{gg,n}^{\rho \rho})^{-1} J_{g,n} \]

\[ = \sigma_{gg,n} n^{-1} T_{gg,n} T_{gg,n} [\alpha_{g,1,n}, \ldots, \alpha_{g,S,n}] (\Psi_{gg,n}^{\rho \rho})^{-1} J_{g,n} (\Psi_{gg,n}^{\rho \rho})^{-1} J_{g,n} \]

\[ \Omega_{\rho \rho}^{gg,n} = (J_{g,n}(\Psi_{gg,n}^{\rho \rho})^{-1} J_{g,n})^{-1}. \]
Observe from part 3 of the proof that
\[ n^{-1} \hat{T}_{g,n} \hat{T}_{g,n} = n^{-1} \hat{Z}_{g,n}(\hat{\rho}_{g,n}) \hat{Z}_{g,n}(\hat{\rho}_{g,n})^{-1} \]
and that \( n^{-1} \hat{T}_{g,n} \hat{T}_{g,n} - n^{-1} T_{g,n} T_{g,n} = o_p(1) \). 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. \[ \square \]

**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 \( \hat{\sigma}_n \)) In light of Lemma A.6 we have
\[ n^{1/2} [\hat{\sigma}_n - \sigma_n] = n^{-1/2} T_n^t e_n + o_p(1) \] with
\[ T_n = (I_G \otimes H_n) P_n, \]
\[ P_n = [\Sigma^{-1} \otimes Q_{HH}] diag_{g=1} [Q_{HZ,g} \rho_{g,n}] \]
\[ \times \{ diag_{g=1} [Q_{HZ,g} \rho_{g,n}] [\Sigma^{-1} \otimes Q_{HH}] diag_{g=1} [Q_{HZ,g} \rho_{g,n}] \}^{-1}. \]

Now let \( T_{gh,n} \) and \( P_{gh,n} \) denote the \((g,h)\)-th block of \( T_n \) and \( P_n \), then \( T_{gh,n} = \hat{F}_{gh,n} P_{gh,n} \) with \( \hat{F}_{gh,n} = H_n \). The remaining conditions of Assumption 10 are then seen to hold in light of Lemma A.6.

**Part 2:** (Specialized Expressions for \( \hat{\Psi}_n \) and the Corresponding Estimator) Specialized expressions for each of the submatrices \( \hat{\Psi}_{n}^{\delta \delta}, \hat{\Psi}_{n}^{\delta \rho}, \hat{\Psi}_{n}^{\rho \rho} \) of \( \hat{\Psi}_n \) defined in (26) are readily found by substituting into those expressions the formulae for \( T_n \) given in (C.2), and by observing that \( \sum_{g=1}^{G} \sum_{v=1}^{G} \sigma_{uv,n} T_{gu,n} T_{hv,n} \) represents the \((g,h)\)-th block of \( \hat{\Psi}_n^{\delta \delta} = T_n^t (\Sigma \otimes I) T_n \).

Next let
\[ \hat{\Psi}_n = (I_G \otimes H_n) \hat{P}_n, \]
\[ \hat{\Psi}_n = [\hat{\Sigma}_n^{-1} \otimes (n^{-1} H_n^t H_n)^{-1}] diag [n^{-1} H_n^t Z_{s,n}(\hat{\rho}_{g,n})] \]
\[ \times [n^{-1} \hat{Z}_{s,n}(\hat{\rho}_{g,n}) (\hat{\Sigma}_n^{-1} \otimes I_n) Z_{s,n}(\hat{\rho}_{g,n})]^{-1}. \]

Let \( \hat{\Psi}_{gh,n} \) and \( \hat{\Psi}_{gh,n} \) denote the \((g,h)\)-th block of \( \hat{\Psi}_n \) and \( \hat{\Psi}_n \), respectively. Then, clearly, \( \hat{\Psi}_{gh,n} = H_n \hat{P}_{gh,n} \). Next observe that
\[ \hat{\Psi}_n^{\delta \delta} = n^{-1} \hat{Z}_{s,n}(\hat{\rho}_{g,n}) (\hat{\Sigma}_n^{-1} \otimes I_n) Z_{s,n}(\hat{\rho}_{g,n})^{-1} = n^{-1} \hat{T}_n^t (\hat{\Sigma}_n \otimes I_n) \hat{T}_n, \]
and thus the \((g,h)\)-th block of \( \hat{\Psi}_n^{\delta \delta} \) is given by
\[ \hat{\Psi}_{gh,n} = \sum_{v=1}^{G} \sigma_{uv,n} T_{gu,n} \hat{T}_{hv,n}. \]

From this we see that the estimators \( \hat{\Psi}_n, \hat{\Psi}_n, \hat{\Psi}_n, \Omega_n, \hat{\Omega}_n, \hat{\Omega}_n \) 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
\[ n^{-1} \hat{T}_{gh,n} \hat{T}_{gl,n} - T'_{gh,n} T_{gl,n} = o_p(1), \]
which verifies that also this condition of Theorem 4 holds.

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

**Part 4:** (Limiting distribution of \( \hat{\delta}_{g,n} \) and \( \hat{\rho}_{g,n} \)) 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
\[ \left[ \begin{array}{c} \hat{\delta}_n - \delta_n \\ \hat{\rho}_n - \rho_n \end{array} \right] \xrightarrow{d} N \left[ \begin{array}{cc} \Omega_{n}^{\hat{\delta}} & \Omega_{n}^{\hat{\rho}} \\ \Omega_{n}^{\hat{\rho}'} & \Omega_{n}^{\hat{\rho}'} \end{array} \right], \tag{C.3} \]
where
\[ \Omega_{n}^{\hat{\delta}} = \Psi_{n}^{\hat{\delta}}, \quad \Omega_{n}^{\hat{\rho}} = \Psi_{n}^{\hat{\rho}} \text{diag}_{g=1}^{G}(J_{g,n}), \quad \Omega_{n}^{pp} = \text{diag}_{g=1}^{G}(J_{g,n}) \Psi_{n}^{pp} \text{diag}_{g=1}^{G}(J_{g,n}) \]
with
\[ \Psi_{n}^{\hat{\rho}} = \Psi_{n}^{\hat{\rho}} \text{diag}_{g=1}^{G}[\alpha_{g,1,n}, \ldots, \alpha_{g,S,n}], \quad J_{g,n} = (\Psi_{n}^{pp})^{-1} J_{g,n} (\Psi_{n}^{pp})^{-1} J_{g,n}^{-1}. \]
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.
References


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

A Supplement to Appendix A

Proof of Lemma A.1: Let $C_A = \sup_n \max_{i=1}^m \sum_{j=1}^n |a_{ij,n}|$, $C_\mu = \sup_n \max_{i=1}^m E |\mu_{i,n}|^p$ and $C_\eta = \sup_n \max_{i=1}^m E |\eta_{i,n}|^p$. Clearly
\[ |\xi_{i,n}| \leq |\mu_{i,n}| + \sum_{j=1}^m |a_{ij,n}| |\eta_{jn}| \leq |\mu_{i,n}| + C_A \sum_{j=1}^m b_{ij,n} |\eta_{jn}| \]
with $b_{ij,n} = |a_{ij,n}| / \left( \sum_{j=1}^m |a_{ij,n}| \right)$ if $\sum_{j=1}^m |a_{ij,n}| > 0$ and $b_{ij,n} = 0$ if $\sum_{j=1}^m |a_{ij,n}| = 0$. Since $0 \leq b_{jn} \leq 1$ and $\sum_{j=1}^m b_{ij,n} = 1$ it follows from Holder’s inequality that
\[ \sum_{j=1}^m b_{ij,n} |\eta_{jn}| \leq \left[ \sum_{j=1}^m b_{ij,n} |\eta_{jn}|^p \right]^{1/p} \]
and consequently
\[ E |\xi_{i,n}|^p \leq 2^p E |\mu_{i,n}|^p + 2^p C_A^p \left[ \sum_{j=1}^m b_{ij,n} |\eta_{jn}| \right]^p \]
\[ \leq 2^p E |\mu_{i,n}|^p + 2^p C_A^p \sum_{j=1}^m b_{ij,n} E |\eta_{jn}|^p \leq 2^p C_\mu + 2^p C_A^p C_\eta < \infty \]
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 $X_n$ is non-stochastic with $\sup_n \sup_{i,k} |x_{ik,n}| < \infty$, and so (A.1) holds trivially if $\xi_{i,n}$ corresponds to an element of $X_n$. Next observe that by (4) and (6) we have
\[ y_n = a_n + A_n \nu_n \]
\[ a_n = (I_n G - B_n)^{-1} C_n x_n, \]
\[ A_n = (I_n G - B_n^*)^{-1} (I_n G - R_n^*)^{-1} (\Sigma^* \otimes I_n). \]
In light of Assumptions 1-3 the absolute elements of $a_n$ are uniformly bounded, and furthermore the row and column sums of the absolute elements of $A_n$ are uniformly bounded; compare, e.g., Remark A.1 in Kelejian and Prucha (2004). By Assumption 4 the elements of $\nu_n$ 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 $Y_n$ are of the form $y_{l,s,n} = W_{s,n} y_{l,n}$. Since by Assumption 1 the row and column sums of the absolute elements of $W_{s,n}$ are uniformly bounded it follows further from Lemma A.1 that $\sup_n \sup_{i,l,s} E |y_{il,s,n}|^4 < \infty$, which completes the proof. ■
**Proof of Lemma A.3:** In light of the proof of Lemma A.2, and observing that 
\( u_n = (I_n - R^*_n)^{-1}(\Sigma'_n \otimes I_n)\nu_n \), it is readily seen that under the maintained assumptions \( u_{g,n} \) and all columns of \( Z_n \) are of the generic form

\[
C_{g,n}\nu_n, c_{g,n} \text{ or } c_{g,n} + C_{g,n}\nu_n,
\]

where \( c_{g,n} \) is an \( n \times 1 \) nonstochastic vector with uniformly bounded elements and \( C_{g,n} \) 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 \( nG \times 1 \) nonstochastic matrix \( \nu_n \) are i.i.d. \( (0,1) \) with finite fourth moments. Given this, it is readily seen that \( n^{-1}u_{h,n}^T A_n u_{g,n} \) and the elements of \( n^{-1}Z_n^T A_n u_{g,n} \) and \( n^{-1}Z_n A_n Z_n \) are of the generic form \( d_n \), \( n^{-1}d_n^T \nu_n \) or \( n^{-1}\nu_n^T D_n \nu_n \), or sums thereof, where \( |d_n| \), the absolute elements of \( d_n \) and the row and column sums of the absolute elements of \( D_n \) are uniformly bounded by some finite constant, say \( K \). In the following let \( \bar{D}_n = (\bar{d}_{ik,n}) = (D_n + D_n^T)/2 \). Observe that \( E[n^{-1}d_n^T \nu_n] = 0 \) and \( E[n^{-1}d_n^T \nu_n, D_n \nu_n] = n^{-1}tr(D_n) = O(1) \). Furthermore, observe that \( \text{var}(n^{-1}d_n^T \nu_n) \leq n^{-1}K^2 = o(1) \), and that in light of, e.g., Lemma A.1 in Kelejian and Prucha (2004) and Remark 2 in Kapoor et al. (2007), we have \( \text{var}(n^{-1}\nu_n^T D_n \nu_n) \leq n^{-2}tr(\bar{D}_n^2) + n^{-2}K^2 \sum_{i=1}^G |E\nu_{ig,n}^4 - 3| \leq n^{-1}K^2 = o(1) \) for some finite constant \( K^* \). Thus clearly \( n^{-1}u_{h,n}^T A_n u_{g,n} = O_p(1) \), \( n^{-1}Z_n^T A_n u_{g,n} = O_p(1) \) and \( n^{-1}Z_n A_n Z_n = O_p(1) \). The third claim in the lemma follows from Chebyshev’s inequality.

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

\[
n^{1/2}(\delta_{g,n} - \delta_{g,n}) = \tilde{P}'_{gg,n} n^{-1/2} F'_{gg,n} \varepsilon_{g,n},
\]

where \( \tilde{P}_{gg,n} \) and \( F_{gg,n} \) are defined in the lemma. Given Assumption 6 clearly \( P_{gg,n} = P_{gg} + o_p(1) \) with \( P_{gg} \) being finite, which establishes (c). Since by Assumption 2 the row and column sums of \( (I_n - R^*_{g,n})^{-1} \) are uniformly bounded in absolute value, and since by Assumption 5 the elements of \( H_n \) are uniformly bounded in absolute value, it follows that the elements of \( F_{gg,n} \) are uniformly bounded in absolute value. By Assumption 4, \( E(\varepsilon_{g,n}) = 0 \) and \( E(\varepsilon_{g,n} \varepsilon_{g,n}') = \sigma_{gg} I_n \). Therefore, \( E n^{-1/2} F'_{gg,n} \varepsilon_{g,n} = 0 \) and the elements of \( VC(n^{-1/2}F_{gg,n} \varepsilon_{g,n}) = \sigma_{gg} n^{-1} F'_{gg,n} F_{gg,n} \) are also uniformly bounded in absolute value. Thus, by Chebyshev’s inequality \( n^{-1/2} F'_{gg,n} \varepsilon_{g,n} = O_p(1) \), and consequently \( n^{1/2}(\delta_{g,n} - \delta_{g,n}) = P_{gg} n^{1/2} F'_{gg,n} \varepsilon_{g,n} + o_p(1) \) and \( P_{gg} n^{1/2} F'_{gg,n} \varepsilon_{g,n} = O_p(1) \). This establishes (a) and (b), recalling that \( T_{gg,n} = F_{gg,n} P_{gg,n} \). Next observe that

\[
\begin{align*}
\lambda_{\min}(n^{-1}T'_{gg,n} T_{gg,n}) \\
\geq \lambda_{\min} \left[ (I_n - R^*_{g,n})^{-1} (I_n - R^*_{g,n})^{-1} \right] \lambda_{\min} \left[ n^{-1}H'_n H_n \right] \\
\lambda_{\min} \left\{ [Q'_{HZ,g} Q_{HH}^{-1} Q_{HZ,g}]^{-1} Q'_{HH} Q_{HH}^{-1} Q_{HZ,g} [Q'_{HZ,g} Q_{HH}^{-1} Q_{HZ,g}]^{-1} \right\} \\
\geq c
\end{align*}
\]

50
for some \( c > 0 \), since in light of Assumptions 1 and 2 the largest eigenvalue of \((I_n - R^*_g)^{-1}(I_n - R^*_g')^{-1}\) is bounded from above, and thus the smallest eigenvalue of \((I_n - R^*_g)^{-1}(I_n - R^*_g')^{-1}\) is bounded away from zero, and since 
\[
\lambda_{\min}\left[n^{-1}H'_nH_n\right] \geq [\lambda_{\min}(Q_{HH})]/2 > 0 \quad \text{for} \quad n \text{ sufficiently large in light of Assumption 6.} \]
This establishes (d).

\[\Box\]

**Proof of Lemma A.5:** Note from (1) and (7) that
\[
y_{g,n}(\hat{\rho}_{g,n}) = Z_{g,n}(\hat{\rho}_{g,n})\delta_{g,n} + \varepsilon_{g,n} - (\hat{R}^*_g - R_{g,n})u_{g,n}
\]
and hence
\[
n^{-1/2}[\delta_{g,n} - \delta_{g,n}] = \left[n^{-1}Z'_{g,n}(\hat{\rho}_{g,n})Z_{g,n}(\hat{\rho}_{g,n})\right]^{-1} - \left[n^{-1/2}Z'_{g,n}(\hat{\rho}_{g,n})\right]^{-1} n^{-1/2}Z'_{g,n}(\hat{\rho}_{g,n}) n^{-1/2}F_{g,g,n}^r \varepsilon_{g,n},
\]
and hence
\[
\begin{align*}
\hat{\rho}_{g,n}^* = R_{g,n}^*(\hat{\rho}_{g,n}) \quad \text{and} \quad \hat{P}_{g,g,n}^* = H_n, \quad \text{since} \quad \hat{P}_{g,g,n}^* = H_n, \quad \text{and} \quad \hat{F}_{g,g,r,n}^* = (I_n - R_{g,n})^{-1} \hat{M}_{r,n}^* H_n, \quad \text{in light of Assumption 6.}
\end{align*}
\]
This establishes (a)-(c) recalling that \( T_{g,g,n}^* = F_{g,g,n}^* P_{g,g,n}^* \).
for some $c_* > 0$ in light of Assumption 6, and observing that $\lambda_{\min} \left[ n^{-1} H_n^T H_n \right] \geq \lambda_{\min}(Q_{HH}) / 2 > 0$ for $n$ sufficiently large. This establishes (d). □

**Proof of Lemma A.6:** Note from (1) and (8) that

$$ y_{sn}(\hat{\rho}_n) = Z_{sn}(\hat{\rho}_n)\delta_n + \varepsilon_n - (\hat{R}_n^* - R_n)u_n $$

and hence

$$ n^{1/2} [\hat{\delta}_n - \delta_n] = \left[ n^{-1} \hat{Z}_{sn}(\hat{\rho}_n)(\hat{\Sigma}_n^{-1} \otimes I_n)Z_{sn}(\hat{\rho}_n) \right]^{-1} n^{-1/2} \hat{z}'_{sn}(\hat{\rho}_n)(\hat{\Sigma}_n^{-1} \otimes I_n) \varepsilon_n - (\hat{R}_n^* - R_n)u_n $$

$$ = \left[ n^{-1} \hat{Z}_{sn}(\hat{\rho}_n)(\hat{\Sigma}_n^{-1} \otimes I_n)Z_{sn}(\hat{\rho}_n) \right]^{-1} \text{diag} \left[ n^{-1} Z_{g,n}(\hat{\rho}_{g,n})H_n \right] Z_{sn}(\hat{\rho}_n) \varepsilon_n - (\hat{R}_n^* - R_n)u_n $$

$$ = \left[ \hat{P}_n \varepsilon_n + \hat{F}_n \varepsilon_n \right] Z_{sn}(\hat{\rho}_n) \delta_n - \text{diag} \left[ Q_{HZ,g*}(\rho_{g,n}) \right] \left[ \Sigma^{-1} \otimes Q_{HH}^{-1} \right] \text{diag} \left[ Q_{HZ,g*}(\rho_{g,n}) \right] $$

$$ n^{-1} \hat{Z}_{sn}(\hat{\rho}_n)(\hat{\Sigma}_n^{-1} \otimes I_n)Z_{sn}(\hat{\rho}_n) - \text{diag} \left[ Q_{HZ,g*}(\rho_{g,n}) \right] \left[ \Sigma^{-1} \otimes Q_{HH}^{-1} \right] \text{diag} \left[ Q_{HZ,g*}(\rho_{g,n}) \right] $$

is

$$ \hat{g}_{gh,n} n^{-1} \hat{Z}_{gh,n}(\hat{\rho}_{gh,n})Z_{gh,n}(\hat{\rho}_{gh,n}) - \sigma_{gh} Q_{HZ,g*}(\rho_{g,n}) Q_{HH}^{-1} Q_{HZ,h*}(\rho_{g,n}) = o_p(1) $$

in light of Assumption 6, and since $\hat{\rho}_n$ and $\hat{\Sigma}_n$ are consistent. By Assumption 6 we have $\text{diag} \left[ Q_{HZ,g*}(\rho_{g,n}) \right] \left[ \Sigma^{-1} \otimes Q_{HH}^{-1} \right] \text{diag} \left[ Q_{HZ,g*}(\rho_{g,n}) \right] = O(1)$ and

$$ \lambda_{\min} \left\{ \text{diag} \left[ Q_{HZ,g*}(\rho_{g,n}) \right] \left[ \Sigma^{-1} \otimes Q_{HH}^{-1} \right] \text{diag} \left[ Q_{HZ,g*}(\rho_{g,n}) \right] \right\} \geq \lambda_{\min} \left\{ \Sigma^{-1} \otimes I \right\} \lambda_{\min} \left\{ \text{diag} \left[ Q_{HZ,g*}(\rho_{g,n}) \right] Q_{HH}^{-1} Q_{HZ,h*}(\rho_{g,n}) \right\} \geq c_* $$

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

$$ \left\{ \text{diag} \left[ Q_{HZ,g*}(\rho_{g,n}) \right] \left[ \Sigma^{-1} \otimes Q_{HH}^{-1} \right] \text{diag} \left[ Q_{HZ,g*}(\rho_{g,n}) \right] \right\}^{-1} = O(1). $$

Consequently

$$ \left\{ n^{-1} \hat{Z}_{sn}(\hat{\rho}_n)(\hat{\Sigma}_n^{-1} \otimes I_n)Z_{sn}(\hat{\rho}_n) \right\}^{-1} - \left\{ \text{diag} \left[ Q_{HZ,g*}(\rho_{g,n}) \right] \left[ \Sigma^{-1} \otimes Q_{HH}^{-1} \right] \text{diag} \left[ Q_{HZ,g*}(\rho_{g,n}) \right] \right\}^{-1} = o_p(1); $$

52
Then, using the triangle inequality and the Cauchy-Schwarz inequality, we have sums of the absolute elements of Remark A.1 in Kelejian and Prucha (2004) - it follows that also the row and this property is preserved under matrix addition and multiplication - see, e.g., Proof of Lemma A.7: for some $\lambda_{\min}(n^{-1/2}T_{n}^{**}T_{n}^{**})$

$$\lambda_{\min}(\Sigma^{-1})\lambda_{\min}[Q_{HHH}^{-1/2}P_{HHH}^{-1/2}]$$

$$\times \lambda_{\min}[\{\text{diag}[Q_{HZ,\rho_{g}}(\rho_{g,n})][\Sigma^{-1} \otimes Q_{HHH}^{-1}] \text{diag}[Q_{HZ,\rho_{g}}(\rho_{g,n})]\}^{-1}]$$

$$\geq \lambda_{\min}(\Sigma^{-1})\lambda_{\min}(Q_{HHH}^{-1})\lambda_{\min}[n^{-1}H_{n}^{T}H_{n}]$$

$$\times \lambda_{\min}[\{\text{diag}[Q_{HZ,\rho_{g}}(\rho_{g,n})][\Sigma^{-1} \otimes Q_{HHH}^{-1}] \text{diag}[Q_{HZ,\rho_{g}}(\rho_{g,n})]\}^{-1}]$$

$$\geq \lambda_{\min}(\Sigma^{-1})\lambda_{\min}(Q_{HHH}^{-1})[\lambda_{\min}(Q_{HHH})/2]$$

$$\times \lambda_{\min}[\{\text{diag}[Q_{HZ,\rho_{g}}(\rho_{g,n})][\Sigma^{-1} \otimes Q_{HHH}^{-1}] \text{diag}[Q_{HZ,\rho_{g}}(\rho_{g,n})]\}^{-1}] \geq c_{*}$$

for some $c_{*} > 0$ in light of Assumption 6, and observing that $\lambda_{\min}[n^{-1}H_{n}^{T}H_{n}] \geq [\lambda_{\min}(Q_{HHH})]/2 > 0$ for $n$ sufficiently large. This establishes (d). $\blacksquare$

**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 $\bar{\vartheta}_{n} = n^{-1/2}u_{n}'A_{u}^{*}u_{n}$ and $\bar{\vartheta}_{n} = n^{-1/2}u_{n}'A_{u}^{*}u_{n}$, then, in light of Assumption A.1, we have $\vartheta_{n} = n^{-1/2}u_{n}'B_{n}^{*}u_{n}$ with $B_{n}^{*} = (1/2)|A_{n}' + A_{n}^{*}|A_{n}^{-1}$. Furthermore, by Assumption A.1, the row and column sums of the matrices $B_{n}$ 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 $B_{n}^{*}$ and $B_{n}^{*}B_{n}^{*}$ 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 $B_{n}$ and $B_{n}^{*}$, and of their respective elements.

Then, using the triangle inequality and the Cauchy-Schwarz inequality, we have

$$E|\vartheta_{n}| = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}|b_{i,j,n}||E|\epsilon_{i,n}| |\epsilon_{j,n}| \leq n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}|b_{i,j,n}| \leq K.$$  

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

$$\text{var}(\vartheta_{n}) = n^{-2}2tr(B_{n}^{*}B_{n}^{*}) + n^{-2}\sum_{i=1}^{n}b_{i,i,n}^{2}[E\epsilon_{i,n}^{4} - 3]$$

$$\leq n^{-1}2K + n^{-1}K^{2}\sup_{i}[E\epsilon_{i,n}^{4} - 3],$$

53
Given that the fourth moments of the $\epsilon_{i,n}$ are uniformly bounded in light of Assumption A.1, this establishes the first two claims of part (a) of the lemma.

We next proof the last claim of part (a) of the lemma. The above discussion implies that $\bar{\theta}_n - \theta_n = o_p(1)$. Hence it suffices to show that $\bar{\theta}_n - \theta_n = o_p(1)$.

By Assumptions A.1 and A.2

$$\bar{\theta}_n - \theta_n = \phi_n + \psi_n$$

with

$$\phi_n = n^{-1} \Delta'_n D'_n (A^*_n + A^*_n) u_n = n^{-1} \Delta'_n D'_n C^*_n \epsilon_n,$$

$$\psi_n = n^{-1} \Delta'_n D'_n A^*_n D_n \Delta_n,$$

and $C^*_n = (c^*_i,n) = (A^*_n + A^*_n) R_n^{-1}$. The row and column sums of the matrices $C^*_n$ are again seen to be uniformly bounded in absolute value. Let $\mathcal{K} < \infty$ denote a uniform bound for the row and column sums of the absolute elements of the matrices $A^*_n$ and $C^*_n$, and let $c^*_i,n$ and $d^*_i,n$ denote the $i$-th row of $C^*_n$ and $D_n$, 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

$$|\phi_n| = \left| n^{-1} \sum_{i=1}^n \Delta'_n D'_n c^*_i,n \epsilon_n \right|$$

$$\leq n^{-1} \|\Delta_n\| \sum_{i=1}^n \|\epsilon_{i,n}\| \left| \sum_{j=1}^n c^*_i,j,n \right| \leq n^{-1} \|\Delta_n\| \sum_{j=1}^n \|\epsilon_{j,n}\| \sum_{i=1}^n \|\epsilon_{i,n}\| \left| c^*_i,j,n \right|$$

$$\leq n^{-1} \|\Delta_n\| \sum_{j=1}^n \|\epsilon_{j,n}\| \left( \sum_{i=1}^n \|\epsilon_{i,n}\| \right)^{1/p} \left( \sum_{i=1}^n \|\epsilon_{i,n}\| \right)^{1/q}$$

$$\leq \mathcal{K} n^{1/p-1/2} \left( n^{1/2} \|\Delta_n\| \right) \left( n^{-1} \sum_{j=1}^n \|\epsilon_{j,n}\| \right) \left( n^{-1} \sum_{i=1}^n \|\epsilon_{i,n}\| \right)$$

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 $\epsilon_{j,n}$ are independent with bounded second moments, it follows that $n^{-1} \sum_{j=1}^n |\epsilon_{j,n}| = O_p(1)$. The terms $n^{1/2} \|\Delta_n\|$ and $n^{-1} \sum_{i=1}^n \|\epsilon_{i,n}\|^p$ are $O_p(1)$ by Assumption A.2. Since $n^{1/p-1/2} \rightarrow 0$ as $n \rightarrow \infty$ it follows that $\phi_n = o_p(1)$. 

54
Again, using the triangle and Hölder inequality yields
\[
|\psi_n| = \left| n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \Delta_n \delta_{i,n}^* a_{i,j,n}^* \delta_{j,n} \Delta_n \right| \quad (A.3)
\]
\[
\leq n^{-1} \|\Delta_n\|^2 \sum_{i=1}^{n} \|\delta_{i,n}\| \sum_{j=1}^{n} \|\delta_{j,n}\| \|a_{i,j,n}^*\| \leq Kn^{1/p} \|\Delta_n\|^2 \left( n^{-1} \sum_{i=1}^{n} \|\delta_{i,n}\|^p \right)^{1/p} \leq Kn^{1/p - 1/2} n^{-1/2} (n^{1/2} \|\Delta_n\|)^2 \left( n^{-1} \sum_{i=1}^{n} \|\delta_{i,n}\|^p \right) \leq Kn^{1/p - 1/2} n^{-1/2} (n^{1/2} \|\Delta_n\|)^2 \left( n^{-1} \sum_{i=1}^{n} \|\delta_{i,n}\|^p \right)
\]
with \( p \) and \( q \) as before. By Assumption A.2 both \( n^{-1} \sum_{i=1}^{n} \|\delta_{i,n}\|^p \) and \( n^{1/2} \|\Delta_n\| \) 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 \( \bar{\varrho}_{s,n} \) denote the \( s \)-th element of \( n^{-1} \mathcal{D}_n^* A_n^* u_n \). Observe \( Ev_n u_n' = \mathcal{R}_n^{-1} \mathcal{R}_n^{-1} \). Then given Assumptions A.1 and A.2 there exists a constant \( K < \infty \) such that \( Ev_n u_n' \leq K \) and \( E|\delta_{s,n}|^p \leq K \). W.l.o.g. assume that the row and column sums of the matrices \( A_n^* \) are uniformly bounded by \( E|\delta_{s,n}|^p \leq K \). Utilizing the Cauchy-Schwarz and Lyapunov inequalities we then have \( E|u_{i,n}| \delta_{s,n} \leq (Ev_n u_n')^{1/2} \leq E|\sigma_{s,n}|^{1/2} \leq E(v_n 
abla 1/p) = E[\delta_{s,n}]^{1/2} \leq E[u_n']^{1/2} \leq Kn^{1/p} \). With p as before and, hence,
\[
E|\bar{\varrho}_{s,n}| = n^{-1} \sum_{i=1}^{n} \|a_{i,n}^*\| \sum_{j=1}^{n} E|u_{i,n}| \delta_{s,n} \leq Kn^{1/p - 1/2} n^{-1} \sum_{i=1}^{n} \|a_{i,n}^*\| \leq Kn^{1/p - 1/2} < \infty,
\]
which shows that indeed \( E|n^{-1} \delta_{s,n}^* A_n^* u_n| = O(1) \) where \( \delta_{s,n} \) denotes the \( s \)-th column of \( \mathcal{D}_n \). Of course, the argument also shows that \( \alpha_n^* = n^{-1} E \mathcal{D}_n^* (A_n^* + A_n') u_n = O(1) \). Next observe that
\[
n^{-1} \mathcal{D}_n^* A_n^* \bar{\psi}_n = n^{-1} \mathcal{D}_n^* A_n^* u_n + \psi_n,
\]
where \( \bar{\psi}_n = n^{-1} \mathcal{D}_n^* A_n^* \mathcal{D}_n \Delta_n \). By argumentation analogous to that employed to demonstrate that \( n^{1/2} \psi_n = o_p(1) \) it follows that also \( \bar{\psi}_n = o_p(1) \), which completes the proof of part (b).

We next prove part (c). In light of the proof of part (a) we have
\[
n^{-1/2} \mathcal{D}_n^* A_n^* \bar{\psi}_n = n^{-1/2} u_n' A_n^* u_n + [n^{-1} u_n' (A_n^* + A_n')] \mathcal{D}_n |n^{1/2} \Delta_n + n^{1/2} \psi_n
\]

55
with \( n^{1/2} \psi_n = o_p(1) \). In light of part (b) and Assumption A.3 we have \( n^{-1/2} u'_n (A_n^* + A_n') D_n - \alpha_n' = o_p(1) \). The claim follows since \( n^{1/2} \Delta_n = O_p(1) \) 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 \( K \) used in proving the last claim of part (a) of the above lemma can be chosen as \( K = 2 c_P c_A \) where \( c_P \) and \( c_A \) denote a bound for the row and column sums of the absolute elements of \( R_n^{-1} \) and \( A_n^* \). Furthermore it proves helpful to observe that in light of (A.2) and (A.3)

\[
|\bar{y}_n - \eta_n| \leq 2 c_P c_A s_n,
\]

where \( s_n = o_p(1) \) does not depend on \( A_n^* \).

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

\[
\begin{align*}
    n^{-1/2} \bar{u}'_n \delta_n' A_n \delta_n \bar{u}_n &= n^{-1/2} u'_n \delta_n' A_n \delta_n u_n + \alpha_n' n^{1/2} \Delta_n + o_p(1) \\
    &= n^{-1/2} \varepsilon_{g,n} A_n \varepsilon_{g,n} + n^{-1/2} \alpha_n' \sum_{h=1}^{G} T'_{h,n} \varepsilon_{h,n} + o_p(1) \\
    &= n^{-1/2} \varepsilon_{g,n} A_n \varepsilon_{g,n} + n^{-1/2} \sum_{h=1}^{G} a'_{h,n} \varepsilon_{h,n} + o_p(1).
\end{align*}
\]

The last inequality holds since \( \alpha_n = O(1) \); see the remark in Lemma A.7(c). Given this and the maintained assumption on \( P_{h,n} \) it follows that \( c_{h,n} = (c_{h1,n}, \ldots, c_{hp,n})' = P_{h,n} \alpha_n = O(1) \). Since \( a_{h,n} = F_{h,n} c_{h,n} \) we have

\[
|a_{h,i,n}|^\eta = \sum_{s=1}^{pF} f_{his,n} c_{h,s,n} \leq p_F^{K} \sum_{s=1}^{pA} f_{his,n} \]

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

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

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).