Badinger, Harald; Egger, Peter

Working Paper

GM estimation of higher-order spatial autoregressive processes in cross-section models with heteroskedastic disturbances

CESifo Working Paper, No. 2356

Provided in Cooperation with:
Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Badinger, Harald; Egger, Peter (2008) : GM estimation of higher-order spatial autoregressive processes in cross-section models with heteroskedastic disturbances, CESifo Working Paper, No. 2356, Center for Economic Studies and ifo Institute (CESifo), Munich

This Version is available at:
https://hdl.handle.net/10419/26401

GM Estimation of Higher-Order Spatial Autoregressive Processes in Cross-Section Models with Heteroskedastic Disturbances

HARALD BADINGER
PETER EGGER

CESIFO WORKING PAPER NO. 2356
CATEGORY 10: EMPIRICAL AND THEORETICAL METHODS
JULY 2008

CESifo *Working Paper No. 2356*

GM Estimation of Higher-Order Spatial Autoregressive Processes in Cross-Section Models with Heteroskedastic Disturbances

Abstract

This paper generalizes the approach to estimating a first-order spatial autoregressive model with spatial autoregressive disturbances (SARAR(1,1)) in a cross-section with heteroskedastic innovations by Kelejian and Prucha (2008) to the case of spatial autoregressive models with spatial autoregressive disturbances of arbitrary (finite) order (SARAR(*R*,*S*)). We derive the moment conditions and the optimal weighting matrix for a generalized moments (GM) estimation procedure of the spatial regressive parameters of the disturbance process and define a generalized two-stages least squares estimator for the regression parameters of the model. We prove consistency of the proposed estimators, derive their (joint) asymptotic distribution, and provide Monte Carlo evidence on their small sample performance.

JEL Code: C21, C31.

Keywords: higher-order spatial dependence, heteroskedasticity, two-stages least squares, generalized moments estimation, asymptotics.

*Harald Badinger**
Department of Economics
Vienna University of Economics and Business Administration
Althanstrasse 39 - 45
1090 Vienna
Austria
harald.badinger@wu-wien.ac.at

Peter Egger
Ifo Institute for Economic Research at the University of Munich
Poschingerstrasse 5
81679 Munich
Germany
egger@ifo.de

July 2008
*This paper was written during my stay as Erwin-Schrödinger Research Fellow at the Center for Economic Studies (CES), University of Munich. Financial support by the Austrian Science Fund (FWF) is gratefully acknowledged.

I. Introduction

In recent years, econometric research started developing estimators for cross-sectional models, where the units of observations are allowed to be correlated. A large class of such models is referred to as *spatial econometric models*, where interdependence occurs through some ex ante known channel. One possible—yet not necessarily the only plausible—channel is geographical distance or *space* as such. The use of and empirical support for the latter gave the corresponding subfield in econometrics its name: *spatial econometrics*. A majority of existing theoretical models and applications follows the general structure introduced by Cliff and Ord (1973, 1981): a continuous endogenous variable is specified as a function of a *spatial lag*, i.e., the spatially weighted average of the endogenous variable, a set of exogenous explanatory variables, and possibly spatially autocorrelated residuals.[1] This framework with both a spatial lag and spatial autoregressive disturbances is commonly referred to as SARAR model.

Almost all theoretical or applied work assumes that the data-generating SARAR process is of first-order, i.e., SARAR(1,1). In principal, this is an unnecessary restriction and it would be surprising if it were generally supported with real data.[2] However, to date a generalized model is not available for a SARAR process including spatial lags up to an order R and spatial dependence of the residuals up to an order S, i.e., SARAR(R,S), with fixed R and S.

This paper derives a generalized moments (GM) estimator and two-stages least squares estimators (TSLS) for the Cliff and Ord-type, cross-sectional model with a SARAR(R,S) structure, generalizing the estimation procedure for a SARAR(1,1) model with heteroskedastic innovations by Kelejian and Prucha (2008). We demonstrate consistency of the proposed estimators and determine the optimal weighting matrix for the moment conditions. Furthermore, we derive the joint asymptotic distribution of the GM estimates of the spatial autoregressive parameters of the disturbance process and the feasible (generalized) TSLS estimates of the regression parameters of the model. The latter provides the basis for Wald statistics which allow the researcher to test the estimated general SARAR(R,S) model

[1] Econometric work on Cliff and Ord (1973) models includes Anselin (1988), Baltagi and Li (2001), Baltagi, Song, and Koh (2003), Conley (1999), Kelejian and Prucha (1999, 2008), Kapoor, Kelejian, and Prucha (2007), Lee (2004, 2007), Pinkse and Slade (1998), Pinkse, Slade, and Brett (2002).

Applications of such models are legion, and they include Audretsch and Feldmann (1996), Baltagi, Egger and Pfaffermayr (2005), Besley and Case (1995), Case, Hines, and Rosen (1993), Cohen and Morrsison Paul (2004), Holtz-Eakin (1994), Shroder (1995), and Topa (2001), to mention only a few.

[2] There are a few empirical studies which allow for higher-order spatial processes. Yet, they are typically based on *either* higher-order spatial autoregressive residuals (see Bell and Bockstael, 2000; Badinger and Egger, 2008; Cohen and Morrison-Paul, 2007) *or* higher-order spatial lags of the endogenous variable (see Egger and Raff, 2008).

against alternatives such as SARAR(1,1), SARAR(0,*S*), or SARAR(*R*,0). We illustrate in a set of Monte Carlo simulations that the proposed estimators work well, even in small samples.

The remainder of the paper is organized as follows. Section II introduces the model specification, some notation, and a set of basic assumptions. Sections III and IV derive the GM estimator and generalized two-stages least squares estimators for the SARAR(*R*,*S*) process, demonstrates consistency and asymptotically normality of the parameter estimates, and formulate a consistent estimator for the variance-covariance matrix of the joint distribution of all model parameter estimates. Section V summarizes the findings from a Monte Carlo simulation exercise with a special emphasis on the point estimates of the spatial parameters and the rejection probabilities of Wald tests of the SARAR(*R*,*S*) model against interesting alternatives such as SARAR(0,*S*), SARAR(*R*,0), and the non-spatial model. Section VI summarizes the results and concludes.

II. Model Specification

In the following, we generalize the specification by Kelejian and Prucha (2008), allowing spatial dependence in the endogenous variable and the disturbances of arbitrary but fixed order *R* and *S*, respectively, i.e., the structure and strength of the cross-sectional interdependence may vary across subsets of the $i = 1,...,N$ cross-sectional units. In matrix notation, the model reads as follows:

$$\mathbf{y}_N = \mathbf{X}_N\boldsymbol{\beta}_N + \sum_{r=1}^{R}\lambda_{r,N}\mathbf{W}_{r,N}\mathbf{y}_N + \mathbf{u}_N, \text{ or} \tag{1a}$$

$$\mathbf{y}_N = \mathbf{Z}_N\boldsymbol{\delta}_N + \mathbf{u}_N, \tag{1b}$$

$$\mathbf{u}_N = \sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N}\mathbf{u}_N + \boldsymbol{\varepsilon}_N, \tag{1c}$$

where $\mathbf{y}_N = (y_{1,N},...,y_{N,N})'$ is the $N \times 1$ vector of observations on the dependent variable. The regressor matrix $\mathbf{X}_N$ is of dimension $N \times K$ and contains the observations on the $k = 1,...,K$ (exogenous) explanatory variables, i.e., $\mathbf{X}_N = (\mathbf{x}_{1,N},...,\mathbf{x}_{K,N})$ with each $N \times 1$ vector $\mathbf{x}_{k,N}$ denoting the observations on the respective explanatory variable. The structure of the spatial dependence in $\mathbf{y}_N$ is determined by the $N \times N$ matrices $\mathbf{W}_{r,N}$, $r = 1,...,R$, whose elements $w_{ij,r,N}$ are assumed to be known. The expression $\overline{\mathbf{y}}_{r,N} = \mathbf{W}_{r,N}\mathbf{y}_N$ is referred to as the *r*-th spatial lag of $\mathbf{y}_N$.

In equation (1b), the $N \times (K + R)$ matrix $\mathbf{Z}_N$ is given by $\mathbf{Z}_N = (\mathbf{X}_N, \overline{\mathbf{Y}}_N)$, with $\overline{\mathbf{Y}}_N = (\overline{\mathbf{y}}_{1,N},...,\overline{\mathbf{y}}_{R,N})$, and $\boldsymbol{\delta}_N = (\boldsymbol{\beta}'_N, \boldsymbol{\lambda}'_N)'$, where the $K \times 1$ parameter vector of the exogenous

variables is given by $\boldsymbol{\beta}_N = (\beta_{1,N},...,\beta_{K,N})'$ and the $R \times 1$ vector of spatial regressive parameters of $\mathbf{y}_N$ is defined as $\boldsymbol{\lambda}_N = (\lambda_{1,N},...,\lambda_{R,N})'$.

The $N \times 1$ vector of error terms $\mathbf{u}_N = (u_{1,N},...,u_{N,N})'$ is assumed to follow a spatial regressive process given by (1c). The structure of the spatial dependence in $\mathbf{u}_N$ is determined by the $N \times N$ matrices $\mathbf{M}_{m,N}$, $m = 1,...,M$, whose elements $m_{ij,m,N}$ are assumed to be known. The expression $\bar{\mathbf{u}}_{m,N} = \mathbf{M}_{m,N}\mathbf{u}_N$ is referred to as the *m*-th spatial lag of $\mathbf{u}_N$. The $S \times 1$ vector of the spatial regressive parameters of $\mathbf{u}_N$ is defined as $\boldsymbol{\rho}_N = (\rho_{1,N},...,\rho_{S,N})'$. Finally, the vector $\boldsymbol{\varepsilon}_N = (\varepsilon_{1,N},...,\varepsilon_{N,N})'$ contains the innovations of the error process, which are assumed to be independently but not necessarily identically distributed, and whose properties will be specified in more detail below.

Note that all variables are allowed to depend on sample size *N*, i.e., to form triangular arrays. Such a specification is consistent, for example, with models where the weights matrix is row-normalized and the number of neighbours of a given cross-sectional unit depends on sample size (see Kapoor, Kelejian, and Prucha, 2007, p. 102). Note that $\mathbf{X}_N$ may also contain spatial lags of exogenous variables, since it is allowed to depend on sample size. As a result, the model specification in equations (1a)-(1c) is fairly general, allowing for higher-order spatial dependence in the dependent variable, the explanatory variables, and the disturbances.

To avoid confusion, a word on notation is in order here. Regarding the spatial lags of the dependent variable, we will always use index $r = 1,...,R$. However, for reasons that will become clear below, we need more than one index to denote the spatial lag of the disturbances. In expressions involving sums as equation (1c), we always use index $m = 1,...,S$. The more natural indexation $s = 1,...,S$ is reserved for the moment conditions and will also be used when the context is clear and there is no danger of confusion.

The following assumptions are maintained throughout the analysis:

Assumption 1.

(a) The diagonal elements of $\mathbf{W}_{r,N}$, $r = 1,...,R$, and $\mathbf{M}_{s,N}$, $s = 1,...,S$, are zero.

(b) Restrictions on admissible parameter space.

$$\lambda_{r,N} \in (-\underline{a}_N^{\lambda_r}, \bar{a}_N^{\lambda_r}), \text{ with } 0 < \underline{a}_N^{\lambda_r}, \bar{a}_N^{\lambda_r} \le a^{\lambda} < \infty, \; r = 1, \ldots, R, \text{ and } \sum_{r=1}^{R} |\lambda_{r,N}| < A_{\lambda} < \infty.$$

The first part of Assumption (1b) simply requires the parameters $\lambda_{r,N}$, $r = 1, \ldots, R$ to be finite; we take a^{λ} such that $a^{\lambda} = \max_{r=1,...,R} a_N^{\lambda,r}$ holds; the expression $\mathbf{a}^{\lambda}$ will be used to denote an $R \times 1$ vector with elements a^{λ}. In the second part of Assumption (1b), the scalar A_{λ}

generally depends on the properties of the weights matrices $\mathbf{W}_{r,N}$. For example, with row-normalized matrices $\mathbf{W}_{r,N}$, $r = 1,...,R$, using $A_\lambda = 1$ ensures that $(\mathbf{I} - \sum_{r=1}^{R} \lambda_{r,N}\mathbf{W}_{r,N})$ is invertible, as required in Assumption (1c). If the matrices $\mathbf{W}_{r,N}$ are not row-normalized, Assumption (1c) is implied by using $A_\lambda = \left(\max_{r=1,...,R} \left\| \mathbf{W}_{r,N} \right\| \right)^{-1}$ for some matrix norm $\| \cdot \|$ (see Horn and Johnson, 1985, p. 301). Analogous assumptions are made for the parameters of the spatial regressive error process:

$$\rho_{s,N} \in (-\underline{a}_N^{\rho_s}, \overline{a}_N^{\rho_s}), \text{ with } 0 < \underline{a}_N^{\rho_s}, \overline{a}_N^{\rho_s} \leq a_N^{\rho} < \infty, \; s = 1, \ldots, S, \text{ and } \sum_{m=1}^{R} \left| \rho_{m,N} \right| < A_\rho < \infty.$$

We take a^ρ such that $a^\rho = \max_{s=1,...,S} a_N^{\rho,s}$ holds; the expression $\mathbf{a}^\rho$ will be used to denote an $S \times 1$ vector with elements a_N^ρ. As above, with row-normalized matrices $\mathbf{M}_{s,N}$, $s = 1,...,S$, the second part of this assumption ensures invertibility of $(\mathbf{I} - \sum_{m=1}^{S} \rho_{m,N}\mathbf{M}_{m,N})$ when $A_\rho = 1$. If the matrices $\mathbf{M}_{s,N}$ are not row-normalized, Assumption (1c) is implied by using $A_\rho = \left(\max_{s=1,...,S} \left\| \mathbf{M}_{s,N} \right\| \right)^{-1}$ for some matrix norm $\| \cdot \|$.

(c) The matrices $(\mathbf{I} - \sum_{r=1}^{R} \lambda_{r,N}\mathbf{W}_{r,N})$ and $(\mathbf{I} - \sum_{m=1}^{S} \rho_{m,N}\mathbf{M}_{m,N})$ are nonsingular for $\lambda_r \in (-\underline{a}_N^{\lambda_r}, \overline{a}_N^{\lambda_r})$ and $\rho_s \in (-\underline{a}_N^{\rho_r}, \overline{a}_N^{\rho_r})$. This ensures that $\mathbf{y}_N$ and $\mathbf{u}_N$ are uniquely identified by (1a) and (1c) through

$$\mathbf{y}_N = (\mathbf{I} - \sum_{r=1}^{R} \lambda_{r,N}\mathbf{W}_N)^{-1}\mathbf{X}_N\boldsymbol{\beta}_N + (\mathbf{I} - \sum_{r=1}^{R} \lambda_{r,N}\mathbf{W}_N)^{-1}\mathbf{u}_N, \tag{2a}$$

$$\mathbf{u}_N = (\mathbf{I} - \sum_{s=1}^{S} \rho_{m,N}\mathbf{M}_{m,N})^{-1}\boldsymbol{\varepsilon}_N. \tag{2b}$$

Assumption 2.

(a) For $1 \leq i \leq N, N \geq 1$ the innovations $\varepsilon_{i,N}$ are (mutually) independently distributed with $E(\varepsilon_{i,N}) = 0$ and $E(\varepsilon_{i,N}^2) = \sigma_{i,N}^2$, where $0 < \underline{a}^\sigma \leq \sigma_{i,N}^2 \leq \overline{a}^\sigma < \infty$, and $\sup_{1 \leq i \leq N, N \geq 1} E\left|\varepsilon_{i,N}\right|^{4+\eta} < \infty$ for some $\eta > 0$. Note that the variance-covariance matrix of $\boldsymbol{\varepsilon}_N$ is given by

$$\boldsymbol{\Omega}_{\varepsilon,N} = E[\boldsymbol{\varepsilon}_N\boldsymbol{\varepsilon}_N'] = diag_{i=1}^{N}[E(\varepsilon_{i,N}^2)] = diag_{i=1}^{N}(\sigma_{i,N}^2). \tag{3a}$$

In line with Kelejian and Prucha (2008), the innovations are allowed to depend on sample size, i.e., to form triangular arrays. Even if the innovations did not depend on N, $\mathbf{y}_N$ and $\mathbf{u}_N$ still would depend on N as can be seen from equations (2a) and (2b).

We assume further that the weighting matrices have the following properties:

Assumption 3.

The row and column sums of the matrices $\mathbf{W}_{r,N}$, $r = 1,...,R$, $\mathbf{M}_{s,N}$, $s = 1,...,S$, $(\mathbf{I}_N - \sum_{r=1}^{R} \lambda_r \mathbf{W}_{r,N})^{-1}$, and $(\mathbf{I}_N - \sum_{m=1}^{S} \rho_m \mathbf{M}_{m,N})^{-1}$ are bounded uniformly in absolute value. (See Remark A.1 in Appendix A for a definition of row and column sum boundedness.)

As Kelejian and Prucha (2008, p. 7) point out, Assumption 3 restricts the extent of neighborliness of the cross-sectional units on the one hand, and the degree of cross-sectional correlation between the model disturbances on the other hand. Such restrictions on the degree of permissible correlations are standard in virtually all large sample theory.

In light of equation (2b) and Remark A.1, Assumptions 2 and 3 imply that $E(\mathbf{u}_N) = 0$ and that the variance-covariance matrix of $\mathbf{u}_N$ is given by

$$\mathbf{\Omega}_{u,N} = E(\mathbf{u}_N \mathbf{u}'_N) = (\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N} \mathbf{M}_{m,N})^{-1} \mathbf{\Omega}_{\varepsilon,N} (\mathbf{I}_N - \sum_{s=1}^{R} \rho_{m,n} \mathbf{M}'_{m,N})^{-1} . \tag{3b}$$

III. GM Estimator for *S*-th Order Spatial Regressive Error Process

In the following, we extend the GM estimator for the spatial autoregressive parameter in Kelejian and Prucha (2008) to the case of an *S*-th order process. In this subsection, we only consider the process in equation (1c) for the disturbances $\mathbf{u}_N$, but not necessarily the one in equation (1a) for $\mathbf{y}_N$. We first derive the moment conditions defining the GM estimator of $\boldsymbol{\rho}_N$ for the case of heteroskedastic innovations and the optimal weighting of the moment conditions. We then prove consistency and derive the asymptotic distribution of the proposed GM estimator.

1. Moment Conditions and Definition of a GM Estimator for $\boldsymbol{\rho}_N$

Kelejian and Prucha (2008) use two moment conditions to derive a generalized moments (GM) estimator for a first-order spatial regressive process ($S = 1$). In case of an *S*-th order process, the GM estimators of the parameters $\rho_{1,N},...,\rho_{S,N}$ are obtained by recognizing that – under Assumptions 1 and 2 – the moment conditions used by Kelejian and Prucha (2008) hold for each matrix $\mathbf{M}_{s,N}$, $s = 1,...,S$. In particular, we define for each matrix $\mathbf{M}_{s,N}$, $s = 1,...,S$,

$$\bar{\boldsymbol{\varepsilon}}_{s,N} = \mathbf{M}_{s,N}\boldsymbol{\varepsilon}_N = \mathbf{M}_{s,N}(\mathbf{u}_N - \sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N}\mathbf{u}_N)\,. \tag{4}$$

The moment conditions $\mathrm{MC}_{1,s}$ to $\mathrm{MC}_{2,s}$, $s = 1,...,S$, associated with matrix $\mathbf{M}_{s,N}$ through equation (4), are given by

$$\mathrm{MC}_{1,s} \quad N^{-1}[E(\bar{\boldsymbol{\varepsilon}}'_{s,N}\bar{\boldsymbol{\varepsilon}}_{s,N}) - Tr\{\mathbf{M}_{s,N}[diag_{i=1}^{N}E(\varepsilon_{i,N}^2)]\mathbf{M}'_{s,N}\}] = 0\,, \tag{5a}$$

$$\mathrm{MC}_{2,s} \quad N^{-1}E(\bar{\boldsymbol{\varepsilon}}'_{s,N}\boldsymbol{\varepsilon}_N) = 0\,. \tag{5b}$$

$\mathrm{MC}_{1,s}$ and $\mathrm{MC}_{2,s}$ can be written alternatively as

$$\mathrm{MC}_{1,s} \quad N^{-1}E(\boldsymbol{\varepsilon}'_N\mathbf{A}_{1s,N}\boldsymbol{\varepsilon}_N) = 0\,, \tag{6a}$$

$$\mathrm{MC}_{2,s} \quad N^{-1}E(\boldsymbol{\varepsilon}'_N\mathbf{A}_{2s,N}\boldsymbol{\varepsilon}_N) = 0\,, \tag{6b}$$

where $\mathbf{A}_{1s,N} = \mathbf{M}'_{s,N}\mathbf{M}_{s,N} - diag_{i=1}^{N}(\mathbf{m}'_{.i,s,N}\mathbf{m}_{.i,s,N})$ with $\mathbf{m}_{.i,s,N}$ denoting the i-th column of $\mathbf{M}_{s,N}$, and $\mathbf{A}_{2s,N} = \mathbf{M}_{s,N}$. It is readily seen that the main diagonal elements of $\mathbf{A}_{1s,N}$ and $\mathbf{A}_{2s,N}$ are zero. Also, note that the row and column sums of $\mathbf{A}_{1s,N}$ and $\mathbf{A}_{2s,N}$ are uniformly bounded in absolute value in light of Remark A.1 in Appendix A.

From the specification of the error term in equation (1c), it follows that

$$\boldsymbol{\varepsilon}_N = \mathbf{u}_N - \sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N}\mathbf{u}_N = \mathbf{u}_N - \sum_{m=1}^{S}\rho_{m,N}\bar{\mathbf{u}}_{m,N} \quad \text{and} \tag{7a}$$

$$\bar{\boldsymbol{\varepsilon}}_{s,N} = \mathbf{M}_{s,N}\boldsymbol{\varepsilon} = \mathbf{M}_{s,N}(\mathbf{u}_N - \sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N}\mathbf{u}_N) = \bar{\mathbf{u}}_{s,N} - \sum_{m=1}^{S}\rho_{m,N}\bar{\bar{\mathbf{u}}}_{sm,N}\,, \tag{7b}$$

where we use the following definitions: $\bar{\mathbf{u}}_{s,N} = \mathbf{M}_{s,N}\mathbf{u}_N$, $\bar{\bar{\mathbf{u}}}_{sm,N} = \mathbf{M}_{s,N}\mathbf{M}_{m,N}\mathbf{u}_N = \mathbf{M}_{s,N}\bar{\mathbf{u}}_{m,N}$.

Substituting (7a) and (7b) into the moment conditions (6a) and (6b), we obtain a $2S$ equation system

$$\boldsymbol{\gamma}_N - \boldsymbol{\Gamma}_N\mathbf{b}_N = 0\,, \tag{8a}$$

where $\mathbf{b}_N$ is a $[2S + S(S\text{-}1)/2\,] \times 1$ vector, given by

$$\mathbf{b}_N = (\rho_{1,N},...,\rho_{S,N},\rho_{1,N}^2,...,\rho_{S,N}^2,\rho_{1,N}\rho_{2,N},...,\rho_{1,N}\rho_{S,N},...,\rho_{S-1,N}\rho_{S,N})'\,,$$

i.e., $\mathbf{b}_N$ contains S linear terms $\rho_{m,N}$ ($m = 1,...,S$), S quadratic terms $\rho_{m,N}^2$ ($m = 1,...,S$), and $S(S-1)/2$ cross products $\rho_{m,N}\rho_{l,N}$ (m = 1, …, S-1, $l = m+1$, … S).

$\boldsymbol{\gamma}_N$ is a $2S \times 1$ vector with elements $(\gamma_{i,N})$, i = 1, …, $2S$, and $\boldsymbol{\Gamma}_N$ is a $2S \times [2S + S(S\text{-}1)/2]$ matrix with elements $(\gamma_{ij,N})$, i = 1, …, $2S$, j = 1, …, $[2S + S(S\text{-}1)/2]$, whose elements will be defined below. The row-index of the elements $\boldsymbol{\gamma}_N$ and $\boldsymbol{\Gamma}_N$ will be chosen such that the equation system (8a) has the following order: the first two rows correspond to the moment conditions $MC_{1,1}$ and $MC_{2,1}$ associated with matrix $\mathbf{M}_{1,N}$ through (4); rows three and four correspond to $MC_{1,2}$ and $MC_{4,2}$ associated with matrix $\mathbf{M}_{2,N}$, and so forth; rows $(2S-1)$ and $2S$ correspond to $MC_{1,S}$ to $MC_{4,S}$ associated with the matrix $\mathbf{M}_{S,N}$. As a result, the equation system can also be written as

$$\begin{bmatrix} \boldsymbol{\gamma}_{1,N} \\ . \\ \boldsymbol{\gamma}_{S,N} \end{bmatrix} - \begin{bmatrix} \boldsymbol{\Gamma}_{1,N} \\ . \\ \boldsymbol{\Gamma}_{S,N} \end{bmatrix} \mathbf{b}_N = \mathbf{0}, \tag{8b}$$

where the 2×1 vectors $\boldsymbol{\gamma}_{s,N}$ and the $2 \times [2S + S(S\text{-}1)/2]$ matrices $\boldsymbol{\Gamma}_{s,N}$, $s = 1,...,S$ are the parts of the equation system (8a) associated with matrix $\mathbf{M}_{s,N}$ (and moment conditions MC_1 and MC_2). Note that the first two rows of equation system (8b) correspond to the equation system (6) in Kelejian and Prucha (2008), which is a special case of equation (8b) under $S = 1$.

The sample analogue of equation system (8a) is given by

$$\widetilde{\boldsymbol{\gamma}}_N - \widetilde{\boldsymbol{\Gamma}}_N \mathbf{b}_N = \vartheta_N(\boldsymbol{\rho}_N), \tag{9}$$

where $\widetilde{\boldsymbol{\gamma}}_N$ and $\widetilde{\boldsymbol{\Gamma}}_N$ are equal to $\boldsymbol{\gamma}_N$ and $\boldsymbol{\Gamma}_N$ with the expectations operator suppressed and the disturbances $\mathbf{u}_N$ replaced by (consistent) estimates $\widetilde{\mathbf{u}}_N$; $\vartheta_N(\boldsymbol{\rho}_N)$ can be viewed as a vector of regression residuals.

The GM estimates of the parameters $(\rho_{1,N},...,\rho_{S,N})$ are then defined as

$$\widetilde{\boldsymbol{\rho}}_N = \widetilde{\boldsymbol{\rho}}_N(\widetilde{\boldsymbol{\Theta}}_N) = \underset{-\mathbf{a}^\rho \le \boldsymbol{\rho} \le \mathbf{a}^\rho}{\arg\min} [(\widetilde{\boldsymbol{\gamma}}_N - \widetilde{\boldsymbol{\Gamma}}_N \mathbf{b})' \widetilde{\boldsymbol{\Theta}}_N (\widetilde{\boldsymbol{\gamma}} - \widetilde{\boldsymbol{\Gamma}}_N \mathbf{b})] = [\vartheta_N(\boldsymbol{\rho})' \widetilde{\boldsymbol{\Theta}}_N (\vartheta_N(\boldsymbol{\rho})], \tag{10}$$

i.e., the parameter estimates can be obtained from a (weighted) nonlinear least squares regression of $\tilde{\boldsymbol{\gamma}}_N$ on the columns of $\tilde{\boldsymbol{\Gamma}}_N$. The optimal choice and estimation of the weighting matrix $\boldsymbol{\Theta}_N$ will be discussed in more detail below.

In the following, we define the elements of $\boldsymbol{\gamma}_N$ and $\boldsymbol{\Gamma}_N$, grouped by the two moment conditions.

Moment Condition M_1 delivers $s = 1, \ldots S$ equations of equation system (8b), appearing in rows 1, 3, …, 2*S*-1. The corresponding elements of $\boldsymbol{\gamma}_N$ and $\boldsymbol{\Gamma}_N$ are given by

$$\gamma_{2(s-1)+1} = N^{-1}E\{\bar{\mathbf{u}}'_{s,N}\bar{\mathbf{u}}_{s,N} - Tr[\mathbf{M}_{s,N}diag_{i=1}^{N}(u_{i,N}^2)\mathbf{M}'_{s,N}]\}, \text{ or} \tag{11}$$
$$\gamma_{2(s-1)+1} = N^{-1}E(\mathbf{u}'_N\mathbf{A}_{1s,N}\mathbf{u}_N).$$

$$\gamma_{2(s-1)+1,m} = 2N^{-1}E\{\bar{\bar{\mathbf{u}}}'_{sm,N}\bar{\mathbf{u}}_{s,N} - Tr[\mathbf{M}_{s,N}diag_{i=1}^{N}(\bar{u}_{m,i,N}u_{i,N})\mathbf{M}'_{s,N}]\}, \text{ or}$$
$$\gamma_{2(s-1)+1,m} = 2N^{-1}E(\mathbf{u}'_N\mathbf{M}'_{m,N}\mathbf{A}_{1s,N}\mathbf{u}_N),\ m = 1,\ldots, S.$$

Note that $\bar{u}_N$ exhibits two subscripts: the first subscript *m* refers to the matrix by which $\mathbf{u}_N$ is premultiplied; the second subscript *i* refers to the unit of observation.

$$\gamma_{2(s-1)+1,S+m} = -N^{-1}E[\bar{\bar{\mathbf{u}}}'_{sm,N}\bar{\bar{\mathbf{u}}}_{sm,N} - Tr[\mathbf{M}_{s,N}diag_{i=1}^{N}(\bar{u}_{m,i,N}^2)\mathbf{M}'_{s,N}]\}, \text{, or}$$
$$\gamma_{2(s-1)+1,S+m} = -N^{-1}E(\mathbf{u}'_N\mathbf{M}'_{m,N}\mathbf{A}_{1s,N}\mathbf{M}_{m,N}\mathbf{u}_N),\ m = 1,\ldots, S.$$

$$\gamma_{2(s-1)+1,S(m+1)-m(m-1)/2+l-m} = -2N^{-1}E\{\bar{\bar{\mathbf{u}}}'_{sm,N}\bar{\bar{\mathbf{u}}}_{sl,N} - Tr[\mathbf{M}_{s,N}diag_{i=1}^{N}(\bar{u}_{l,i,N}\bar{u}_{m,i,N})\mathbf{M}'_{s,N}]\},$$
or
$$\gamma_{2(s-1)+1,S(m+1)-m(m-1)/2+l-m} = -2N^{-1}E(\mathbf{u}'_N\mathbf{M}'_{m,N}\mathbf{A}_{1s,N}\mathbf{M}_{l,N}\mathbf{u}_N),\ m = 1,\ldots,S-1,\ l = m+1,\ldots,S$$

Moment Condition M_2 delivers $s = 1,\ldots,S$ equations of system (8b), appearing in rows 2, 4, …., 2*S*. The corresponding elements of $\boldsymbol{\gamma}_N$ and $\boldsymbol{\Gamma}_N$ are given by

$$\gamma_{2(s-1)+2} = N^{-1}E(\mathbf{u}'_N\bar{\mathbf{u}}_{s,N}), \text{ or}$$
$$\gamma_{2(s-1)+2} = N^{-1}E(\mathbf{u}'_N\mathbf{A}_{2s,N}\mathbf{u}_N).$$

$$\gamma_{2(s-1)+2,m} = N^{-1}E[\mathbf{u}'_N\bar{\bar{\mathbf{u}}}_{sm,N} + \bar{\mathbf{u}}'_{m,N}\bar{\mathbf{u}}_{s,N}], \text{ associated with } \rho_m, \text{ or}$$
$$\gamma_{2(s-1)+2,m} = N^{-1}E[\mathbf{u}'_N\mathbf{M}'_{m,N}(\mathbf{A}'_{2s,N} + \mathbf{A}_{2s,N})\mathbf{u}_N],\ m = 1,\ldots, S.$$

$$\gamma_{2(s-1)+2,S+m} = -N^{-1}E[\overline{\overline{\mathbf{u}}}'_{sm,N}\overline{\mathbf{u}}_{m,N}] = -N^{-1}E[\overline{\mathbf{u}}'_{m,N}\overline{\overline{\mathbf{u}}}_{sm,N}], \text{ or}$$

$$\gamma_{2(s-1)+2,S+m} = -N^{-1}E(\mathbf{u}'_N\mathbf{M}'_{m,N}\mathbf{A}_{2s,N}\mathbf{M}_{m,N}\mathbf{u}_N),\ m=1,...,S.$$

$$\gamma_{2(s-1)+2,S(m+1)-m(m-1)/2+l-m} = -N^{-1}E(\overline{\overline{\mathbf{u}}}'_{sl,N}\overline{\mathbf{u}}_{m,N} + \overline{\overline{\mathbf{u}}}'_{sm,N}\overline{\mathbf{u}}_{l,N}), \text{ or}$$

$$\gamma_{2(s-1)+2,S(m+1)-m(m-1)/2+l-m} = = -N^{-1}E[\mathbf{u}'_N\mathbf{M}'_{l,N}(\mathbf{A}'_{2s,N} + \mathbf{A}_{2s,N})\mathbf{M}_{m,N}\mathbf{u}_N],\ m=1,...,S-1,$$

$l = m+1,...,S$.

2. Asymptotic Properties of the GM Estimator for $\boldsymbol{\rho}_N$

2.1 Consistency

In order to prove consistency, the following additional assumptions are required:

Assumption 4.

Let $\widetilde{u}_{i,N}$ denote the *i*-th element of $\widetilde{\mathbf{u}}_N$. We then assume that

$$\widetilde{u}_{i,N} - u_{i,N} = \mathbf{d}_{i.,N}\boldsymbol{\Delta}_N,$$

where $\mathbf{d}_{i.,N}$ is an $1\times P$ vector and $\boldsymbol{\Delta}_N$ is a $P\times 1$ vector. Let $d_{ij,N}$ be the *j*-th element of $\mathbf{d}_{i.,N}$. Then we assume that for some $\delta > 0$, $E\left|d_{ij,N}\right|^{2+\delta} \le c_d < \infty$, where c_d does not depend on *N*, and that $N^{1/2}\left\|\boldsymbol{\Delta}_N\right\| = O_p(1)$.

Assumption 4 will typically be fulfilled in linear spatial models, where the estimates of $\widetilde{u}_{i,N}$ are based on $N^{1/2}$-consistent estimates of the model parameters. This is not different from the first-order case (Kelejian and Prucha, 2008, p.11) and ensures that $\widetilde{\boldsymbol{\gamma}}_N$ and $\widetilde{\boldsymbol{\Gamma}}_N$ converge in probability to $\boldsymbol{\gamma}_N$ and $\boldsymbol{\Gamma}_N$. To be more specific, consider the linear model in equation (1a) without endogenous regressors. Then, $\mathbf{d}_{i.,N}$ is the *i*-th row of the regressor matrix $\mathbf{X}_N$ and $\boldsymbol{\Delta}_N$ denotes the difference between the parameter estimator and the true parameter values, i.e., $(\widetilde{\boldsymbol{\beta}}_N - \boldsymbol{\beta}_N)$. In that case, consistency of least squares ensures that Assumption 4 holds. As stated here, Assumption 4 will also be fulfilled in more general settings, e.g., if model (1a) contains endogenous variables (such as spatial lags of **y**) and is estimated using an instrumental variable procedure. Under certain conditions, Assumption 4 will also be satisfied if model (1a) involves a nonlinear specification (see Kelejian and Prucha, 2008, p. 12).

Assumption 5.

(a) The smallest eigenvalue of $\boldsymbol{\Gamma}'_N\boldsymbol{\Gamma}_N$ is uniformly bounded away from zero.

(b) $\widetilde{\boldsymbol{\Theta}}_N - \boldsymbol{\Theta}_N = o_p(1)$, where $\boldsymbol{\Theta}_N$ are $2S\times 2S$ nonstochastic symmetric, positive definite

matrices. (c) The largest eigenvalues of $\boldsymbol{\Theta}_N$ are uniformly bounded from above, and the smallest eigenvalues of $\boldsymbol{\Theta}_N$ are uniformly bounded away from zero.

As we will show in Appendix A, Assumption 5 also implies that the smallest eigenvalues of $\boldsymbol{\Gamma}'_N\boldsymbol{\Theta}_N\boldsymbol{\Gamma}_N$ are uniformly bounded away from zero, ensuring that the true parameter vector $\boldsymbol{\rho}_N$ is identifiable unique. By the equivalence of matrix norms, Assumption 5 also implies that $\boldsymbol{\Theta}_N$ and $\boldsymbol{\Theta}_N^{-1}$ are $O(1)$.

Assumptions 1 to 3 (maintained throughout) together with Assumptions 4 and 5 ensure consistency of the estimator $\widetilde{\boldsymbol{\rho}}_N$. We summarize this result in the following Theorem, which is proved in the Appendix C.

Theorem 1.[3]
Suppose Assumptions 1-5 hold. Then, provided the optimization space contains the parameter space, the GM estimators $\widetilde{\boldsymbol{\rho}}_N(\widetilde{\boldsymbol{\Theta}}_N) = [\widetilde{\rho}_{1,N}(\widetilde{\boldsymbol{\Theta}}_N),...,\widetilde{\rho}_{S,N}(\widetilde{\boldsymbol{\Theta}}_N)]'$ defined by (10) are consistent for $\rho_{1,N},...,\rho_{S,N}$, i.e.,

$$\widetilde{\rho}_{s,N} - \rho_{s,N} \xrightarrow{p} 0 \text{ as } N \to \infty,\ s = 1,...,S.$$

2.2 Asymptotic Distribution of GM Estimator for $\boldsymbol{\rho}_N$

To establish asymptotic normality of $\widetilde{\boldsymbol{\rho}}_N$, we need some additional assumptions.

Let $\mathbf{D}_N = (\mathbf{d}'_{1.,N},...,\mathbf{d}'_{N.,N})'$, with $\mathbf{d}_{i.,N}$ defined as in Assumption 4, such that $\widetilde{\mathbf{u}}_N - \mathbf{u}_N = \mathbf{D}_N\boldsymbol{\Delta}_N$.

Assumption 6.
For any real $N \times N$ matrix $\mathbf{A}_N$, whose row and column sums are bounded uniformly in absolute value, it holds that

$$N^{-1}\mathbf{D}'_N\mathbf{A}_N\mathbf{u}_N - N^{-1}E(\mathbf{D}'_N\mathbf{A}_N\mathbf{u}_N) = o_p(1).$$

A sufficient condition for Assumption 6 is, e.g., that the columns of $\mathbf{D}_N$ are of the form $\boldsymbol{\pi}_N + \boldsymbol{\Pi}_N\boldsymbol{\varepsilon}_N$, where the elements of $\boldsymbol{\pi}_N$ are uniformly bounded in absolute value and the row and column sums of $\boldsymbol{\Pi}_N$ are uniformly bounded in absolute value (compare Lemma C.2 in Kelejian and Prucha, 2008). This will be the case in many applications, e.g., for the model in

[3] It is assumed that $\widetilde{\boldsymbol{\rho}}_N$ exists and measurable. In the present setting, this is ensured, for example, by Lemma 2 in Jennrich (1969), which is a special case of Lemma 3.4 in Pötscher and Prucha (1997) when the parameter space is a compact subset of the Euclidian space.

equations (1) and (2), when $\mathbf{D}_N$ equals (the negative of) the matrix $\mathbf{Z}_N$ (compare Lemma 1 in Section IV).

Assumption 7.
Let $\mathbf{\Delta}_N$ be defined as in Assumption 4. Then

$$N^{1/2}\mathbf{\Delta}_N = N^{-1/2}\mathbf{T}_N'\boldsymbol{\varepsilon}_N + o_p(1),$$

where $\mathbf{T}_N$ is an $N \times P$-dimensional real nonstochastic matrix whose elements are uniformly bounded in absolute value. As remarked above, $\mathbf{\Delta}_N$ typically denotes the difference between the parameter estimates and the true parameter values. Assumption 7 will be satisfied by many estimators. In Section IV, we verify that this assumption hold when the model in equation (1) is estimated by two-stages least squares (TSLS).

We summarize the results regarding the asymptotic distribution of $\widetilde{\boldsymbol{\rho}}_N$ in the following theorem, which is proved in Appendix C.

Theorem 2. (Asymptotic Normality of $\widetilde{\boldsymbol{\rho}}_N$)
Let $\widetilde{\boldsymbol{\rho}}_N$ be the GM estimator defined by (10). Suppose Assumptions 1-7 hold and, furthermore, that $\lambda_{\min}(\mathbf{\Psi}_N) \geq c_{\mathbf{\Psi}}^* > 0$. Then, provided the optimization space contains the parameter space, we have

$$N^{1/2}(\widetilde{\boldsymbol{\rho}}_N - \boldsymbol{\rho}_N) = (\mathbf{J}_N'\mathbf{\Theta}_N\mathbf{J}_N)^{-1}\mathbf{J}_N'\mathbf{\Theta}_N\mathbf{\Psi}_N^{1/2}\boldsymbol{\xi}_N + o_p(1), \text{ with}$$

$$\mathbf{J}_N = \frac{\partial}{\partial \boldsymbol{\rho}'}\mathbf{\Gamma}_N\mathbf{b}_N = \mathbf{\Gamma}_N\mathbf{B}_N, \text{ and}$$

$$\boldsymbol{\xi}_N = \mathbf{\Psi}_N^{-1/2}\mathbf{v}_N \xrightarrow{d} N(0, \mathbf{I}_{2S}),$$

where $\mathbf{\Psi}_N = E[\mathbf{v}_N\mathbf{v}_N']$ and $\mathbf{\Psi}_N = (\mathbf{\Psi}_N^{1/2})(\mathbf{\Psi}_N^{1/2})'$.

Furthermore $N^{1/2}(\widetilde{\boldsymbol{\rho}}_N - \boldsymbol{\rho}_N) = O_p(1)$ and

$$\mathbf{\Omega}_{\widetilde{\rho}_N}(\mathbf{\Theta}_N) = (\mathbf{J}_N'\mathbf{\Theta}_N\mathbf{J}_N)^{-1}\mathbf{J}_N'\mathbf{\Theta}_N\mathbf{\Psi}_N\mathbf{\Theta}_N\mathbf{J}_N(\mathbf{J}_N'\mathbf{\Theta}_N\mathbf{J}_N)^{-1},$$

where $\mathbf{\Omega}_{\widetilde{\rho}_N}$ is positive definite.

Theorem 2 implies that the difference between the cumulative distribution function of $N^{1/2}(\widetilde{\boldsymbol{\rho}}_N - \boldsymbol{\rho}_N)$ and that of $N(\mathbf{0}, \boldsymbol{\Omega}_{\widetilde{\rho}_N})$ converges pointwise to zero, which justifies the use of the latter as an approximation of the former.[4]

The elements of $\mathbf{v}_N$ in Theorem 2 are given by

$$\mathbf{v}_N = \begin{bmatrix} \mathbf{v}_{1,N} \\ . \\ \mathbf{v}_{S,N} \end{bmatrix} = (\mathbf{v}_{s,N}) \;\; s = 1,...,S \text{, where} \tag{12}$$

$$\mathbf{v}_{s,N} = \begin{bmatrix} v_{1s,N} \\ v_{2s,N} \end{bmatrix} = N^{-1/2} \begin{bmatrix} [\tfrac{1}{2}\boldsymbol{\varepsilon}'_N(\mathbf{A}_{1s,N} + \mathbf{A}'_{1s,N})\boldsymbol{\varepsilon}_N + \mathbf{a}'_{1s,N}\boldsymbol{\varepsilon}_N] \\ [\tfrac{1}{2}\boldsymbol{\varepsilon}'_N(\mathbf{A}_{2s,N} + \mathbf{A}'_{2s,N})\boldsymbol{\varepsilon}_N + \mathbf{a}'_{2s,N}\boldsymbol{\varepsilon}_N] \end{bmatrix}, \;\; s = 1,...,S .$$

The $N \times 1$ vectors $\mathbf{a}_{1s,N}$ and $\mathbf{a}_{2s,N}$ are defined as

$$\mathbf{a}_{1s,N} = \mathbf{T}_N \boldsymbol{\alpha}_{1s,N} \tag{13a}$$

$$\mathbf{a}_{2s,N} = \mathbf{T}_N \boldsymbol{\alpha}_{2s,N} , \tag{13b}$$

where

$$\boldsymbol{\alpha}_{1s,N} = N^{-1} E[\mathbf{D}'_N(\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N}\mathbf{M}'_{m,N})(\mathbf{A}_{1s,N} + \mathbf{A}'_{1s,N})(\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N}\mathbf{M}_{m,N})\mathbf{u}_N]$$

$$\boldsymbol{\alpha}_{2s,N} = N^{-1} E[\mathbf{D}'_N(\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N}\mathbf{M}'_{m,N})(\mathbf{A}_{2s,N} + \mathbf{A}'_{2s,N})(\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N}\mathbf{M}_{m,N})\mathbf{u}_N] .$$

The $2S \times 2S$ (limiting) variance-covariance matrix of $\mathbf{v}_N$, denoted as $\boldsymbol{\Psi}_N$, takes the following form:

$$\boldsymbol{\Psi}_N = E(\mathbf{v}_N \mathbf{v}'_N) = E\begin{bmatrix} \mathbf{v}_{1,N}\mathbf{v}'_{1,N} & & \mathbf{v}_{1,N}\mathbf{v}'_{S,N} \\ & & \\ \mathbf{v}_{S,N}\mathbf{v}'_{1,N} & & \mathbf{v}_{S,N}\mathbf{v}'_{S,N} \end{bmatrix} = E(\mathbf{v}_{p,N}\mathbf{v}'_{q,N}) , \;\; p,q = 1,...,S . \tag{14}$$

It is made up of S^2 submatrices of dimension 2×2, defined as

[4] Compare Corollary F4 in Pötscher and Prucha (1997) (see Appendix B).

$$E(\mathbf{v}_{p,N}\mathbf{v}'_{q,N}) = E\begin{bmatrix} v_{1p,N}v'_{1q,N} & v_{1p,N}v'_{2q,N} \\ v_{2p,N}v'_{1q,N} & v_{2q,N}v'_{2q,N} \end{bmatrix}.$$

Hence, $\mathbf{\Psi}_N$ can also be written as

$$\mathbf{\Psi}_N = (\boldsymbol{\psi}_{pq,N}),\ p,q = 1,...,S,$$

where the 2×2 elements

$$\boldsymbol{\psi}_{pq,N} = \begin{bmatrix} \psi^{11}_{pq,N} & \psi^{12}_{pq,N} \\ \psi^{21}_{pq,N} & \psi^{22}_{pq,N} \end{bmatrix}$$

are defined as

$$\psi^{11}_{pq,N} = E(v_{1p,N}v'_{1q,N}) = \tag{15a}$$
$$= \frac{1}{2}N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}(a_{ij,1p,N} + a_{ji,1p,N})(a_{ij,1q,N} + a_{ji,1q,N})\sigma^2_{i,N}\sigma^2_{j,N} + N^{-1}\sum_{i=1}^{N}a_{i,1p,N}a_{i,1q,N}\sigma^2_{i,N},$$

$$\psi^{12}_{pq,N} = E(v_{1p,N}v'_{2q,N}) =$$
$$= \frac{1}{2}N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}(a_{ij,1p,N} + a_{ji,1p,N})(a_{ij,2q,N} + a_{ji,2q,N})\sigma^2_{i,N}\sigma^2_{j,N} + N^{-1}\sum_{i=1}^{N}a_{i,1p,N}a_{i,2q,N}\sigma^2_{i,N},$$

$$\psi^{21}_{pq,N} = E(v_{2p,N}v'_{1q,N}) =$$
$$= \frac{1}{2}N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}(a_{ij,2p,N} + a_{ji,2p,N})(a_{ij,1q,N} + a_{ji,1q,N})\sigma^2_{i,N}\sigma^2_{j,N} + N^{-1}\sum_{i=1}^{N}a_{i,2p,N}a_{i,1q,N}\sigma^2_{i,N},$$

$$\psi^{22}_{pq,N} = E(v_{2p,N}v'_{2q,N}) =$$
$$= \frac{1}{2}N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}(a_{ij,2p,N} + a_{ji,2p,N})(a_{ij,2q,N} + a_{ji,2q,N})\sigma^2_{i,N}\sigma^2_{j,N} + N^{-1}\sum_{i=1}^{N}a_{i,2p,N}a_{i,2q,N}\sigma^2_{i,N},$$

or, in matrix notation,

$$\psi^{11}_{pq,N} = \frac{1}{2}N^{-1}Tr[(\mathbf{A}_{1p,N} + \mathbf{A}'_{1p,N})\mathbf{\Sigma}_N(\mathbf{A}_{1q,N} + \mathbf{A}'_{1q,N})\mathbf{\Sigma}_N] + N^{-1}\mathbf{a}'_{1p,N}\mathbf{\Sigma}_N\mathbf{a}_{1q,N}, \tag{15b}$$
$$\psi^{12}_{pq,N} = \frac{1}{2}N^{-1}Tr[(\mathbf{A}_{1p,N} + \mathbf{A}'_{1p,N})\mathbf{\Sigma}_N(\mathbf{A}_{2q,N} + \mathbf{A}'_{2q,N})\mathbf{\Sigma}_N] + N^{-1}\mathbf{a}'_{1p,N}\mathbf{\Sigma}_N\mathbf{a}_{2q,N},$$
$$\psi^{21}_{pq,N} = \frac{1}{2}N^{-1}Tr[(\mathbf{A}_{2p,N} + \mathbf{A}'_{2p,N})\mathbf{\Sigma}_N(\mathbf{A}_{1q,N} + \mathbf{A}'_{1q,N})\mathbf{\Sigma}_N] + N^{-1}\mathbf{a}'_{2p,N}\mathbf{\Sigma}_N\mathbf{a}_{1q,N},$$

$$\psi_{pq,N}^{22} = \frac{1}{2} N^{-1} Tr[(\mathbf{A}_{2p,N} + \mathbf{A}'_{2p,N})\boldsymbol{\Sigma}_N(\mathbf{A}_{2q,N} + \mathbf{A}'_{2q,N})\boldsymbol{\Sigma}_N] + N^{-1}\mathbf{a}'_{2p,N}\boldsymbol{\Sigma}_N\mathbf{a}_{2q,N}.$$

Note that $\boldsymbol{\Omega}_{\tilde{\rho}_N}(\boldsymbol{\Psi}_N^{-1}) = (\mathbf{J}'_N\boldsymbol{\Psi}_N^{-1}\mathbf{J}_N)^{-1}$ and $\boldsymbol{\Omega}_{\tilde{\rho}_N}(\boldsymbol{\Theta}_N) - \boldsymbol{\Omega}_{\tilde{\rho}_N}(\boldsymbol{\Psi}_N^{-1})$ is positive semidefinite. Thus, using a consistent estimator of $\boldsymbol{\Psi}_N^{-1}$ (which will be derived below) as the weighting matrix $\boldsymbol{\Theta}_N$ leads to the efficient GM estimator. By assumption, $\lambda_{\min}(\boldsymbol{\Psi}_N) \geq c_{\boldsymbol{\Psi}}^{*} > 0$. Moreover, the elements of $\boldsymbol{\Psi}_N$ are uniformly bounded in absolute value, such that $\lambda_{\max}(\boldsymbol{\Psi}_N) \leq c_{\boldsymbol{\Psi}}^{**} < \infty$ by the equivalence of matrix norms. Hence, $\boldsymbol{\Psi}_N^{-1}$ automatically satisfies the assumptions made with respect to $\boldsymbol{\Theta}_N$ in Assumption 5. Note that $\boldsymbol{\Psi}_N$ is generally not identical to the variance-covariance matrix of the moment vector unless $\mathbf{a}_{1s,N}$ and $\mathbf{a}_{2s,N}$ are equal to zero. This is due to the fact that the GM estimator is based on estimated rather than the true disturbances and the presence of endogenous right-hand side variables included in equation (1). In the absence of an endogenous right-hand side variable, $\mathbf{a}_{1s,N} = \mathbf{a}_{2s,N} = \mathbf{0}$. Apart from this fact, the variance-covariance matrix of the GM estimator of $\boldsymbol{\rho}_N$ is of the usual 'sandwich form'.

2.3 Estimation of Variance-Covariance Matrix of $\boldsymbol{\rho}_N$

In the following we develop a consistent estimator for the variance-covariance matrix of $\tilde{\boldsymbol{\rho}}_N$. Define

$$\tilde{\mathbf{J}}_N = \tilde{\boldsymbol{\Gamma}}_N\tilde{\mathbf{B}}_N, \text{ and} \tag{16a}$$

$$\tilde{\boldsymbol{\Sigma}}_N = diag_{i=1}^{N}(\tilde{\varepsilon}_{i,N}^2), \tag{16b}$$

where $\tilde{\boldsymbol{\varepsilon}}_N = (\mathbf{I}_N - \sum_{m=1}^{S}\tilde{\rho}_{m,N}\mathbf{M}_{m,N})\tilde{\mathbf{u}}_N$.

We next specify an estimator for $\mathbf{a}_{1s,N} = \mathbf{T}_N\boldsymbol{\alpha}_{1s,N}$ and $\mathbf{a}_{2s,N} = \mathbf{T}_N\boldsymbol{\alpha}_{2s,N}$. The matrix $\mathbf{T}_N$ will in many applications be of the form

$$\mathbf{T}_N = \mathbf{F}_N\mathbf{P}_N \text{ with } \mathbf{F}_N = (\mathbf{I}_N - \sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N})\mathbf{H}_N \text{ or } \mathbf{F}_N = (\mathbf{I}_N - \sum_{m=1}^{S}\rho_{m,N}\mathbf{M}'_{m,N})^{-1}\mathbf{H}_N, \tag{17}$$

where $\mathbf{H}_N$ is a real nonstochastic $N \times P_*$ matrix of instruments, and $\mathbf{P}_N$ is a real nonstochastic $P_* \times P$ matrix, with P as in Assumption 7.

To be more specific, consider a TSLS estimator of the model in equation (1a). In that case, $\boldsymbol{\Delta}_N = (\tilde{\boldsymbol{\delta}}_N - \boldsymbol{\delta}_N)$ and the matrix $\mathbf{P}_N$ will be of the structure as defined above and can be

estimated consistently by some estimator $\widetilde{\mathbf{P}}_N$ (see Section IV). The estimators for $\mathbf{T}_N$ are defined as

$$\widetilde{\mathbf{T}}_N = \widetilde{\mathbf{F}}_N \widetilde{\mathbf{P}}_N \quad \text{with } \widetilde{\mathbf{F}}_N = (\mathbf{I}_N - \sum_{m=1}^{S} \widetilde{\rho}_{m,N} \mathbf{M}_{m,N}) \mathbf{H}_N \text{ or } \widetilde{\mathbf{F}}_N = (\mathbf{I}_N - \sum_{m=1}^{S} \widetilde{\rho}_{m,N} \mathbf{M}'_{m,N})^{+} \mathbf{H}_N . \tag{18}$$

The estimators of $\mathbf{a}_{1s,N} = \mathbf{T}_N \boldsymbol{\alpha}_{1s,N}$ and $\mathbf{a}_{2s,N} = \mathbf{T}_N \boldsymbol{\alpha}_{2s,N}$ are then given by

$$\widetilde{\mathbf{a}}_{1s,N} = \widetilde{\mathbf{T}}_N \widetilde{\boldsymbol{\alpha}}_{1s,N} \text{ and} \tag{19a}$$

$$\widetilde{\mathbf{a}}_{2s,N} = \widetilde{\mathbf{T}}_N \widetilde{\boldsymbol{\alpha}}_{2s,N}, \tag{19b}$$

with

$$\widetilde{\boldsymbol{\alpha}}_{1s,N} = N^{-1}[\mathbf{D}'_N (\mathbf{I}_N - \sum_{m=1}^{S} \widetilde{\rho}_{m,N} \mathbf{M}'_{m,N})(\mathbf{A}_{1s,N} + \mathbf{A}'_{1s,N})(\mathbf{I}_N - \sum_{m=1}^{S} \widetilde{\rho}_{m,N} \mathbf{M}_{m,N}) \widetilde{\mathbf{u}}_N] \text{ and} \tag{20a}$$

$$\widetilde{\boldsymbol{\alpha}}_{2s,N} = N^{-1}[\mathbf{D}'_N (\mathbf{I}_N - \sum_{m=1}^{S} \widetilde{\rho}_{m,N} \mathbf{M}'_{m,N})(\mathbf{A}_{2s,N} + \mathbf{A}'_{2s,N})(\mathbf{I}_N - \sum_{m=1}^{S} \widetilde{\rho}_{m,N} \mathbf{M}_{m,N}) \widetilde{\mathbf{u}}_N]. \tag{20b}$$

In matrix form, the elements of the estimated $2S \times 2S$ matrix $\widetilde{\boldsymbol{\Psi}}_N$ are defined as:

$$\widetilde{\psi}^{11}_{pq,N} = \frac{1}{2} N^{-1} Tr[(\mathbf{A}_{1p,N} + \mathbf{A}'_{1p,N}) \widetilde{\boldsymbol{\Sigma}}_N (\mathbf{A}_{1q,N} + \mathbf{A}'_{1q,N}) \widetilde{\boldsymbol{\Sigma}}_N] + N^{-1} \widetilde{\mathbf{a}}'_{1p,N} \widetilde{\boldsymbol{\Sigma}}_N \widetilde{\mathbf{a}}_{1q,N}, \tag{21}$$

$$\widetilde{\psi}^{12}_{pq,N} = \frac{1}{2} N^{-1} Tr[(\mathbf{A}_{1p,N} + \mathbf{A}'_{1p,N}) \widetilde{\boldsymbol{\Sigma}}_N (\mathbf{A}_{2q,N} + \mathbf{A}'_{2q,N}) \widetilde{\boldsymbol{\Sigma}}_N] + N^{-1} \widetilde{\mathbf{a}}'_{1p,N} \widetilde{\boldsymbol{\Sigma}}_N \widetilde{\mathbf{a}}_{2q,N},$$

$$\widetilde{\psi}^{21}_{pq,N} = \frac{1}{2} N^{-1} Tr[(\mathbf{A}_{2p,N} + \mathbf{A}'_{2p,N}) \widetilde{\boldsymbol{\Sigma}}_N (\mathbf{A}_{1q,N} + \mathbf{A}'_{1q,N}) \widetilde{\boldsymbol{\Sigma}}_N] + N^{-1} \widetilde{\mathbf{a}}'_{2p,N} \widetilde{\boldsymbol{\Sigma}}_N \widetilde{\mathbf{a}}_{1q,N},$$

$$\widetilde{\psi}^{22}_{pq,N} = \frac{1}{2} N^{-1} Tr[(\mathbf{A}_{2p,N} + \mathbf{A}'_{2p,N}) \widetilde{\boldsymbol{\Sigma}}_N (\mathbf{A}_{2q,N} + \mathbf{A}'_{2q,N}) \widetilde{\boldsymbol{\Sigma}}_N] + N^{-1} \widetilde{\mathbf{a}}'_{2p,N} \widetilde{\boldsymbol{\Sigma}}_N \widetilde{\mathbf{a}}_{2q,N},$$

for $p,q = 1,...,S$. Based on $\widetilde{\boldsymbol{\Psi}}_N$, we can now define the estimator for $\boldsymbol{\Omega}_{\tilde{\boldsymbol{\rho}}_N}$ as

$$\widetilde{\boldsymbol{\Omega}}_{\tilde{\boldsymbol{\rho}}_N}(\widetilde{\boldsymbol{\Theta}}_N) = (\widetilde{\mathbf{J}}'_N \widetilde{\boldsymbol{\Theta}}_N \widetilde{\mathbf{J}}_N)^{+} \widetilde{\mathbf{J}}'_N \widetilde{\boldsymbol{\Theta}}_N \widetilde{\boldsymbol{\Psi}}_N \widetilde{\boldsymbol{\Theta}}_N \widetilde{\mathbf{J}}_N (\widetilde{\mathbf{J}}'_N \widetilde{\boldsymbol{\Theta}}_N \widetilde{\mathbf{J}}_N)^{+}. \tag{22}$$

The following theorem establishes the consistency of $\widetilde{\boldsymbol{\Psi}}_N$ and $\widetilde{\boldsymbol{\Omega}}_{\tilde{\boldsymbol{\rho}}_N}$.

Theorem 3. (Variance-Covariance Matrix Estimation).
Suppose all of the assumptions of Theorem 2, apart from Assumption 5, hold and that additionally all of the fourth moments of the elements of $\mathbf{D}_N$ are uniformly bounded. Suppose

furthermore (a) that the elements of the nonstochastic matrices $\mathbf{H}_N$ are uniformly bounded in absolute value, (b) $\sup_N \sum_{s=1}^{S} |\rho_{s,N}| < 1$ and that the row and column sums of $\mathbf{M}_N$ are bounded uniformly in absolute value by one and some finite constant, respectively, and (c) $\widetilde{\mathbf{P}}_N - \mathbf{P}_N = o_p(1)$ with $\mathbf{P}_N = O(1)$. Then,

$$\widetilde{\mathbf{\Psi}}_N - \mathbf{\Psi}_N = o_p(1) \text{ and } \widetilde{\mathbf{\Psi}}_N^{-1} - \mathbf{\Psi}_N^{-1} = o_p(1).$$

Furthermore, if Assumption 5 holds, then also

$$\widetilde{\mathbf{\Omega}}_{\widetilde{\rho}_N} - \mathbf{\Omega}_{\widetilde{\rho}_N} = o_p(1).$$

Remark 1.
As in Kelejian and Prucha (2008, p. 17), Theorem 3 also holds, if $\widetilde{\boldsymbol{\rho}}_N$ is replaced by some other estimator $N^{1/2}(\breve{\boldsymbol{\rho}}_N - \boldsymbol{\rho}_N) = O_p(1)$. In case that $\mathbf{F}_N = (\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N}\mathbf{M}_{m,N})\mathbf{H}_N$, condition (b) can be dropped. The consistency result for $\widetilde{\mathbf{\Psi}}_N^{-1}$ verifies that this estimator for $\mathbf{\Psi}_N^{-1}$ can indeed by used in the formulation of an efficient GM estimator.

3. Joint Distribution of the GM Estimator for $\boldsymbol{\rho}_N$ and Estimators of Other Model Parameters

Note that $N^{1/2}(\widetilde{\boldsymbol{\rho}}_N - \boldsymbol{\rho}_N)$ depends on a vector of linear quadratic forms in the innovations $\boldsymbol{\varepsilon}_N$ plus a term of order $o_p(1)$. By Assumption 7, $N^{1/2}\boldsymbol{\Delta}_N$ is asymptotically linear in $\boldsymbol{\varepsilon}_N$. Hence, the joint distribution of the vector $[N^{1/2}\boldsymbol{\Delta}_N', N^{1/2}(\widetilde{\boldsymbol{\rho}}_N - \boldsymbol{\rho}_N)']'$ can be derived invoking the central limit theorem for vectors of quadratic forms by Kelejian and Prucha (2008); see Appendix B.

Consider the $(P_* + 2S)\times 1$ vector of linear and linear quadratic forms in $\boldsymbol{\varepsilon}_N$:

$$\mathbf{w}_N = \begin{bmatrix} N^{-1/2}\mathbf{F}_N'\boldsymbol{\varepsilon}_N \\ \mathbf{v}_N \end{bmatrix}. \tag{23}$$

Using Lemma A.1 in Kelejian and Prucha (2008) (see Appendix B), its variance-covariance matrix is of dimension $(P_* + 2S)\times(P_* + 2S)$ and given by

$$Var(\mathbf{w}_N) = \mathbf{\Psi}_{o,N} = E\begin{bmatrix} N^{-1}\mathbf{F}_N'\boldsymbol{\varepsilon}_N\boldsymbol{\varepsilon}_N'\mathbf{F}_N & N^{-1/2}\mathbf{F}_N'\boldsymbol{\varepsilon}_N\mathbf{v}_N' \\ N^{-1/2}\mathbf{v}_N\boldsymbol{\varepsilon}_N\mathbf{F}_N & \mathbf{v}_N\mathbf{v}_N' \end{bmatrix} = \begin{bmatrix} \mathbf{\Psi}_{\Delta\Delta,N} & \mathbf{\Psi}_{\Delta\rho,N} \\ \mathbf{\Psi}_{\Delta\rho,N}' & \mathbf{\Psi}_N \end{bmatrix}, \tag{24}$$

where

$$\boldsymbol{\Psi}_{\Delta\Delta,N} = N^{-1}\mathbf{F}_N'\boldsymbol{\Sigma}_N\mathbf{F}_N \text{ and}$$
$$\boldsymbol{\Psi}_{\Delta\rho,N} = N^{-1}E(\mathbf{F}_N'\boldsymbol{\varepsilon}_N\mathbf{v}_N') = N^{-1}\mathbf{F}_N'\boldsymbol{\Sigma}_N(\mathbf{a}_{11,N},\mathbf{a}_{21,N},....,\mathbf{a}_{1S,N},\mathbf{a}_{2S,N}), \text{ and}$$

$\boldsymbol{\Psi}_N$ is defined in equation (14).

As we demonstrate in Appendix C, the matrix $\boldsymbol{\Psi}_{o,N}$ can be estimated consistently by

$$\widetilde{\boldsymbol{\Psi}}_{o,N} = \begin{bmatrix} \widetilde{\boldsymbol{\Psi}}_{\Delta\Delta,N} & \widetilde{\boldsymbol{\Psi}}_{\Delta\rho,N} \\ \widetilde{\boldsymbol{\Psi}}_{\Delta\rho,N}' & \widetilde{\boldsymbol{\Psi}}_N \end{bmatrix}, \text{ where} \tag{25}$$
$$\widetilde{\boldsymbol{\Psi}}_{\Delta\Delta,N} = N^{-1}\widetilde{\mathbf{F}}_N'\widetilde{\boldsymbol{\Sigma}}_N\widetilde{\mathbf{F}}_N,$$
$$\widetilde{\boldsymbol{\Psi}}_{\Delta\rho,N} = N^{-1}\widetilde{\mathbf{F}}_N'\widetilde{\boldsymbol{\Sigma}}_N(\widetilde{\mathbf{a}}_{11,N},\widetilde{\mathbf{a}}_{21,N},....,\widetilde{\mathbf{a}}_{1S,N},\widetilde{\mathbf{a}}_{2S,N}), \text{ and}$$

$\widetilde{\boldsymbol{\Psi}}_N$ is defined in equation (21).

Regarding the joint limiting distribution of $N^{1/2}(\widetilde{\boldsymbol{\rho}}_N - \boldsymbol{\rho}_N)$ and $N^{1/2}\boldsymbol{\Delta}_N$, we now have the following result:

Theorem 4. (Joint Distribution of $\widetilde{\boldsymbol{\rho}}_N$ and Other Model Parameters)
Suppose all assumptions used in Theorem 3 hold and $\lambda_{\min}(\boldsymbol{\Psi}_{o,N}) \geq c^*_{\boldsymbol{\Psi}_o} > 0$. Then,

$$\begin{bmatrix} N^{1/2}\boldsymbol{\Delta}_N \\ N^{1/2}(\widetilde{\boldsymbol{\rho}}_N - \boldsymbol{\rho}_N) \end{bmatrix} = \begin{bmatrix} \mathbf{P}_N' & \mathbf{0} \\ \mathbf{0} & (\mathbf{J}_N'\boldsymbol{\Theta}_N\mathbf{J}_N)^{-1}\mathbf{J}_N'\boldsymbol{\Theta}_N \end{bmatrix}\boldsymbol{\Psi}_{o,N}^{1/2}\boldsymbol{\xi}_{o,N} + o_p(1), \text{ with} \tag{26}$$
$$\boldsymbol{\xi}_{o,N} = \boldsymbol{\Psi}_{o,N}^{-1/2}(N^{-1/2}\boldsymbol{\varepsilon}_N'\mathbf{F}_N, \mathbf{v}_N')' \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_{P^*+2S}).$$

Furthermore, let

$$\boldsymbol{\Omega}_{o,N} = \begin{bmatrix} \mathbf{P}_N' & \mathbf{0} \\ \mathbf{0} & (\mathbf{J}_N'\boldsymbol{\Theta}_N\mathbf{J}_N)^{-1}\mathbf{J}_N'\boldsymbol{\Theta}_N \end{bmatrix}\boldsymbol{\Psi}_{o,N}\begin{bmatrix} \mathbf{P}_N & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Theta}_N\mathbf{J}_N(\mathbf{J}_N'\boldsymbol{\Theta}_N\mathbf{J}_N)^{-1} \end{bmatrix}, \text{ and} \tag{27}$$

$$\widetilde{\boldsymbol{\Omega}}_{o,N} = \begin{bmatrix} \widetilde{\mathbf{P}}_N' & \mathbf{0} \\ \mathbf{0} & (\widetilde{\mathbf{J}}_N'\widetilde{\boldsymbol{\Theta}}_N\widetilde{\mathbf{J}}_N)^{+}\widetilde{\mathbf{J}}_N'\widetilde{\boldsymbol{\Theta}}_N \end{bmatrix}\widetilde{\boldsymbol{\Psi}}_{o,N}\begin{bmatrix} \widetilde{\mathbf{P}}_N & \mathbf{0} \\ \mathbf{0} & \widetilde{\boldsymbol{\Theta}}_N\widetilde{\mathbf{J}}_N(\widetilde{\mathbf{J}}_N'\widetilde{\boldsymbol{\Theta}}_N\widetilde{\mathbf{J}}_N)^{+} \end{bmatrix}. \tag{28}$$

Then,

$$\widetilde{\boldsymbol{\Psi}}_{o,N} - \boldsymbol{\Psi}_{o,N} = o_p(1),\ \widetilde{\boldsymbol{\Omega}}_{o,N} - \boldsymbol{\Omega}_{o,N} = o_p(1), \text{ and } \boldsymbol{\Psi}_{o,N} = O(1),\ \boldsymbol{\Omega}_{o,N} = O(1).$$

Theorem 4 implies that the difference between the joint cumulative distribution function of $[N^{1/2}\boldsymbol{\Delta}_N', N^{1/2}(\widetilde{\boldsymbol{\rho}}_N - \boldsymbol{\rho}_N)']'$ and that of $N(\mathbf{0}, \boldsymbol{\Omega}_{o,N})$ converges pointwise to zero, which justifies the use of the latter distribution as an approximation of the former. The theorem also states that $\widetilde{\boldsymbol{\Omega}}_{o,N}$ is a consistent estimator of $\boldsymbol{\Omega}_{o,N}$. The proof of Theorem 4 is given in Appendix C.

Remark 2.

Theorem 4 can also be used to obtain the joint distribution of $(\widetilde{\boldsymbol{\rho}}_N - \boldsymbol{\rho}_N)$ and some other estimator $\boldsymbol{\Delta}_N^*$, where $N^{1/2}\boldsymbol{\Delta}_N^* = N^{1/2}\mathbf{T}_N^{*\prime}\boldsymbol{\varepsilon}_N + o_p(1)$, $\mathbf{T}_N^* = \mathbf{F}_N^*\mathbf{P}_N^*$, $\widetilde{\mathbf{T}}_N^* = \widetilde{\mathbf{F}}_N^*\widetilde{\mathbf{P}}_N^*$, assuming that analogous assumptions are maintained for this estimator. In particular, the results remain valid, but with $\boldsymbol{\Psi}_{\Delta\Delta,N} = N^{-1}\mathbf{F}_N^{*\prime}\boldsymbol{\Sigma}_N\mathbf{F}_N^*$, $\boldsymbol{\Psi}_{\Delta\rho,N} = N^{-1}\mathbf{F}_N^{*\prime}\boldsymbol{\Sigma}_N(\mathbf{a}_{11,N}, \mathbf{a}_{21,N}, ..., \mathbf{a}_{1S,N}, \mathbf{a}_{2S,N})$, $\widetilde{\boldsymbol{\Psi}}_{\Delta\Delta,N} = N^{-1}\widetilde{\mathbf{F}}_N^{*\prime}\widetilde{\boldsymbol{\Sigma}}_N\widetilde{\mathbf{F}}_N^*$,
$\widetilde{\boldsymbol{\Psi}}_{\Delta\rho,N} = N^{-1}\widetilde{\mathbf{F}}_N^{*\prime}\widetilde{\boldsymbol{\Sigma}}_N(\widetilde{\mathbf{a}}_{11,N}, \widetilde{\mathbf{a}}_{21,N}, ..., \widetilde{\mathbf{a}}_{1S,N}, \widetilde{\mathbf{a}}_{2S,N})$, and with $\mathbf{P}_N, \widetilde{\mathbf{P}}_N$ replaced by $\mathbf{P}_N^*, \widetilde{\mathbf{P}}_N^*$.

IV. Two-Stages Least Squares (TSLS) Estimator for $\boldsymbol{\delta}_N$

1. Instruments

It is evident from model (1), that $E(\overline{\mathbf{Y}}_N\mathbf{u}_N') \neq 0$. In line with Kelejian and Prucha (2008), we consider a TSLS procedure to obtain consistent estimates of the parameters $\boldsymbol{\delta}_N$.

The following assumptions are maintained.

Assumption 8.

The regressor matrix $\mathbf{X}_N$ has full column rank (for N large enough). Furthermore, the elements of $\mathbf{X}_N$ are uniformly bounded in absolute value.

Assumption 9.

The instrument matrix $\mathbf{H}_N$ has full column rank $P_* \geq K + R$ (for N large enough). Furthermore, the elements of $\mathbf{H}_N$ are uniformly bounded in absolute value.

Assumption 10.

The instruments $\mathbf{H}_N$ satisfy:

$\mathbf{Q_{HH}} = \lim_{N\to\infty}(N^{-1}\mathbf{H}_N'\mathbf{H}_N)$ is finite and nonsingular.

$\mathbf{Q_{HZ}} = \operatorname{plim}_{N\to\infty}(N^{-1}\mathbf{H}_N'\mathbf{Z}_N)$ is finite and has full column rank.

Regarding the choice of instruments, note that

$$E(\sum_{r=1}^{R}\mathbf{W}_{r,N}\mathbf{y}_N)=\sum_{r=1}^{R}\mathbf{W}_{r,N}E(\mathbf{y}_N)=\sum_{r=1}^{R}\mathbf{W}_{r,N}E[(\mathbf{I}-\sum_{r'=1}^{R}\lambda_{r',N}\mathbf{W}_{r',N})^{-1}\mathbf{X}_N\boldsymbol{\beta}_N]$$
$$=\sum_{r=1}^{R}\mathbf{W}_{r,N}[\mathbf{I}_N+\sum_{i=1}^{\infty}(\sum_{r'=1}^{R}\lambda_{r',N}\mathbf{W}_N)^i]\mathbf{X}_N\boldsymbol{\beta}_N, \tag{29}$$

provided that $\left\|\sum_{r'=1}^{R}\lambda_{r',N}\mathbf{W}_{r',N}\right\|<1$ for some matrix norm $\|\cdot\|$ (compare Horn and Johnson, 1985, p. 301). The instrument matrices $\mathbf{H}_N$ are used to instrument $\mathbf{Z}_N=(\mathbf{X}_N,\overline{\mathbf{Y}}_N)$ in terms of their predicted values from a least squares regression on $\mathbf{H}_N$, i.e., $\hat{\mathbf{Z}}_N=\mathbf{P}_{\mathbf{H}_N}\mathbf{Z}_N$, where $\mathbf{P}_{\mathbf{H}_N}=\mathbf{H}_N(\mathbf{H}'_N\mathbf{H}_N)^{-1}\mathbf{H}'_N$. In light of (29) it is reasonable to select $\mathbf{H}_N$ to include $\mathbf{X}_N$ and a subset of the linearly independent columns of terms of the sum

$$\sum_{i=1}^{Q}(\sum_{r'=1}^{R}\mathbf{W}_{r',N})^i\mathbf{X}_N, \tag{30}$$

where Q is some predefined constant.[5]

Note that such a choice of $\mathbf{H}_N$ implies that Assumption 9 will be fulfilled (by Assumptions 3 and 8). This choice also ensures that $\mathbf{X}_N$ is instrumented by itself.

2. Definition of TSLS Estimator and Asymptotic Results

As in Kelejian and Prucha (2008), estimation of the model in equation (1) proceeds in three steps. In the first step, model (1a) is estimated by TSLS using the instruments $\mathbf{H}_N$. In the second step, the spatial regressive parameters $\rho_{1,N},...,\rho_{S,N}$ can be estimated using the GM estimator defined in Section III, based on consistent estimates of $\mathbf{u}_N$ from the first step. In the third step, model (1a) is re-estimated by feasible generalized two-stages least squares (FGTSLS), which is equivalent to performing a TSLS estimation on a transformed version of equation (1). We outline each of these steps in more detail in the following.

The TSLS estimator of model (1a) is defined as

$$\tilde{\boldsymbol{\delta}}_N=(\hat{\mathbf{Z}}'_N\mathbf{Z}_N)^{-1}\hat{\mathbf{Z}}'_N\mathbf{y}_N, \text{ where} \tag{31}$$

[5] Kelejian, Prucha, and Yuzefovich (2004) consider alternative sets of instruments in the estimation of SARAR(1,1) models. Their Monte Carlo simulation results suggest that choosing $Q=2$ will be sufficient in many applications.

$$\hat{\mathbf{Z}}_N = \mathbf{P}_{\mathbf{H}_N}\mathbf{Z}_N = (\mathbf{X}_N, \hat{\overline{\mathbf{Y}}}_N) \text{ and} \tag{32a}$$

$$\hat{\overline{\mathbf{Y}}}_N = \mathbf{P}_{\mathbf{H}_N}\overline{\mathbf{Y}}_N . \tag{32b}$$

In the second step, the parameters $\rho_{s,N}$, $s = 1,...,S$, are estimated using the GM estimator defined in equation (10), based on the first step residuals $\tilde{\mathbf{u}}_N = \mathbf{y}_n - \mathbf{Z}_N\tilde{\boldsymbol{\delta}}_N$. As above these estimators are denoted as $\tilde{\rho}_{s,N}$, $s = 1,...,S$.

Lemma 1 shows that the various assumptions maintained in Section III are automatically satisfied by the TSLS estimator $\tilde{\boldsymbol{\delta}}_N$ and the corresponding residuals $\tilde{\mathbf{u}}_N$.

Lemma 1.[6]

Suppose that Assumptions 1-3 and 8-10 hold, and that $\sup_N \|\boldsymbol{\beta}_N\| \leq b < \infty$. Let $\mathbf{D}_N = -\mathbf{Z}_N$, then, the fourth moments of the elements of $\mathbf{D}_N$ are uniformly bounded in absolute value, Assumption 6 holds, and

(a) $N^{1/2}(\tilde{\boldsymbol{\delta}}_N - \boldsymbol{\delta}_N) = N^{-1/2}\mathbf{T}_N'\boldsymbol{\varepsilon}_N + o_p(1)$ with $\mathbf{T}_N = \mathbf{F}_N\mathbf{P}_N$ and where

$\mathbf{P}_N = \mathbf{Q}_{\mathbf{HH}}^{-1}\mathbf{Q}_{\mathbf{HZ}}(\mathbf{Q}_{\mathbf{HZ}}'\mathbf{Q}_{\mathbf{HH}}^{-1}\mathbf{Q}_{\mathbf{HZ}})^{-1}$ and

$\mathbf{F}_N = (\mathbf{I}_N - \sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N}')^{-1}\mathbf{H}_N$;

(b) $N^{-1/2}\mathbf{T}_N'\boldsymbol{\varepsilon}_N = O_p(1)$;

(c) $\mathbf{P}_N = O_p(1)$ and $\tilde{\mathbf{P}}_N - \mathbf{P}_N = o_p(1)$ for

$\tilde{\mathbf{P}}_N = (N^{-1}\mathbf{H}_N'\mathbf{H}_N)^{-1}(N^{-1}\mathbf{H}_N'\mathbf{Z}_N)[(N^{-1}\mathbf{Z}_N'\mathbf{H}_N)(N^{-1}\mathbf{H}_N'\mathbf{H}_N)^{-1}(N^{-1}\mathbf{H}_N'\mathbf{Z}_N)]^{-1}$.

The condition $\sup_N \|\boldsymbol{\beta}_N\| \leq b < \infty$ is trivially satisfied if $\boldsymbol{\beta}_N = \boldsymbol{\beta}$. Note that (a) and (b) together imply that $\tilde{\boldsymbol{\delta}}_N$ is a $N^{1/2}$-consistent estimator of $\boldsymbol{\delta}_N$.

Regarding Assumption 4, we now have $\tilde{\mathbf{u}}_N - \mathbf{u}_N = \mathbf{D}_N\boldsymbol{\Delta}_N$, where $\mathbf{D}_N = -\mathbf{Z}_N$ and $\boldsymbol{\Delta}_N = \tilde{\boldsymbol{\delta}}_N - \boldsymbol{\delta}_N$. Lemma 1 shows that under Assumptions 1-3 and 8-10 the TSLS residuals automatically satisfy the conditions postulated in Assumptions 4, 6, and 7 with respect to $\mathbf{D}_N$, $\boldsymbol{\Delta}_N$, and $\mathbf{T}_N$. Hence, Theorems 1 and 2 apply to the GM estimator $\tilde{\boldsymbol{\rho}}_N$, which is based on TSLS residuals. The Lemma also establishes that the elements of $\mathbf{D}_N$ are uniformly bounded in absolute value, gives explicit expressions for $\mathbf{P}_N$ and $\tilde{\mathbf{P}}_N$, and verifies that the conditions concerning these matrices made in Theorems 3 and 4 are fulfilled. Hence, Theorems 3 and 4

[6] The above Lemma corresponds to Lemma 3 in Kelejian and Prucha (2008) and is adapted here to apply to the higher-order case.

cover the GM estimator $\widetilde{\boldsymbol{\rho}}_N$ and the TSLS estimator $\widetilde{\boldsymbol{\delta}}_N$. In particular, Theorem 4 gives the joint limiting distribution of $N^{1/2}(\widetilde{\boldsymbol{\rho}}_N - \boldsymbol{\rho}_N)$ and $N^{1/2}(\widetilde{\boldsymbol{\delta}}_N - \boldsymbol{\delta}_N)$, where $\mathbf{D}_N = -\mathbf{Z}_N$, the matrices $\mathbf{P}_{N,}\mathbf{F}_N, \widetilde{\mathbf{P}}_N$ are as in Lemma 1, and $\widetilde{\mathbf{F}}_N = (\mathbf{I}_N - \sum_{m=1}^{S} \widetilde{\rho}_{m,N}\mathbf{M}'_{m,N})^{+}\mathbf{H}_N$.

We now turn to the third step. A Cochrane-Orcutt transformation to (1) is:

$$\mathbf{y}^*_N = \mathbf{Z}^*_N\boldsymbol{\delta}_N + \boldsymbol{\varepsilon}_N, \text{ where} \tag{33}$$

$$\mathbf{y}^*_N = (\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N}\mathbf{M}_{m,N})\mathbf{y}_N,$$

$$\mathbf{Z}^*_N = (\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N}\mathbf{M}_{m,N})\mathbf{Z}_N, \text{ and}$$

$$\mathbf{u}^*_N = (\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N}\mathbf{M}_{m,N})\mathbf{u}_N = \boldsymbol{\varepsilon}_N.$$

The FGTSLS estimator, denoted as $\hat{\hat{\boldsymbol{\delta}}}_N$, is then obtained as a two-stages least squares estimator applied to the transformed model (33), using the transformed instruments $\mathbf{H}^*_N = (\mathbf{I} - \sum_{m=1}^{S} \rho_{m,N}\mathbf{M}_{m,N})\mathbf{H}_N$, after replacing $\boldsymbol{\rho}_N$ by $\widetilde{\boldsymbol{\rho}}_N$, i.e.,

$$\hat{\hat{\boldsymbol{\delta}}}_N = (\hat{\hat{\mathbf{Z}}}^{*\prime}_N \widetilde{\mathbf{Z}}^*_N)^{1}\ \hat{\hat{\mathbf{Z}}}^{*\prime}_N \widetilde{\mathbf{y}}^*_N, \text{ where} \tag{34}$$

$$\hat{\hat{\mathbf{Z}}}^*_N = \mathbf{P}_{\widetilde{\mathbf{H}}^*_N}\widetilde{\mathbf{Z}}^*_N, \text{ with } \mathbf{P}_{\mathbf{H}_N} = \widetilde{\mathbf{H}}^*_N(\widetilde{\mathbf{H}}^{*\prime}_N \widetilde{\mathbf{H}}^*_N)^{-1}\widetilde{\mathbf{H}}^{*\prime}_N \text{ and } \widetilde{\mathbf{H}}^*_N = (\mathbf{I} - \sum_{m=1}^{S} \widetilde{\rho}_{m,N}\mathbf{M}_{m,N})\mathbf{H}_N,$$

$$\widetilde{\mathbf{Z}}^*_N = (\mathbf{I}_N - \sum_{m=1}^{S} \widetilde{\rho}_{m,N}\mathbf{M}_{m,N})\mathbf{Z}_N \text{ with,}$$

$$\widetilde{\mathbf{y}}^*_N = (\mathbf{I}_N - \sum_{m=1}^{S} \widetilde{\rho}_{m,N}\mathbf{M}_{m,N})\mathbf{y}_N.$$

The advantage of this approach as compared to the use of heteroskedasticity-and-autocorrelation-consistent estimates is that joint hypotheses about $\boldsymbol{\lambda}_N$ and $\boldsymbol{\rho}_N$ may be formulated and tested.

Kelejian and Prucha (2008) and Arraiz, Drukker, Kelejian and Prucha (2007) use the untransformed instrument matrix $\mathbf{H}_N$ in the FGTSLS estimation of SARAR(1,1) models. While this choice does not affect consistency, it has implications for the efficiency of the estimates. In light of (29), the ideal instruments matrix for $\mathbf{W}\overline{\mathbf{Y}}^*$ in the transformed model is given by $\mathbf{H}^*_N = (\mathbf{I} - \sum_{m=1}^{S} \rho_{m,N}\mathbf{M}_{m,N})\mathbf{H}_N$. In fact, the Monte Carlos analysis below suggests that

using the estimates of the transformed instruments $\widetilde{\mathbf{H}}_N^* = (\mathbf{I} - \sum_{m=1}^{S} \widetilde{\rho}_{m,N} \mathbf{M}_{m,N}) \mathbf{H}_N$ instead of $\mathbf{H}_N$ leads to smaller standard errors and produces slightly better results in small samples, in particular, with respect to the size of tests.

Lemma 2 shows that the various assumptions maintained in Section III are automatically satisfied by the (feasible) generalized TSLS estimator $\breve{\tilde{\boldsymbol{\delta}}}_N$ and the corresponding residuals.

Lemma 2.[7]

Suppose the Assumptions of Lemma 1 hold[8] and let $\breve{\tilde{\boldsymbol{\delta}}}_N$ be defined as in equation (34), where $\breve{\boldsymbol{\rho}}_N$ is any $N^{1/2}$-consistent estimator of $\boldsymbol{\rho}_N$ (such as the GM estimator $\widetilde{\boldsymbol{\rho}}_N$ based on TSLS residuals). Then

a) $N^{1/2}[\breve{\tilde{\boldsymbol{\delta}}}_N(\breve{\boldsymbol{\rho}}_N) - \boldsymbol{\delta}_N] = N^{-1/2}\mathbf{T}_N^{*\prime}\boldsymbol{\varepsilon}_N + o_p(1)$ with $\mathbf{T}_N^* = \mathbf{F}_N^*\mathbf{P}_N^*$ and where

$\mathbf{P}_N^* = \mathbf{Q}_{\mathbf{H}^*\mathbf{H}^*}^{-1}\mathbf{Q}_{\mathbf{H}^*\mathbf{Z}^*}[\mathbf{Q}'_{\mathbf{H}^*\mathbf{Z}^*}\mathbf{Q}_{\mathbf{H}^*\mathbf{H}^*}^{-1}\mathbf{Q}_{\mathbf{H}^*\mathbf{Z}^*}]^{-1}$ and $\mathbf{F}_N^* = \mathbf{H}_N^*$;

(b) $N^{-1/2}\mathbf{T}_N^{*\prime}\boldsymbol{\varepsilon}_N = O_p(1)$;

(c) $\mathbf{P}_N^* = O(1)$ and $\breve{\mathbf{P}}_N^* - \mathbf{P}_N^* = o_p(1)$ for

$\breve{\mathbf{P}}_N^* = (N^{-1}\breve{\mathbf{H}}_N^{*\prime}\breve{\mathbf{H}}_N^*)^{-1}(N^{-1}\breve{\mathbf{H}}_N^{*\prime}\breve{\mathbf{Z}}_N^*)\times[(N^{-1}\breve{\mathbf{Z}}_N^{*\prime}\breve{\mathbf{H}}_N^*)(N^{-1}\breve{\mathbf{H}}_N^{*\prime}\breve{\mathbf{H}}_N^*)^{-1}(N^{-1}\breve{\mathbf{H}}_N^{*\prime}\breve{\mathbf{Z}}_N^*)]^{-1}$.

In light of Lemmata 1 and 2 the joint limiting distribution of the (feasible) generalized spatial TSLS estimator $\breve{\tilde{\boldsymbol{\delta}}}_N$ and the GM estimator $\widetilde{\boldsymbol{\rho}}_N$ follows from Theorem 4 and the discussion thereafter, with $\boldsymbol{\Delta}_N^* = \breve{\tilde{\boldsymbol{\delta}}}_N - \boldsymbol{\delta}_N$. The asymptotic variance-covariance matrix and its corresponding estimator are given by (27) and (28) with the modifications as described in Remark 2 after Theorem 4.

Note that in light of Lemma 2 the residuals $\hat{\mathbf{u}}_N^* = \mathbf{y}_N - \mathbf{Z}_N\breve{\tilde{\boldsymbol{\delta}}}_N = \mathbf{u}_N + \mathbf{D}_N\boldsymbol{\Delta}_N^*$ can be used to estimate $\boldsymbol{\rho}_N$ by the GM estimator defined by (10), where the discussion surrounding Lemma 2 would also apply here. Taking this argument one step further, $\boldsymbol{\rho}_N$ and $\boldsymbol{\delta}_N$ can also be estimated by an iterative procedure.

---

[7] The above Lemma corresponds to Lemma 4 in Kelejian and Prucha (2008) and is adapted here to apply to the higher-order case.

[8] In light of the properties maintained with respect to the matrix $(\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N}\mathbf{M}_{m,N})\mathbf{u}_N$, this implies that Assumptions 9 and 10 will be satisfied for the transformed instruments $\mathbf{H}_N^*$.

As a final point, note that the above theory carries over to cases where the regressor matrix $\mathbf{X}_N$ includes endogenous variables, provided that suitable instruments are available. To be more specific, let $\mathbf{X}_N = (\overline{\mathbf{X}}_N, \mathbf{E}_N)$ and $\mathbf{D}_N = -\mathbf{Z}_N = -(\overline{\mathbf{X}}_N, \mathbf{E}_N, \mathbf{Y}_N)$, where $\overline{\mathbf{X}}_N$ satisfies Assumptions 8-10 with $\mathbf{X}_N$ replaced by $\overline{\mathbf{X}}_N$ (including in the formulation of the instruments), and where $\mathbf{E}_N$ is a matrix of endogenous variables. Then, given the fourth moments of $\mathbf{D}_N$ are uniformly bounded, and Assumption 6 holds, parts (a), (b), and (c) of Lemma 1 and 2 still hold, but with

$$\hat{\mathbf{Z}}_N = (\overline{\mathbf{X}}_N, \mathbf{P}_{\mathbf{H}_N}\mathbf{E}_N, \mathbf{P}_{\mathbf{H}_N}\overline{\mathbf{Y}}_N) \text{ and} \tag{35a}$$

$$\hat{\mathbf{Z}}^*_N = \mathbf{P}_{\mathbf{H}^*_N}\mathbf{Z}^*_N . \tag{35b}$$

V. Monte Carlo Evidence

In this section, we consider a Monte Carlos experiment for a SARAR(3,3) specification and restricted versions thereof. We assume that $\mathbf{W}_N = \mathbf{M}_N$ and that the matrix $\mathbf{X}$ includes two explanatory variables. Hence we have[9]

$$\mathbf{y} = \mathbf{x}_1\beta_1 + \mathbf{x}_2\beta_2 + \sum_{r=1}^{3}\lambda_r\mathbf{W}_r\mathbf{y} + \mathbf{u} , \tag{36a}$$

$$\mathbf{u} = \sum_{m=1}^{3}\rho_m\mathbf{W}_m\mathbf{u} + \boldsymbol{\varepsilon} . \tag{36b}$$

We consider three sample sizes: $N = 100$, $N = 250$, and $N = 500$. The explanatory variables $\mathbf{x}_1$ and $\mathbf{x}_2$ are generated as random draws from a standard normal distribution, scaled with a factor of five, and treated as fixed in repeated samples. Their parameters β_1 and β_2 are assumed to be unity in all Monte Carlo experiments considered.

For our basic setup of the weights matrix, we follow Kelejian and Prucha (1999) and use a binary 'up to 9 ahead and up to 9 behind' contiguity specification. This means that the elements of the time-invariant, raw weights matrix $\mathbf{W}^0$ are defined such that the *i*-th cross-section element is related to the 9 elements after it and the 9 elements before it.

The unnormalized $N \times N$ matrix $\mathbf{W}^0$ is then split up into three $N \times N$ matrices $\mathbf{W}^0_1, \mathbf{W}^0_2$, and $\mathbf{W}^0_3$, where $\mathbf{W}^0_1 + \mathbf{W}^0_2 + \mathbf{W}^0_3 = \mathbf{W}^0$. The matrices $\mathbf{W}^0_1$, $\mathbf{W}^0_2$, and $\mathbf{W}^0_3$ are specified such that they contain the elements of $\mathbf{W}^0$ for a different band of neighbours each. Otherwise, they have zero elements. We choose a design, where $\mathbf{W}^0_1$ corresponds to an 'up to 3 ahead and up

[9] For simplicity of notation, the subscript N is suppressed in the following.

to 3 behind' specification, $\mathbf{W}_2^0$ corresponds to a '4 to 6 ahead and 4 to 6 behind' specification, and $\mathbf{W}_3^0$ corresponds to a '7 to 9 ahead and 7 to 9 behind' specification. $\mathbf{W}_1^0$, $\mathbf{W}_2^0$, and $\mathbf{W}_3^0$ have typical elements $w_{1,ij}^0$, $w_{2,ij}^0$, and $w_{3,ij}^0$, respectively, where subscripts i and j indicate that the corresponding element captures the possible contiguity of unit i with j. $w_{1,ij}^0$, $w_{2,ij}^0$, and $w_{3,ij}^0$ are either unity or zero. By design, at most one of the three elements, $w_{1,ij}^0$, $w_{2,ij}^0$, or $w_{3,ij}^0$, can be unity. The final weights matrices $\mathbf{W}_1$, $\mathbf{W}_2$, and $\mathbf{W}_3$ are obtained by separately row-normalizing $\mathbf{W}_1^0$, $\mathbf{W}_2^0$, and $\mathbf{W}_3^0$, that is, by dividing their typical elements $w_{1,ij}^0$, $w_{2,ij}^0$, and $w_{3,ij}^0$ through the corresponding row sum, respectively.

With three row-normalized matrices $\mathbf{W}_1$, $\mathbf{W}_2$, and $\mathbf{W}_3$, the parameter space for $\boldsymbol{\lambda}$ and $\boldsymbol{\rho}$ must satisify $0 \le |\lambda_1| + |\lambda_2| + |\lambda_3| < 1$ and $0 \le |\rho_1| + |\rho_2| + |\rho_3| < 1$. We consider 12 parameter constellations, assuming that the spatial regressive parameters are non-increasing in the order of neighbourhood, i.e., we always have $\lambda_1 \ge \lambda_2 \ge \lambda_3$ and $\rho_1 \ge \rho_2 \ge \rho_3$. In parameter constellations (1a) through (2c), we assume that the spatial dependence in the endogenous variable **y** is at least as strong as that in the disturbances **u**, without loss of generality. We consider cases, where we have nonzero spatial dependence in both **y** and **u** (parameter constellations (1a) through (1c)), as well as ones where spatial dependence shows up exclusively in **y** (parameter constellations (2a) through (2c)) or exclusively in **u** (parameter constellations (3a) through (3c)). This setting should be informative about the performance of the GM estimator in discriminating between alternative specifications of the spatial dependence. Parameter constellations (3a) through (3c) and (4) consider cases where the spatial dependence in the disturbances **u** is stronger than that in **y**. Parameter constellation (5a) considers zero dependence parameters for all spatial lags in **y** and **u**. Finally, constellation (5b) assumes homogeneous but nonzero spatial dependence parameters for spatial lags in **y** and **u**.

Regarding the choice of instruments, we include linearly independent terms of up to the second order in equation (30b). In particular, the matrix of untransformed instruments **H** contains 18 columns and is given by

$$\mathbf{H} = (\mathbf{X}, \mathbf{W}_1\mathbf{X}, \mathbf{W}_2\mathbf{X}, \mathbf{W}_3\mathbf{X}, \mathbf{W}_1^2\mathbf{X}, \mathbf{W}_2^2\mathbf{X}, \mathbf{W}_3^2\mathbf{X}, \mathbf{W}_{12}\mathbf{X}, \mathbf{W}_{23}\mathbf{X}), \tag{37}$$

where $\mathbf{W}_{ij} = \mathbf{W}_i\mathbf{W}_j$.

The innovations $\boldsymbol{\varepsilon}$ are assumed to be heteroskedastic and generated as follows. Let ζ_i denote a draw from a standard normal distribution. The pattern of heteroskedasticity is drawn from a uniform distribution with support $[0.2, 1.8]$. Then, the innovations are generated as $\varepsilon_i = \sigma_{\varepsilon,i}\zeta_i$, where $\sqrt{0.2} \le \sigma_{\varepsilon,i} \le \sqrt{1.8}$.

Table 1. Parameter Constellations in Monte Carlo Experiments

Parameter constellation	λ_1	λ_2	λ_3	ρ_1	ρ_2	ρ_3
(1a)	0.5	0.3	0.1	0.4	0.2	0.2
(1b)	0.5	0.3	0	0.4	0.2	0
(1c)	0.5	0	0	0.4	0	0
(2a)	0.5	0.3	0.1	0	0	0
(2b)	0.5	0.3	0	0	0	0
(2c)	0.5	0	0	0	0	0
(3a)	0	0	0	0.4	0.2	0.2
(3b)	0	0	0	0.4	0.2	0
(3c)	0	0	0	0.4	0	0
(4)	0.2	0.1	0.1	0.5	0.3	0.1
(5a)	0	0	0	0	0	0
(5b)	0.2	0.2	0.2	0.2	0.2	0.2

Note: $\beta_1 = \beta_2 = 1$ under all parameter constellations.

For each Monte Carlo experiment, we consider 2000 draws. To ensure comparability, the same draws of ζ_i and $\sigma_{\varepsilon,i}$ are used for each of the 12 parameter constellations. Results for the estimates of $\rho_{1,N}, \rho_{2,N},$ and $\rho_{3,N}$ are obtained by the GM estimator defined in equation (10), using the optimal weighting matrix as given in equation (21).[10] The estimates reported for the regression parameters are FGTSLS estimates, based on the transformed model as given by equation (34) using the transformed set of instruments given in (37). For sample sizes of $N = 100$ and $N = 250$ we also report the results from regressions using the untransformed set of instruments $\mathbf{H}$ instead of $\mathbf{H}^*$ as in Kelejian and Prucha (2008) and Arriaz, Drukker, Kelejian, and Prucha (2007).

We calculate the average bias and root mean squared error for each parameter constellation. Moreover, based on the estimated approximate joint distribution of the vector of the spatial autoregressive and regression parameters, i.e., $\tilde{\mathbf{q}} = (\tilde{\lambda}_1, \tilde{\lambda}_2, \tilde{\lambda}_3, \tilde{\beta}_1, \tilde{\beta}_2, \tilde{\rho}_1, \tilde{\rho}_2, \tilde{\rho}_3)'$, we report rejection probabilities for Wald tests about a set of hypotheses of interest, using a nominal significance level of 5 percent.

[10] We use the identity matrix in an initial step to obtain consistent initial estimates of $\rho_{1,N}, \rho_{2,N},$ and $\rho_{3,N}$ and $\boldsymbol{\varepsilon}_N$, which are required to calculate the optimal weighting matrix $\tilde{\boldsymbol{\Psi}}_N$.

i) For all parameter constellations, we test the hypothesis that each single coefficient is equal to the true parameter value (this corresponds to a t-test as in Kelejian and Prucha, 2008). Hence, the corresponding rejection rates reflect the size of the test.
ii) For parameter constellation (1c) we report the test of the SARAR(3,3) against the SARAR(1,1) model, using $H_0^{\lambda,\rho,*}: \lambda_2 = \lambda_3 = \rho_2 = \rho_3 = 0$
iii) For parameter constellations (2a) and (2b), where spatial dependence occurs only in **y**, we test the joint hypothesis $H_0^{\rho}: \rho_1 = \rho_2 = \rho_3 = 0$. Similarly, for parameter constellations (3a) and (3b), where spatial dependence occurs only in **u**, we test the joint hypothesis $H_0^{\lambda}: \lambda_1 = \lambda_2 = \lambda_3 = 0$. These tests should be informative about the performance of the proposed estimator in discriminating between alternative spatial dependence in the endogenous variable versus the disturbances.
iv) Finally, the joint hypothesis $H_0^{\lambda,\rho}: \lambda_1 = \lambda_2 = \lambda_3 = \rho_1 = \rho_2 = \rho_3 = 0$ is also reported for the non-spatial model under parameter constellation (5a).

Using Theorem 4, the approximation of the small sample distribution of $\tilde{\mathbf{q}}$ is given by
$\tilde{\mathbf{q}} \sim N(\mathbf{q}, \mathbf{Q})$, where
$\mathbf{q} = (\lambda_1, \lambda_2, \lambda_3, \beta_1, \beta_2, \rho_1, \rho_2, \rho_3)'$ and
$\mathbf{Q} = Var(\tilde{\mathbf{q}})$,
which can be estimated using $\tilde{\mathbf{Q}} = N^{-1}\tilde{\mathbf{\Omega}}_o$.

Tests referring to a single parameter are carried out using a standard t-test:
E.g., $H_0^{\rho}: \tilde{\rho}_1 = \rho_1$ is tested using $t = \frac{\tilde{\rho}_1 - \rho_1}{\tilde{\sigma}_{\tilde{\rho}_1}}$, where $\sigma_{\tilde{\rho}_1}$ is the corresponding main diagonal element of $\tilde{\mathbf{Q}}$.

Tests regarding joint hypotheses are carried out using Wald tests. Generally, we have (e.g., Greene, 2003, pp. 95, 487):

$H_0: \mathbf{Rq} - \mathbf{t} = \mathbf{0}$ against $H_1: \mathbf{Rq} - \mathbf{t} \neq \mathbf{0}$.

Define the discrepancy vector: $\mathbf{m} = \mathbf{R}\,\tilde{\mathbf{q}} - \mathbf{t}$. The null hypothesis can the be tested using

$$\mathbf{m}'(\mathbf{R}\,\tilde{\mathbf{Q}}\,\mathbf{R}')^{-1}\mathbf{m} \sim \chi^2_G,$$

where G is the number of restrictions (the number of rows of $\mathbf{R}$).

In the present context, we have

for $H_0^{\lambda,\rho,*}: \lambda_2 = \lambda_3 = \rho_2 = \rho_3 = 0$, $\mathbf{R} = \begin{bmatrix} 0 & \mathbf{I}_2 & 0 & \mathbf{0} \\ 0 & \mathbf{0} & 0 & \mathbf{I}_2 \end{bmatrix}$ and $G = 4$;

for $H_0^{\rho}: \rho_1 = \rho_2 = \rho_3 = 0$, $\mathbf{R} = (\mathbf{A}, \mathbf{I}_3)$ where $\mathbf{A}$ is a 3×5 matrix of zeros and $G = 3$;

for $H_0^{\lambda}: \lambda_1 = \lambda_2 = \lambda_3 = 0$, $\mathbf{R} = (\mathbf{I}_3, \mathbf{A})$, where $\mathbf{A}$ is a 3×5 matrix of zeros and $G = 3$;

for $H_0^{\lambda,\rho}: \lambda_1 = \lambda_2 = \lambda_3 = \rho_1 = \rho_2 = \rho_3 = 0$, $\mathbf{R} = \begin{bmatrix} \mathbf{I}_3 & 0 & 0 & \mathbf{0} \\ \mathbf{0} & 0 & 0 & \mathbf{I}_3 \end{bmatrix}$ and $G = 6$;

Table 2 to 4 report the results of the Monte Carlo analysis for the three sample sizes considered, using the matrix $\mathbf{H}^*$.

In terms of bias and RMSE, the estimator performs well, even in the small sample with $N = 100$. On average over all parameter constellation the bias and RMSE amount to 0.0052 and 0.0426 for the estimates of $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_3)'$ and to 0.0314 and 0.2017 for the estimates of $\boldsymbol{\rho} = (\rho_1, ..., \rho_3)'$. Regarding the size of the tests, the performance of the GM estimates of $\boldsymbol{\rho}$ the disturbances process is quite well. Even for the small sample of $N = 100$, the size of the tests is not too far away from the nominal size. This holds true for the size of the rejections rates of the tests involving only 1 parameter (average: 0.0794) and to a smaller extent also for tests of the joint tests involving $\boldsymbol{\rho}$ only (average size: 0.0703). The performance of the FGTSLS estimates of $\boldsymbol{\lambda}$ is worse, with an average size of 0.0947 for the single tests and an average size of 0.1516 for the joint tests (involving $\boldsymbol{\lambda}$ only).[11]

However, performance improves quickly with growing sample size. For $N = 250$, the average bias and RMSE of the estimates of $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_3)'$ shrink to 0.0016 and 0.0252, those of $\boldsymbol{\rho} = (\rho_1, ..., \rho_3)'$ shrink to 0.0153 and 0.1034. Also, the size of the tests improves and approaches the nominal size of 5 percent. Regarding the GM estimates of $\boldsymbol{\rho}$, the average size of the tests involving only one parameter amounts to 0.0604, that for the joint tests involving $\boldsymbol{\rho}$ only to 0.0542. For the FGTSLS estimates of $\boldsymbol{\lambda}$, the average size is 0.0677 for the single tests and 0.0909 for the joint tests (involving $\boldsymbol{\lambda}$ only).[12]

[11] Results when using the untransformed instrument matrix $\mathbf{H}$ instead of $\mathbf{H}^*$ are as follows: bias of $\boldsymbol{\rho}$: -0.0308; RMSE of $\boldsymbol{\rho}$: 0.2112; size of $\boldsymbol{\rho}$ for single tests: 0.0764; size of $\boldsymbol{\rho}$ for joint tests: 0.1042; bias of $\boldsymbol{\lambda}$: 0.0054, RMSE of $\boldsymbol{\lambda}$: 0.0416; size of single tests for $\boldsymbol{\lambda}$: 0.1021; size of joint tests for $\boldsymbol{\lambda}$: 0.1738. See Table A.1 in Appendix E for details.

[12] Results when using the untransformed instrument matrix $\mathbf{H}$ instead of $\mathbf{H}^*$ are as follows: bias of $\boldsymbol{\rho}$: 0.0148; RMSE of $\boldsymbol{\rho}$: 0.1037; size of $\boldsymbol{\rho}$ for single tests: 0.0524; size of $\boldsymbol{\rho}$ for joint tests: 0.0577; bias of $\boldsymbol{\lambda}$: 0.0026, RMSE of $\boldsymbol{\lambda}$: 0.0250; size of single tests for $\boldsymbol{\lambda}$: 0.0720; size of joint tests for $\boldsymbol{\lambda}$: 0.0946. See Table A.2 in Appendix E for details.

Table 2. Monte Carlo Results, $N = 100$, 2000 draws, instrument matrix $\mathbf{H}^*$

Constellation[1]	(1a)	(1b)	(1c)	(2a)	(2b)	(2c)	(3a)	(3b)	(3c)	(4)	(5a)	(5b)	average [2]
λ_1	0.5	0.5	0.5	0.5	0.5	0.5	0	0	0	0.2	0	0.2	0.2833
Bias	0.0044	0.0050	0.0059	0.0030	0.0028	0.0033	0.0165	0.0096	0.0064	0.0367	0.0034	0.0046	0.0085
RMSE	0.0349	0.0354	0.0388	0.0275	0.0270	0.0289	0.0794	0.0694	0.0570	0.0913	0.0407	0.0421	0.0477
Rej. Rate	0.0800	0.0825	0.0830	0.0785	0.0740	0.0730	0.1260	0.1075	0.0900	0.2045	0.0715	0.0750	0.0955
λ_2	0.3	0.3	0	0.3	0.3	0	0	0	0	0.1	0	0.2	0.1250
Bias	0.0001	0.0001	-0.0002	-0.0011	-0.0008	-0.0005	0.0112	0.0045	0.0010	0.0269	0.0001	0.0018	0.0040
RMSE	0.0292	0.0306	0.0333	0.0269	0.0278	0.0276	0.0644	0.0550	0.0481	0.0715	0.0362	0.0366	0.0406
Rej. Rate	0.0700	0.0745	0.0845	0.0675	0.0720	0.0785	0.1345	0.1085	0.0965	0.2005	0.0865	0.0905	0.0970
λ_3	0.1	0	0	0.1	0	0	0	0	0	0.1	0	0.2	0.0417
Bias	-0.0021	-0.0027	-0.0027	-0.0024	-0.0022	-0.0017	0.0056	0.0005	-0.0019	0.0155	-0.0016	0.0001	0.0032
RMSE	0.0358	0.0372	0.0377	0.0284	0.0306	0.0319	0.0495	0.0457	0.0457	0.0533	0.0397	0.0393	0.0396
Rej. Rate	0.0880	0.0915	0.0890	0.0900	0.0925	0.0875	0.0935	0.0835	0.0855	0.1255	0.0875	0.0840	0.0915
β_1	1	1	1	1	1	1	1	1	1	1	1	1	1.0000
Bias	0.0000	0.0001	0.0003	-0.0003	-0.0004	-0.0004	0.0021	0.0011	0.0007	0.0051	-0.0001	0.0000	0.0009
RMSE	0.0193	0.0194	0.0193	0.0199	0.0199	0.0197	0.0224	0.0216	0.0208	0.0233	0.0201	0.0199	0.0205
Rej. Rate	0.0680	0.0670	0.0685	0.0735	0.0755	0.0715	0.0760	0.0740	0.0695	0.0850	0.0685	0.0665	0.0720
β_2	1	1	1	1	1	1	1	1	1	1	1	1	1.0000
Bias	-0.0007	-0.0006	-0.0004	-0.0008	-0.0009	-0.0010	0.0014	0.0003	-0.0001	0.0040	-0.0007	-0.0006	0.0010
RMSE	0.0216	0.0215	0.0218	0.0223	0.0223	0.0225	0.0240	0.0232	0.0227	0.0244	0.0224	0.0222	0.0226
Rej. Rate	0.0665	0.0685	0.0680	0.0685	0.0695	0.0715	0.0660	0.0660	0.0675	0.0680	0.0715	0.0690	0.0684
ρ_1	0.4	0.4	0.4	0	0	0	0.4	0.4	0.4	0.5	0	0.2	0.2583
Bias	-0.0378	-0.0446	-0.0572	-0.0583	-0.0614	-0.0583	-0.0277	-0.0397	-0.0510	-0.0125	-0.0539	-0.0364	0.0449
RMSE	0.1805	0.1855	0.2142	0.2771	0.2742	0.2647	0.1645	0.1746	0.2016	0.1493	0.2410	0.1881	0.2096
Rej. Rate	0.0795	0.0835	0.0930	0.0800	0.0810	0.0815	0.0650	0.0800	0.0865	0.0595	0.0690	0.0700	0.0774
ρ_2	0.2	0.2	0	0	0	0	0.2	0.2	0	0.3	0	0.2	0.1083
Bias	-0.0303	-0.0312	-0.0076	-0.0443	-0.0446	-0.0307	-0.0156	-0.0194	-0.0063	-0.0355	-0.0348	-0.0334	0.0278
RMSE	0.1810	0.1855	0.1927	0.2545	0.2493	0.2410	0.1682	0.1813	0.1914	0.1686	0.2247	0.1890	0.2023
Rej. Rate	0.0755	0.0765	0.0765	0.0735	0.0710	0.0710	0.0650	0.0720	0.0795	0.0620	0.0635	0.0690	0.0713
ρ_3	0.1	0	0	0	0	0	0.2	0	0	0.1	0	0.2	0.0500
Bias	-0.0248	-0.0090	-0.0105	-0.0252	-0.0216	-0.0177	-0.0306	-0.0088	-0.0110	-0.0374	-0.0230	-0.0382	0.0215
RMSE	0.1783	0.1810	0.1912	0.2224	0.2179	0.2138	0.1737	0.1823	0.1935	0.1659	0.2059	0.1915	0.1931
Rej. Rate	0.0650	0.0660	0.0620	0.0575	0.0590	0.0570	0.0625	0.0650	0.0670	0.0695	0.0535	0.0615	0.0621
Joint Tests [3]													
Rej. Rate			0.1230	0.0970	0.0970	0.0985	0.1645	0.1510	0.1510		0.1400		0.1278

Note: [1] Each column corresponds to one parameter constellation (see Table 1). [2] Average of absolute row values. [3] Rejections rates for the following hypotheses. (1c): $H_0^{\lambda,\rho,*}: \lambda_2 = \lambda_3 = \rho_2 = \rho_3 = 0$, (2a), (2b), (2c): $H_0^{\rho}: \rho_1 = \rho_2 = \rho_3 = 0$, (3a), (3b), (3c): $H_0^{\lambda}: \lambda_1 = \lambda_2 = \lambda_3 = 0$, (5a): $H_0^{\lambda,\rho}: \lambda_1 = \lambda_2 = \lambda_3 = \rho_1 = \rho_2 = \rho_3 = 0$.

Table 3. Monte Carlo Results, $N = 250$, 2000 draws, instrument matrix $\mathbf{H}^*$

Constellation[1]	(1a)	(1b)	(1c)	(2a)	(2b)	(2c)	(3a)	(3b)	(3c)	(4)	(5a)	(5b)	average [2]
λ_1	0.5	0.5	0.5	0.5	0.5	0.5	0	0	0	0.2	0	0.2	0.2833
Bias	0.0010	0.0012	0.0011	0.0007	0.0006	0.0007	0.0049	0.0029	0.0013	0.0123	0.0003	0.0012	0.0023
RMSE	0.0183	0.0191	0.0215	0.0156	0.0153	0.0164	0.0428	0.0383	0.0316	0.0487	0.0232	0.0232	0.0262
Rej. Rate	0.0565	0.0600	0.0675	0.0560	0.0585	0.0640	0.0805	0.0765	0.0690	0.1140	0.0655	0.0585	0.0689
λ_2	0.3	0.3	0	0.3	0.3	0	0	0	0	0.1	0	0.2	0.1250
Bias	0.0001	0.0003	-0.0002	-0.0005	-0.0004	-0.0004	0.0039	0.0017	0.0003	0.0094	-0.0002	0.0003	0.0015
RMSE	0.0201	0.0208	0.0220	0.0203	0.0205	0.0196	0.0380	0.0331	0.0292	0.0419	0.0233	0.0234	0.0260
Rej. Rate	0.0560	0.0570	0.0585	0.0540	0.0565	0.0555	0.0855	0.0745	0.0690	0.1165	0.0590	0.0585	0.0667
λ_3	0.1	0	0	0.1	0	0	0	0	0	0.1	0	0.2	0.0417
Bias	-0.0001	-0.0002	0.0002	-0.0002	-0.0001	0.0004	0.0030	0.0010	-0.0001	0.0065	0.0003	0.0009	0.0011
RMSE	0.0197	0.0204	0.0217	0.0163	0.0168	0.0183	0.0311	0.0287	0.0270	0.0333	0.0230	0.0234	0.0233
Rej. Rate	0.0560	0.0565	0.0600	0.0655	0.0655	0.0600	0.0790	0.0760	0.0700	0.0925	0.0605	0.0685	0.0675
β_1	1	1	1	1	1	1	1	1	1	1	1	1	1.0000
Bias	-0.0001	-0.0002	-0.0003	-0.0001	-0.0001	-0.0003	0.0006	0.0001	-0.0001	0.0017	-0.0002	0.0000	0.0003
RMSE	0.0125	0.0127	0.0128	0.0129	0.0131	0.0129	0.0140	0.0135	0.0131	0.0142	0.0130	0.0128	0.0131
Rej. Rate	0.0590	0.0625	0.0625	0.0620	0.0610	0.0590	0.0640	0.0600	0.0575	0.0650	0.0630	0.0620	0.0615
β_2	1	1	1	1	1	1	1	1	1	1	1	1	1.0000
Bias	0.0001	0.0002	0.0001	0.0001	0.0000	0.0001	0.0008	0.0006	0.0003	0.0019	0.0001	0.0002	0.0004
RMSE	0.0123	0.0124	0.0125	0.0128	0.0129	0.0127	0.0140	0.0135	0.0130	0.0141	0.0127	0.0126	0.0130
Rej. Rate	0.0545	0.0525	0.0565	0.0545	0.0610	0.0550	0.0595	0.0550	0.0515	0.0585	0.0570	0.0540	0.0558
ρ_1	0.4	0.4	0.4	0	0	0	0.4	0.4	0.4	0.5	0	0.2	0.2583
Bias	-0.0186	-0.0215	-0.0293	-0.0308	-0.0315	-0.0307	-0.0130	-0.0200	-0.0264	-0.0111	-0.0267	-0.0172	0.0231
RMSE	0.0884	0.0879	0.0957	0.1198	0.1206	0.1196	0.0899	0.0878	0.0946	0.0826	0.1184	0.1000	0.1004
Rej. Rate	0.0575	0.0590	0.0625	0.0505	0.0565	0.0590	0.0600	0.0585	0.0575	0.0570	0.0570	0.0555	0.0575
ρ_2	0.2	0.2	0	0	0	0	0.2	0.2	0	0.3	0	0.2	0.1083
Bias	-0.0123	-0.0143	-0.0057	-0.0215	-0.0215	-0.0162	-0.0082	-0.0110	-0.0061	-0.0137	-0.0161	-0.0140	0.0134
RMSE	0.1027	0.1036	0.1023	0.1189	0.1190	0.1180	0.1015	0.1036	0.1026	0.1044	0.1173	0.1062	0.1083
Rej. Rate	0.0600	0.0590	0.0585	0.0585	0.0570	0.0560	0.0590	0.0585	0.0625	0.0575	0.0555	0.0630	0.0587
ρ_3	0.1	0	0	0	0	0	0.2	0	0	0.1	0	0.2	0.0500
Bias	-0.0101	-0.0048	-0.0059	-0.0114	-0.0097	-0.0084	-0.0140	-0.0045	-0.0052	-0.0144	-0.0078	-0.0157	0.0093
RMSE	0.0963	0.0954	0.0956	0.1126	0.1123	0.1121	0.0952	0.0945	0.0954	0.0920	0.1119	0.1041	0.1014
Rej. Rate	0.0510	0.0445	0.0445	0.0415	0.0410	0.0415	0.0535	0.0460	0.0405	0.0660	0.0410	0.0455	0.0464
Joint Tests [3]													
Rej. Rate			0.0680	0.0490	0.0515	0.0525	0.0990	0.0935	0.0915		0.0795		0.0731

Note: [1] Each column corresponds to one parameter constellation (see Table 1). [2] Average of absolute row values. [3] Rejections rates for the following hypotheses. (1c): $H_0^{\lambda,\rho,*}: \lambda_2 = \lambda_3 = \rho_2 = \rho_3 = 0$, (2a), (2b), (2c): $H_0^{\rho}: \rho_1 = \rho_2 = \rho_3 = 0$, (3a), (3b), (3c): $H_0^{\lambda}: \lambda_1 = \lambda_2 = \lambda_3 = 0$, (5a): $H_0^{\lambda,\rho}: \lambda_1 = \lambda_2 = \lambda_3 = \rho_1 = \rho_2 = \rho_3 = 0$.

Table 4. Monte Carlo Results, $N = 500$, 2000 draws, instrument matrix $\mathbf{H}^*$

Constellation[1]	(1a)	(1b)	(1c)	(2a)	(2b)	(2c)	(3a)	(3b)	(3c)	(4)	(5a)	(5b)	average [2]
λ_1	0.5	0.5	0.5	0.5	0.5	0.5	0	0	0	0.2	0	0.2	0.2833
Bias	0.0007	0.0008	0.0008	0.0007	0.0005	0.0006	0.0024	0.0013	0.0011	0.0062	0.0003	0.0007	0.0013
RMSE	0.0126	0.0130	0.0144	0.0109	0.0107	0.0113	0.0290	0.0259	0.0215	0.0332	0.0157	0.0158	0.0178
Rej. Rate	0.0515	0.0555	0.0540	0.0555	0.0605	0.0525	0.0685	0.0595	0.0595	0.0885	0.0505	0.0545	0.0592
λ_2	0.3	0.3	0	0.3	0.3	0	0	0	0	0.1	0	0.2	0.1250
Bias	-0.0004	-0.0003	-0.0005	-0.0007	-0.0005	-0.0006	0.0014	0.0001	-0.0006	0.0041	-0.0002	0.0000	0.0008
RMSE	0.0139	0.0143	0.0152	0.0141	0.0141	0.0134	0.0256	0.0223	0.0200	0.0274	0.0160	0.0162	0.0177
Rej. Rate	0.0505	0.0560	0.0610	0.0535	0.0530	0.0545	0.0675	0.0600	0.0595	0.0835	0.0555	0.0590	0.0595
λ_3	0.1	0	0	0.1	0	0	0	0	0	0.1	0	0.2	0.0417
Bias	0.0000	-0.0001	-0.0003	0.0000	-0.0001	0.0000	0.0010	0.0000	-0.0004	0.0030	-0.0004	0.0000	0.0005
RMSE	0.0140	0.0143	0.0152	0.0112	0.0114	0.0125	0.0215	0.0199	0.0193	0.0228	0.0162	0.0162	0.0162
Rej. Rate	0.0580	0.0525	0.0555	0.0540	0.0575	0.0575	0.0645	0.0665	0.0700	0.0775	0.0600	0.0550	0.0607
β_1	1	1	1	1	1	1	1	1	1	1	1	1	1.0000
Bias	-0.0001	0.0001	0.0001	0.0001	0.0001	0.0000	0.0004	0.0001	0.0002	0.0011	0.0001	0.0001	0.0002
RMSE	0.0090	0.0088	0.0089	0.0091	0.0091	0.0092	0.0098	0.0094	0.0092	0.0100	0.0090	0.0090	0.0092
Rej. Rate	0.0530	0.0475	0.0500	0.0515	0.0500	0.0540	0.0505	0.0450	0.0475	0.0515	0.0525	0.0560	0.0507
β_2	1	1	1	1	1	1	1	1	1	1	1	1	1.0000
Bias	0.0001	0.0002	0.0001	0.0000	0.0000	0.0000	0.0005	0.0002	0.0001	0.0010	0.0001	0.0000	0.0002
RMSE	0.0086	0.0087	0.0087	0.0089	0.0089	0.0089	0.0096	0.0093	0.0090	0.0098	0.0088	0.0089	0.0090
Rej. Rate	0.0480	0.0510	0.0505	0.0490	0.0475	0.0490	0.0530	0.0485	0.0440	0.0535	0.0480	0.0500	0.0493
ρ_1	0.4	0.4	0.4	0	0	0	0.4	0.4	0.4	0.5	0	0.2	0.2583
Bias	-0.0087	-0.0100	-0.0139	-0.0147	-0.0148	-0.0149	-0.0055	-0.0092	-0.0129	-0.0047	-0.0127	-0.0078	0.0108
RMSE	0.0640	0.0631	0.0676	0.0856	0.0853	0.0850	0.0650	0.0631	0.0673	0.0590	0.0847	0.0720	0.0718
Rej. Rate	0.0600	0.0660	0.0690	0.0615	0.0610	0.0580	0.0590	0.0630	0.0680	0.0590	0.0590	0.0570	0.0617
ρ_2	0.2	0.2	0	0	0	0	0.2	0.2	0	0.3	0	0.2	0.1083
Bias	-0.0061	-0.0072	-0.0034	-0.0106	-0.0106	-0.0083	-0.0048	-0.0058	-0.0034	-0.0081	-0.0086	-0.0072	0.0070
RMSE	0.0709	0.0718	0.0703	0.0808	0.0810	0.0804	0.0702	0.0714	0.0702	0.0713	0.0805	0.0726	0.0743
Rej. Rate	0.0535	0.0565	0.0565	0.0555	0.0565	0.0535	0.0535	0.0530	0.0580	0.0515	0.0535	0.0545	0.0547
ρ_3	0.1	0	0	0	0	0	0.2	0	0	0.1	0	0.2	0.0500
Bias	-0.0053	-0.0024	-0.0025	-0.0052	-0.0044	-0.0036	-0.0073	-0.0021	-0.0025	-0.0082	-0.0035	-0.0074	0.0045
RMSE	0.0671	0.0670	0.0675	0.0795	0.0793	0.0794	0.0649	0.0661	0.0676	0.0610	0.0797	0.0725	0.0710
Rej. Rate	0.0510	0.0530	0.0540	0.0465	0.0500	0.0490	0.0535	0.0545	0.0575	0.0615	0.0495	0.0530	0.0527
Joint Tests [3]													
Rej. Rate			0.0810	0.0725	0.0750	0.0710	0.0855	0.0890	0.0890		0.0785		0.0802

Note: [1] Each column corresponds to one parameter constellation (see Table 1). [2] Average of absolute row values. [3] Rejections rates for the following hypotheses. (1c): $H_0^{\lambda,\rho,*}: \lambda_2 = \lambda_3 = \rho_2 = \rho_3 = 0$, (2a), (2b), (2c): $H_0^{\rho}: \rho_1 = \rho_2 = \rho_3 = 0$, (3a), (3b), (3c): $H_0^{\lambda}: \lambda_1 = \lambda_2 = \lambda_3 = 0$, (5a): $H_0^{\lambda,\rho}: \lambda_1 = \lambda_2 = \lambda_3 = \rho_1 = \rho_2 = \rho_3 = 0$.

For $N = 500$ we have the following results. Across all parameter constellations, the average bias and RMSE amount to 0.0010 and 0.0172, respectively, for the estimates of $\boldsymbol{\lambda} = (\lambda_1,...,\lambda_3)'$ and to 0.0075 and 0.0724, respectively, for the estimates of $\boldsymbol{\rho}$. The significance levels of the tests are very similar between the $N = 250$ and the $N = 500$ experiments across all tests.

Overall, the Monte Carlo experiments illustrate that the proposed estimators work reasonably well, even in small samples. This is true for both the point estimates and the variance-covariance matrix of the parameter estimates. The single as well as joint tests are properly sized and may be recommended for specification tests about the lag- and error-structure and the order of spatial dependence in medium to large samples.

VI. Conclusions

This paper derives generalized moments (GM) and two-stages least squares (TSLS) estimators for spatial autoregressive models with spatial regressive disturbances and heteroskedastic innovations, allowing for an arbitrary (but finite) order of spatial dependence both in the dependent variable and the disturbances. We prove consistency of the proposed estimators and derive the (joint) asymptotic distribution of the GM estimates of the spatial autoregressive parameters of the disturbance process and the feasible (generalized) TSLS estimates of the regression parameters of the model. The variance-covariance matrix of all model parameters can be used to formulate joint tests about the form and order of spatial dependence, e.g., tests of the general SARAR(*R*,*S*) model against interesting alternatives such as SARAR(0,*S*), SARAR(*R*,0), and the non-spatial model. A comprehensive Monte Carlo simulation exercise suggests that the estimators perform reasonably well in terms of bias and root means squared errors, even in small samples with 100 observations. The rejection rates of the single and joint tests approach the nominal size as the number of observations grows larger and can be used for specification tests in medium to large samples in order to empirically determine the proper specification and order of spatial dependence.

The SARAR(*R*,*S*) framework developed in this paper allows the applied econometrician to study the strength of interdependence between cross-sectional units more flexibly than in existing SARAR(1,1) models. For instance, with the suggested model one may allow first, second, and higher orders of bands of neighbours to exert a different impact on each other, allowing a better approximation of the (possibly nonlinear, discontinuous) functional form of the decay of spatial interdependence.

Moreover, one may allow for several alternative channels or concepts of interdependence in space, which may be non-geographical, and the SARAR(*R*,*S*) framework can be readily adapted to non-geographical models by replacing the notion of geographical distance with, e.g., economic, socio-economic, cultural, or political distance.

References

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

Arraiz, I., Drukker, D.M., Kelejian, H., and Prucha, I. (2007). A spatial Cliff-Ord-type model with heteroskedastic innovations: Small and large sample results. Unpublished manuscript.

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

Badinger, H. and Egger, P. (2008). Intra- and inter-industry productivity spillovers in OECD manufacturing: A spatial econometric perspective. *CESIfo Working Paper*, No. 2181.

Baltagi, B.H. and Li, D. (2001). LM test for functional form and spatial error correlation. *International Regional Science Review*, 24, 194-225.

Baltagi, B.H., Egger, P., and Pfaffermayr, M. (2005). Estimating models of complex FDI: Are there third-country effects? *Journal of Econometrics*, 127(1), 260-281

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

Bell, K.P. and Bockstael, N.E. (2000). Applying the generalized-moments estimation apporoach to spatial problems involving microlevel data. *The Review of Economics and Statistics,* 82(1), 72–82.

Besley, T. and Case, A. (1995). Incumbent behavior: Vote-seeking, tax-setting, and yardstick competition. *American Economic Review*, 85, 25-45.

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

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

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

Cohen, J.P. and Morrison Paul, C.J. (2004). Public infrastructure investment, interstate spatial spillovers, and manufacturing costs. *The Review of Economics and Statistics,* 86(2), 551-560.

Cohen, J.P. and Morrison Paul, C. (2007). The impacts of transportation infrastructure on property values: A higher order spatial econometrics approach. *Journal of Regional Science*, 47(3), 457-478.

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

Egger, P. and Raff, H. (2008). Tax rate and tax base competition for foreign direct investment. Unpublished manuscript, Christian-Albrechts-University of Kiel.

Greene, W.H. (2003). *Econometric Analysis*, fifth edition. Pearson, Upper Saddle River, New Jersey.

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

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

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

Kelejian, H.H. and Prucha, I.R. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *Journal of Real Estate Finance and Economics*, 17, 99-121.

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

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

Kelejian, H.H. and Prucha, I.R. (2007). HAC estimation in a spatial framework. *Journal of Econometrics*, 140(1), 131-154.

Kelejian, H.H. and Prucha, I.R. (2008). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics*, forthcoming.

Kelejian, H.H., Prucha, I.R. and Yuzefovich, E. (2004). Instrumental variable estimation of a spatial autoregressive model with autoregressive disturbances: Large and small sample results. In: LeSage, J. and Pace, K. (eds.), *Advances in Econometrics: Spatial and Spatiotemporal Econometrics*. Elsevier, New York, 163-198.

Lee, L.F. (2004). Asymptotic distributions of maximum likelihood estimators for spatial autoregressive models. *Econometrica*, 72, 1899-1925.

Lee, L.F. (2007). GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics*, 137, 489-514.

Mittelhammer, R.C. (1996). *Mathematical Statistics for Economics and Business*. New York: Springer.

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

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

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

Rao, C.R. (1973). *Linear Statistical Inference and its Applications*, 2nd edition. New York: Wiley.

Resnik, S. (1999). *A Probability Path*. Boston: Birkhäuser.

Shroder, M. (1995). Games the States don't play: Welfare benefits and the theory of fiscal federalism. *Review of Economics and Statistics*, 77, 183-191.

Topa, G. (2001). Social interactions, local spillovers and unemployment. *Review of Economic Studies*, 68, 261-295.

APPENDIX (Not intended for publication in full length.)

In the following, we give a proof of Theorems 1-4 and state several results that are repeatedly used in this paper. The proofs as given here proceed closely along the lines of Kelejian and Prucha (2008) for the SARAR(1,1) model and are adapted to apply the more general case of an SARAR(*R*,*S*) model.

APPENDIX A

Notation

We adopt the standard convention to refer to matrices and vectors with acronyms in boldface. Let $\mathbf{A}_N$ denote some matrix. Its elements are referred to as $a_{ij,N}$; $\mathbf{a}_{i.,N}$ and $\mathbf{a}_{.i,N}$ denote the *i*-th row and the *i*-th column of $\mathbf{A}_N$ respectively. If $\mathbf{A}_N$ is a square matrix, $\mathbf{A}_N^{-1}$ denotes its inverse; if $\mathbf{A}_N$ is singular, $\mathbf{A}_N^{+}$ denotes its generalized inverse. If $\mathbf{A}_N$ is a square, symmetric and positive definite matrix, $\mathbf{A}_N^{1/2}$ denotes the unique positive definite square root of $\mathbf{A}_N$ and $\mathbf{A}_N^{-1/2}$ denotes $(\mathbf{A}_N^{-1})^{1/2}$. Finally, define the matrix norm $\|\mathbf{A}_N\| = [Tr(\mathbf{A}_N'\mathbf{A}_N)]^{1/2}$, where *Tr* is the trace operator. Note that the norm is submultiplicative, i.e., $\|\mathbf{A}_N\mathbf{B}_N\| \le \|\mathbf{A}_N\|\|\mathbf{B}_N\|$.

Remark A.1[13]

Let $\mathbf{A}_N$, $N \ge 1$, be some sequence of $N \times N$ matrices. We will then say that the row and column sums of the (sequence of) matrices $\mathbf{A}_N$ are bounded uniformly in absolute value if there exists a constant $c_A < \infty$, which does not depend on *N*, such that

$$\max_{1\le i\le N}\sum_{j=1}^{N}|a_{ij,N}| \le c_A, \text{ and } \max_{1\le j\le N}\sum_{i=1}^{N}|a_{ij,N}| \le c_A \text{ for all } N \ge 1.$$

The following results will be used repeatedly in the proofs:

- If $\mathbf{A}_N$ and $\mathbf{B}_N$ are (sequences of) $N \times N$ matrices, whose row and column sums are bounded uniformly in absolute value (say by c_A and c_B), then so are the row and column sums of $\mathbf{A}_N\mathbf{B}_N$ and $\mathbf{A}_N + \mathbf{B}_N$ by c_Ac_B and $c_A + c_B$, respectively (Kelejian and Prucha, 1999, p. 526).
- If $\mathbf{Z}_N$ is a (sequence of) $N \times P$ matrices whose elements are uniformly bounded in absolute value (say by c_Z), then so are the elements of $\mathbf{A}_N\mathbf{Z}_N$ (by c_Ac_Z) and $N^{-1}\mathbf{Z}_N'\mathbf{A}_N\mathbf{Z}_N$ (Kelejian and Prucha, 2004, Remark A.1). (This also covers the case $N^{-1}\mathbf{Z}_N'\mathbf{Z}_N$ for $\mathbf{A}_N = \mathbf{I}_N$.)

[13] Compare Kelejian and Prucha (2008, p. 5).

- Suppose that the row and columns sums of the $NP \times NP$ matrices $\mathbf{A}_N = (a_{ij,N})$ are uniformly bounded in absolute value by some finite constant c_A, then $\sum_{i=1}^{NP} \left| a_{ij,N} \right|^q \leq c_A^q$ for $q > 1$ (Kelejian and Prucha, 2008, Remark C.1).
- Let $\boldsymbol{\xi}_N$ and $\boldsymbol{\eta}_N$ be $N \times 1$ random vectors, where, for each *N*, the elements are independently distributed with zero mean and finite variances. Then the elements of $N^{-1/2}\mathbf{Z}_N'\boldsymbol{\xi}_N$ are $O_p(1)$ and $N^{-1}\boldsymbol{\xi}_N'\mathbf{A}_N\boldsymbol{\eta}_N$ is $O_p(1)$.[14]
- Let $\boldsymbol{\zeta}_N$ be an $N \times 1$ random vector, where, for each *N*, the elements are independently distributed with zero mean and finite fourth moments. Let $\boldsymbol{\pi}_N$ be some nonstochastic $N \times 1$ vector, whose elements are uniformly bounded in absolute value and let $\boldsymbol{\Pi}_N$ be an $N \times N$ nonstochastic matrix whose row and columns sums are uniformly bounded in absolute value. Define the column vector $\mathbf{d}_N = \boldsymbol{\pi}_N + \boldsymbol{\Pi}_N\boldsymbol{\zeta}_N$. It follows that the elements of $d_{ij,N}$ have finite fourth moments. (See Kelejian and Prucha, 2008, Lemma C.2 for the proof.)

Remark A.2 (Identifiable Uniqueness in Higher-Order Case)
Assumption 5 states that the smallest eigenvalue of $\boldsymbol{\Gamma}_N'\boldsymbol{\Gamma}_N$ is uniformly bounded away from zero. To show what this assumption requires in the higher-order case, write

$$\boldsymbol{\Gamma}_N = [\boldsymbol{\Gamma}_{1,N}' \quad . \quad \boldsymbol{\Gamma}_{S,N}']' . \tag{A.1}$$

Hence, for any $N \times 1$ vector $\mathbf{x} \neq \mathbf{0}$

$$\mathbf{x}'\boldsymbol{\Gamma}_N'\boldsymbol{\Gamma}_N\mathbf{x} = \mathbf{x}'(\boldsymbol{\Gamma}_{1,N}'\boldsymbol{\Gamma}_{1,N} + ... + \boldsymbol{\Gamma}_{S,N}'\boldsymbol{\Gamma}_{S,N})\mathbf{x} . \tag{A.2}$$

In light of Rao (1973, p.62),[15] we have

$$\mathbf{x}'\boldsymbol{\Gamma}_{1,N}'\boldsymbol{\Gamma}_{1,N}\mathbf{x} + ... + \mathbf{x}'\boldsymbol{\Gamma}_{S,N}'\boldsymbol{\Gamma}_{S,N}\mathbf{x} \geq \lambda_{\min}(\boldsymbol{\Gamma}_{1,N}'\boldsymbol{\Gamma}_{1,N})\mathbf{x}'\mathbf{x} + ... + \lambda_{\min}(\boldsymbol{\Gamma}_{S,N}'\boldsymbol{\Gamma}_{S,N})\mathbf{x}'\mathbf{x} . \tag{A.3}$$

This expression is strictly larger than zero if $\lambda_{\min}(\boldsymbol{\Gamma}_{s,N}'\boldsymbol{\Gamma}_{s,N}) \geq \lambda_* > 0$ for some *s*, $s = 1, \ldots, S$.

[14] Compare Kelejian and Prucha (2004), who consider homoskedastic random variables. It is readily observed from the proof that the result also holds under heteroskedasticity, as long as the variances of the elements of $\boldsymbol{\xi}_N$ and $\boldsymbol{\eta}_N$ are uniformly bounded.

[15] See Remark B.1 in Appendix B.

We then have

$$\mathbf{x}'\boldsymbol{\Gamma}'_N\boldsymbol{\Gamma}_N\mathbf{x} \geq \lambda_*\mathbf{x}'\mathbf{x} > 0\,. \tag{A.4}$$

Next, using Rao (1973, p. 62) again

$$\lambda_{\min}(\boldsymbol{\Gamma}'_N\boldsymbol{\Gamma}_N) = \inf_{\mathbf{x}} \frac{\mathbf{x}'\boldsymbol{\Gamma}'_N\boldsymbol{\Gamma}_N\mathbf{x}}{\mathbf{x}'\mathbf{x}} \geq \lambda_* > 0\,. \tag{A.5}$$

As can be seen from (A.5), Assumption 5 in the higher-order case requires that the assumption made by Kelejian and Prucha (2008) for the first-order case is fulfilled for at least one subset of moment conditions associated with one of the weights matrices. Note, however, that all weighting matrices enter the elements of each $\boldsymbol{\Gamma}_{s,N}$, $s = 1,...,S$. If two weights matrices are collinear, for example, none of the matrices $\boldsymbol{\Gamma}_{s,N}$ would have a smallest eigenvalue that is strictly positive and Assumption 5 would be hurt.

APPENDIX B.

For the convenience of the reader, Appendix B lists some Lemmata and Theorems as used in the subsequent proofs.

Remark B.1.

Let **A** be a symmetric $N \times N$ matrix. It holds that (see Rao, 1973, p. 62)

$$\lambda_{\min}(\mathbf{A}) = \min_{\substack{\mathbf{x}\in R^n \\ \mathbf{x}\neq 0}} \frac{\mathbf{x}'\mathbf{A}\mathbf{x}}{\mathbf{x}'\mathbf{x}} \text{ and } \lambda_{\max}(\mathbf{A}) = \max_{\substack{\mathbf{x}\in R^n \\ \mathbf{x}\neq 0}} \frac{\mathbf{x}'\mathbf{A}\mathbf{x}}{\mathbf{x}'\mathbf{x}}.$$

Let **A** and **B** be symmetric, positive semidefinite matrices of dimension $N \times N$. Then $\lambda_S(\mathbf{A})Tr(\mathbf{B}) \leq Tr(\mathbf{AB}) \leq \lambda_L(\mathbf{A})Tr(\mathbf{B})$, where λ_L and λ_S denote the largest and smallest eigenvalue of **A**, respectively (Mittelhammer, 1996, p. 254).

Lemma A.1 in Kelejian and Prucha (2008)

Let $\boldsymbol{\varepsilon}$ be a random $N \times 1$ vector with zero mean and positive definite variance-covariance matrix $\boldsymbol{\Sigma}$, let **A** and **B** be a symmetric, nonstochastic $N \times N$ matrices, and let **a** and **b** be real nonstochastic $N \times 1$ vectors. Consider the decomposition $\boldsymbol{\Sigma} = \mathbf{S}\mathbf{S}'$. Let $\mathbf{A}^* = (a_{ij,*}) = \mathbf{S}'\mathbf{A}\mathbf{S}$ and $\mathbf{B}^* = (b_{ij,*}) = \mathbf{S}'\mathbf{B}\mathbf{S}$, and let $\mathbf{a}^* = (a_{i,*}) = \mathbf{S}'\mathbf{a}$ and $\mathbf{b}^* = (b_{i,*}) = \mathbf{S}'\mathbf{b}$. Furthermore, let $\boldsymbol{\eta} = \mathbf{S}^{-1}\boldsymbol{\varepsilon}$. Then assuming that the elements of $\boldsymbol{\eta}$ are independently

distributed with zero mean, unit variances and finite third and fourth moments $E(\eta_i^3) = \mu_{\eta_i}^{(3)}$ and $E(\eta_i^4) = \mu_{\eta_i}^{(4)}$, we have

$E(\boldsymbol{\varepsilon}'\mathbf{A}\boldsymbol{\varepsilon} + \mathbf{a}'\boldsymbol{\varepsilon}) = Tr(\mathbf{A}^*) = Tr(\mathbf{A}\boldsymbol{\Sigma})$ and

$$Cov(\boldsymbol{\varepsilon}'\mathbf{A}\boldsymbol{\varepsilon} + \mathbf{a}'\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon}'\mathbf{B}\boldsymbol{\varepsilon} + \mathbf{b}'\boldsymbol{\varepsilon}) = 2Tr(\mathbf{A}\boldsymbol{\Sigma}\mathbf{B}\boldsymbol{\Sigma}) + \mathbf{a}'\boldsymbol{\Sigma}\mathbf{b} + \sum_{i=1}^{N} a_{ii,*}b_{ii,*}[\mu_{\eta_i}^{(4)} - 3] + \sum_{i=1}^{N}(a_{i,*}b_{ii,*} + a_{ii,*}b_{i,*})\mu_{\eta_i}^{(3)}.$$

Note that when $\mathbf{A}$ and $\mathbf{B}$ have zero main diagonal elements the last two terms of the expression for the covariance drop out.

Theorem A.1 in Kelejian and Prucha (2008)

Define the $M \times 1$ vector of quadratic forms $\mathbf{x}_N = (\boldsymbol{\varepsilon}_N'\mathbf{A}_{r,N}\boldsymbol{\varepsilon}_N + \mathbf{a}_{r,N}'\boldsymbol{\varepsilon}_N)$, where $\mathbf{A}_{r,N}$, $m = 1,...,M$, are real nonstochastic $N \times N$ matrices, $\mathbf{a}_{r,N}$, $r = 1,...,M$, are real nonstochastic $N \times 1$ vectors, and $\boldsymbol{\varepsilon}_N$ is a $N \times 1$ random vector. Suppose the following assumptions hold:

Assumption A.1

The real valued random variables of the array $\{\varepsilon_{i,N} : 1 \le i \le N, N \ge 1\}$ satisfy $E(\varepsilon_{i,N}) = 0$. Furthermore, for each $N \ge 1$, the random variables $\varepsilon_{1,N},...,\varepsilon_{N,N}$ are totally independent.

Assumption A.2

For $r = 1,...,M$

a) the elements of the array of real numbers $\{a_{ij,r,N} : 1 \le i, j \le N, N \ge 1\}$ satisfy $a_{ij,r,N} = a_{ji,r,N}$ and $\sup_{1 \le j \le N, N \ge 1} \sum_{i=1}^{N} |a_{ij,r,N}| < \infty$.

b) the elements of the array of real numbers $\{a_{i,r,N} : 1 \le i, N \ge 1\}$ satisfy $\sup_N N^{-1} \sum_{i=1}^{N} |a_{i,r,N}|^{2+\eta_1} < \infty$ for some $\eta_1 > 0$.

Assumption A.3

For $r = 1,...,M$ one of the following two conditions holds:

a) $\sup_{1 \le i \le N, N \ge 1} E|\varepsilon_{i,N}|^{2+\eta_2} < \infty$ for some $\eta_2 > 0$ and $a_{ii,r,N} = 0$.

b) $\sup_{1 \le i \le N, N \ge 1} E|\varepsilon_{i,N}|^{4+\eta_2} < \infty$ for some $\eta_2 > 0$ (but possibly $a_{ii,r,N} \ne 0$).

Denote the expectation of $\mathbf{x}_N$ as $\boldsymbol{\mu}_N = E(\mathbf{x}_N)$ and its variance-covariance matrix as $\Sigma_{\mathbf{x}_N} = E(\mathbf{x}_N\mathbf{x}_N')$, which can be derived using Lemma A.1 in Kelejian and Prucha (2008). It then follows under Assumptions A.1-A.3, and provided that $N^{-1}\lambda_{\min}(\boldsymbol{\Sigma}_{\mathbf{x}_N}) \ge c > 0$ holds, that

$\boldsymbol{\Sigma}_{\mathbf{x}_N}^{-1/2}(\mathbf{x}_n - \boldsymbol{\mu}_{\mathbf{x}_N}) \xrightarrow{d} (0, \mathbf{I}_M)$ as $N \to \infty$.

Lemma F1 in Pötscher and Prucha (1997)

Let $\mathbf{A}_N$ and $\mathbf{B}_N$ be real square random matrices. Let $\mathbf{B}_N$ be non-singular with probability approaching 1. Let $\mathbf{A}_N - \mathbf{B}_N \xrightarrow{p} \mathbf{0}$ as $N \to \infty$ and let the sequences $\mathbf{B}_N$ and $\mathbf{B}_N^+$ be bounded normwise in probability. Then the sequences $\mathbf{A}_N$ and $\mathbf{A}_N^+$ are bounded normwise in probability, $\mathbf{A}_N$ is non-singular with probability approaching 1, and $\mathbf{A}_N^+ - \mathbf{B}_N^+ \xrightarrow{p} \mathbf{0}$ as $N \to \infty$.

Corollary F4 in Pötscher and Prucha (1997)

Assume that $\boldsymbol{\eta}_N$ and $\boldsymbol{\zeta}_N$ are sequences of random vectors in $\mathbf{R}^p$ and $\mathbf{R}^q$ respectively, and let $\mathbf{A}_N$ be a sequence of bounded non-random $p \times q$ matrices. Suppose $\boldsymbol{\eta}_N = \mathbf{A}_N \boldsymbol{\zeta}_N + o_p(1)$ and that $\boldsymbol{\zeta}_N \xrightarrow{d} \boldsymbol{\zeta} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma}$ being positive definite. Define $\boldsymbol{\xi}_N = \mathbf{A}_N \boldsymbol{\zeta}_N$ and $\boldsymbol{\psi}_N = \mathbf{A}_N \boldsymbol{\zeta} \sim N(\mathbf{A}_N \boldsymbol{\mu}_N, \mathbf{A}_N \boldsymbol{\Sigma} \mathbf{A}_N')$. Let $F_N^{\eta}, F_N^{\xi},$ and F_N^{ψ} be the cumulative distribution functions of $\boldsymbol{\eta}_N, \boldsymbol{\xi}_N,$ and $\boldsymbol{\psi}_N$, respectively. ($F_N^{\psi}(x)$ is the cdf of a normal distribution with mean $\mathbf{A}_N \boldsymbol{\mu}_N$ and variance-covariance matrix $\mathbf{A}_N \boldsymbol{\Sigma} \mathbf{A}_N'$.) Assume further that $\liminf_{N \to \infty} \lambda_{\min}(\mathbf{A}_N \mathbf{A}_N') > 0$ holds. Then $F_N^{\eta}(x) - F_N^{\xi}(x) \to 0$ as $N \to \infty$ (i.e., the difference between the cdf of $\boldsymbol{\eta}_N$ and $\boldsymbol{\xi}_N$ converges to zero at all continuity points of the cdf of $\boldsymbol{\xi}_N$), and $F_N^{\eta}(x) - F_N^{\psi}(x) \to 0$ as $N \to \infty$. (i.e., the difference between the cdf of $\boldsymbol{\eta}_N$ and $\boldsymbol{\psi}_N$ converges to zero at all continuity points of the cdf of $\boldsymbol{\psi}_N$).

APPENDIX C

I. Proof of Theorem 1 (Consistency of $\widetilde{\boldsymbol{\rho}}_N$)

As a preliminary step, we now give a version of Lemma C.1 and Remark C.2 in Kelejian and Prucha (2008) that is applicable to the higher-order case.

Lemma C.1[16]

Suppose the row and the column sums of the real nonstochastic $N \times N$ matrices $\mathbf{A}_N$ are bounded uniformly in absolute value. Let $\mathbf{u}_N$ be defined by (1c) and let $\widetilde{\mathbf{u}}_N$ denote a predictor for $\mathbf{u}_N$. Suppose Assumptions 1-4 hold. Then

(a) $N^{-1} E\left|\mathbf{u}_N' \mathbf{A}_N \mathbf{u}_N\right| = O(1)$ and $Var(N^{-1} \mathbf{u}_N' \mathbf{A}_N \mathbf{u}_N) = o(1)$, and

$N^{-1}(\widetilde{\mathbf{u}}_N' \mathbf{A}_N \widetilde{\mathbf{u}}_N) - N^{-1} E(\mathbf{u}_N' \mathbf{A}_N \mathbf{u}_N) = o_p(1)$.

[16] Compare Kelejian and Prucha (2008, p. 32).

(b) $N^{-1}E\left|\mathbf{d}'_{.j,N}\mathbf{A}_N\mathbf{u}_N\right| = O(1)$, $j = 1,...,P$, where $\mathbf{d}_{.j,N}$ is the j-th column of the $N \times P$ matrix $\mathbf{D}_N$, and $N^{-1}\mathbf{D}'_N\mathbf{A}_N\widetilde{\mathbf{u}}_N - N^{-1}E(\mathbf{D}'_N\mathbf{A}_N\mathbf{u}_N) = o_p(1)$.

(c) Furthermore, if Assumption 6 holds, then
$N^{-1/2}\widetilde{\mathbf{u}}'_N\mathbf{A}_N\widetilde{\mathbf{u}}_N = N^{-1/2}\mathbf{u}'_N\mathbf{A}_N\mathbf{u}_N + \boldsymbol{\alpha}'_N N^{1/2}\boldsymbol{\Delta}_N + o_p(1)$ with $\boldsymbol{\alpha}_N = N^{-1}E[\mathbf{D}'_N(\mathbf{A}_N + \mathbf{A}'_N)\mathbf{u}_N]$.
In light of (b), we have $\boldsymbol{\alpha}_N = O(1)$ and $N^{-1}\mathbf{D}'_N(\mathbf{A}_N + \mathbf{A}'_N)\widetilde{\mathbf{u}}_N - \boldsymbol{\alpha}_N = o_p(1)$.

Proof of part (a)
Let

$$\vartheta_N = N^{-1}\mathbf{u}'_N\mathbf{A}_N\mathbf{u}_N \text{ and } \widetilde{\vartheta}_N = N^{-1}\widetilde{\mathbf{u}}'_N\mathbf{A}_N\widetilde{\mathbf{u}}_N \tag{C.1}$$

then given (1c), we have $\vartheta_N = N^{-1}\boldsymbol{\varepsilon}'_N\mathbf{B}_N\boldsymbol{\varepsilon}_N$, with[17]

$$\mathbf{B}_N = (1/2)(\mathbf{I}_N - \sum_{m=1}^{S}\rho_{m,N}\mathbf{M}'_{m,N})^{-1}(\mathbf{A}_N + \mathbf{A}'_N)(\mathbf{I}_N - \sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N})^{-1}. \tag{C.2}$$

By Assumption 3 and Remark A.1 in Appendix A, the row and column sums of the matrices $\mathbf{B}_N$ are uniformly bounded in absolute value. Let $\boldsymbol{\Sigma}_N = diag_{i=1}^{N}(\sigma^2_{1,N},...,\sigma^2_{N,N})$, then given Assumption 2 it follows that the row and column sums of the matrices $\mathbf{B}_N\boldsymbol{\Sigma}_N\mathbf{B}_N\boldsymbol{\Sigma}_N$ are uniformly bounded in absolute value.

In the following let $K < \infty$ be a common bound for the row and column sums of the absolute elements of $\mathbf{B}_N, \boldsymbol{\Sigma}_N$, and $\mathbf{B}_N\boldsymbol{\Sigma}_N\mathbf{B}_N\boldsymbol{\Sigma}_N$ and of their respective elements. Then, using Lemma A.1 in Kelejian and Prucha (2008), we have

$$\begin{aligned} E|\vartheta_N| &= E\left|N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}b_{ij,N}\varepsilon_{i,N}\varepsilon_{j,N}\right| \qquad (C.3)\\ &\le N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}|b_{ij,N}|E|\varepsilon_{i,N}\|\varepsilon_{j,N}| \\ &\le N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}|b_{ij,N}|\sigma_{i,N}\sigma_{j,N} \\ &\le K^3, \end{aligned}$$

[17] We use the fact that $\boldsymbol{\varepsilon}'_N\mathbf{A}_N\boldsymbol{\varepsilon}_N = \boldsymbol{\varepsilon}'_N\mathbf{A}'_N\boldsymbol{\varepsilon}_N = \boldsymbol{\varepsilon}'_N(\mathbf{A}_N + \mathbf{A}'_N)\boldsymbol{\varepsilon}_N/2$, which is a quadratic form in the symmetric matrix $(\mathbf{A}_N + \mathbf{A}'_N)/2$.

where we used Hölder's inequality in the last step. This proves that $E|\vartheta_N|$ is O(1).

Now consider $Var(\vartheta_N)$, invoking Lemma A.1 in Kelejian and Prucha (2008):

$$
\begin{aligned}
&Var(\vartheta_N) = Cov(N^{-1}\boldsymbol{\varepsilon}'_N\mathbf{B}_N\boldsymbol{\varepsilon}_N, N^{-1}\boldsymbol{\varepsilon}'_N\mathbf{B}_N\boldsymbol{\varepsilon}_N) \qquad \text{(C.4)}\\
&= 2N^{-2}Tr(\mathbf{B}_N\boldsymbol{\Sigma}_N\mathbf{B}_N\boldsymbol{\Sigma}_N) + N^{-2}\sum_{i=1}^{N} b_{ii,N}^2\sigma_{i,N}^4[E(\frac{\varepsilon_{i,N}^4}{\sigma_{i,N}^4}) - 3]\\
&= 2N^{-2}Tr(\mathbf{B}_N\boldsymbol{\Sigma}_N\mathbf{B}_N\boldsymbol{\Sigma}_N) + N^{-2}\sum_{i=1}^{N} b_{ii,N}^2\sigma_{i,N}^4[\frac{E(\varepsilon_{i,N}^4) - 3\sigma_{i,N}^4}{\sigma_{i,N}^4}]\\
&= 2N^{-2}Tr(\mathbf{B}_N\boldsymbol{\Sigma}_N\mathbf{B}_N\boldsymbol{\Sigma}_N) + N^{-2}\sum_{i=1}^{N} b_{ii,N}^2[E(\varepsilon_{i,N}^4) - 3\sigma_{i,N}^4]\\
&\le 2N^{-1}K + N^{-1}K^2\sup_{i=1,...,N}[E(\varepsilon_{i,N}^4) - 3\sigma_{i,N}^4].
\end{aligned}
$$

Since the fourth moments of $\varepsilon_{i,N}$ are uniformly bounded by Assumption 2, it follows that both terms converge to zero as $N \to \infty$. This establishes the claims in part (a) of Lemma C.1 that $N^{-1}(\mathbf{u}'_N\mathbf{A}_N\mathbf{u}_N) - N^{-1}E(\mathbf{u}'_N\mathbf{A}_N\mathbf{u}_N) \xrightarrow{p} \mathbf{0}$.

We now prove the second part of (a), i.e., $N^{-1}(\tilde{\mathbf{u}}'_N\mathbf{A}_N\tilde{\mathbf{u}}_N) - N^{-1}E(\mathbf{u}'_N\mathbf{A}_N\mathbf{u}_N) \xrightarrow{p} 0$. Since $\vartheta_N - E(\vartheta_N) = o_p(1)$, it is sufficient to show that $\tilde{\vartheta}_N - \vartheta_N = o_p(1)$. By Assumption 4, we have $\tilde{\mathbf{u}}_N - \mathbf{u}_N = \mathbf{D}_N\boldsymbol{\Delta}_N$, where $\mathbf{D}_N = (\mathbf{d}'_{1.,N},...,\mathbf{d}'_{N.,N})'$. Substituting $\tilde{\mathbf{u}}_N = \mathbf{u}_N + \mathbf{D}_N\boldsymbol{\Delta}_N$ into the expression for $\tilde{\vartheta}_N$ in (C.5), we obtain

$$
\begin{aligned}
&\tilde{\vartheta}_N - \vartheta_N = N^{-1}(\mathbf{u}'_N + \boldsymbol{\Delta}'_N\mathbf{D}'_N)\mathbf{A}_N(\mathbf{u}_N + \mathbf{D}_N\boldsymbol{\Delta}_N) - N^{-1}\mathbf{u}'_N\mathbf{A}_N\mathbf{u}_N \qquad \text{(C.5)}\\
&= N^{-1}(\mathbf{u}'_N\mathbf{A}_N\mathbf{u}_N + \boldsymbol{\Delta}'_N\mathbf{D}'_N\mathbf{A}_N\mathbf{u}_N + \mathbf{u}'_N\mathbf{A}_N\mathbf{D}_N\boldsymbol{\Delta}_N + \boldsymbol{\Delta}'_N\mathbf{D}'_N\mathbf{A}_N\mathbf{D}_N\boldsymbol{\Delta}_N - \mathbf{u}'_N\mathbf{A}_N\mathbf{u}_N)\\
&= N^{-1}(\boldsymbol{\Delta}'_N\mathbf{D}'_N\mathbf{A}_N\mathbf{u}_N + \mathbf{u}'_N\mathbf{A}_N\mathbf{D}_N\boldsymbol{\Delta}_N + \boldsymbol{\Delta}'_N\mathbf{D}'_N\mathbf{A}_N\mathbf{D}_N\boldsymbol{\Delta}_N)\\
&= N^{-1}(\boldsymbol{\Delta}'_N\mathbf{D}'_N\mathbf{A}_N\mathbf{u}_N + \boldsymbol{\Delta}'_N\mathbf{D}'_N\mathbf{A}'_N\mathbf{u}_N + \boldsymbol{\Delta}'_N\mathbf{D}'_N\mathbf{A}_N\mathbf{D}_N\boldsymbol{\Delta}_N)\\
&= N^{-1}[\boldsymbol{\Delta}'_N\mathbf{D}'_N(\mathbf{A}_N + \mathbf{A}'_N)\mathbf{u}_N + \boldsymbol{\Delta}'_N\mathbf{D}'_N\mathbf{A}_N\mathbf{D}_N\boldsymbol{\Delta}_N]\\
&= \phi_N + \psi_N,
\end{aligned}
$$

where

$$
\begin{aligned}
&\phi_N = N^{-1}[\boldsymbol{\Delta}'_N\mathbf{D}'_N(\mathbf{A}_N + \mathbf{A}'_N)\mathbf{u}_N] = N^{-1}[\boldsymbol{\Delta}'_N\mathbf{D}'_N(\mathbf{A}_N + \mathbf{A}'_N)(\mathbf{I}_N - \sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N})^{-1}\boldsymbol{\varepsilon}_N], \qquad \text{(C.6)}\\
&= N^{-1}(\boldsymbol{\Delta}'_N\mathbf{D}'_N\mathbf{C}_N\boldsymbol{\varepsilon}_N), \text{ where } \mathbf{C}_N = (\mathbf{A}_N + \mathbf{A}'_N)(\mathbf{I}_N - \sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N})^{-1} = (\mathbf{c}'_{1.,N},..,\mathbf{c}'_{N.,N})'
\end{aligned}
$$

and

$$\psi_N = N^{-1}\boldsymbol{\Delta}_N'\mathbf{D}_N'\mathbf{A}_N\mathbf{D}_N\boldsymbol{\Delta}_N . \tag{C.7}$$

By Assumption 3 and Remark A.1, the row and column sums of $\mathbf{C}_N$ are uniformly bounded in absolute value. Denote by K the uniform bound for the row and column sums of the matrices $\mathbf{A}_N$ and $\mathbf{C}_N$. We next prove that $\phi_N = o_p(1)$ and $\psi_N = o_p(1)$.

Proof that $\phi_N = o_p(1)$:

$$\begin{aligned}
|\phi_N| &= \left|N^{-1}\boldsymbol{\Delta}_N'\mathbf{D}_N'\mathbf{C}_N\boldsymbol{\varepsilon}_N\right| \qquad (C.8)\\
&= \left\|N^{-1}\sum_{i=1}^{N}\boldsymbol{\Delta}_N'\mathbf{d}_{i.,N}'\mathbf{c}_{i.,N}\boldsymbol{\varepsilon}_N\right\| \\
&\le \sum_{i=1}^{N}N^{-1}\left\|\boldsymbol{\Delta}_N'\right\|\left\|\mathbf{d}_{i.,N}'\right\|\left\|\mathbf{c}_{i.,N}\boldsymbol{\varepsilon}_N\right\| \\
&= N^{-1}\left\|\boldsymbol{\Delta}_N'\right\|\sum_{i=1}^{N}\left\|\mathbf{d}_{i.,N}'\right\|\left\|\sum_{j=1}^{N}c_{ij,N}\varepsilon_{j,N}\right\| \\
&\le N^{-1}\left\|\boldsymbol{\Delta}_N'\right\|\sum_{i=1}^{N}\left\|\mathbf{d}_{i.,N}'\right\|\sum_{j=1}^{N}\left|c_{ij,N}\varepsilon_{j,N}\right| \\
&= N^{-1}\left\|\boldsymbol{\Delta}_N'\right\|\sum_{i=1}^{N}\left\|\mathbf{d}_{i.,N}'\right\|\sum_{j=1}^{N}\left|c_{ij,N}\right|\left|\varepsilon_{j,N}\right| \\
&= N^{-1}\left\|\boldsymbol{\Delta}_N'\right\|\sum_{j=1}^{N}\left|\varepsilon_{j,N}\right|\sum_{i=1}^{N}\left\|\mathbf{d}_{i.,N}'\right\|\left|c_{ij,N}\right| \\
&\le N^{-1}\left\|\boldsymbol{\Delta}_N'\right\|\sum_{j=1}^{N}\left|\varepsilon_{j,N}\right|\left(\sum_{i=1}^{N}\left\|\mathbf{d}_{i.,N}'\right\|^{p}\right)^{1/p}\left(\sum_{i=1}^{N}\left|c_{ij,N}\right|^{q}\right)^{1/q} \\
&= N^{-1}N^{-1/2}N^{1}N^{1/p}\left(N^{1/2}\left\|\boldsymbol{\Delta}_N'\right\|\right)\left(N^{-1}\sum_{j=1}^{N}\left|\varepsilon_{j,N}\right|\right)\left(N^{-1}\sum_{i=1}^{N}\left\|\mathbf{d}_{i.,N}'\right\|^{p}\right)^{1/p}\left(\sum_{i=1}^{N}\left|c_{ij,N}\right|^{q}\right)^{1/q} \\
&= N^{1/p-1/2}\left(N^{1/2}\left\|\boldsymbol{\Delta}_N'\right\|\right)\left(N^{-1}\sum_{j=1}^{N}\left|\varepsilon_{j,N}\right|\right)\left(N^{-1}\sum_{i=1}^{N}\left\|\mathbf{d}_{i.,N}'\right\|^{p}\right)^{1/p}\left(\sum_{i=1}^{N}\left|c_{ij,N}\right|^{q}\right)^{1/q} .
\end{aligned}$$

Note that $\left(\sum_{i=1}^{N}\left|c_{ij,N}\right|\right) \le \overline{K}$ by Assumption. From Remark C.1 in Kelejian and Prucha (2008), see Appendix A, it follows that $\left(\sum_{i=1}^{N}\left|c_{ij,N}\right|^{q}\right) \le \overline{K}^{q}$ (for $q>1$) and thus $\left(\sum_{i=1}^{N}\left\|c_{ij,N}\right\|^{q}\right)^{1/q} \le \overline{K}$.

Factoring $\overline{K}$ out of the sum yields the final expression

$$|\phi_N| \le \overline{K} N^{1/p-1/2}\left(N^{1/2}\|\boldsymbol{\Delta}'_N\|\right)\left(N^{-1}\sum_{j=1}^{N}|\varepsilon_{j,N}|\right)\left(N^{-1}\sum_{i=1}^{N}\|\mathbf{d}'_{i.,N}\|^p\right)^{1/p}.$$

This holds for $p = 2+\delta$ for some $\delta > 0$ as in Assumption 4 and $1/p + 1/q = 1$. The innovations ε_i are independent and have bounded second moments by Assumption 2. It follows that $N^{-1}\sum_{j=1}^{N}|\varepsilon_{j,N}| = O_p(1)$.

Moreover, it follows from Assumption 4 that $\left(N^{1/2}\|\boldsymbol{\Delta}'_N\|\right) = O_p(1)$ and $\left(N^{-1}\sum_{i=1}^{N}\|\mathbf{d}'_{i.,N}\|^p\right)^{1/p} = O_p(1)$ for $p = 2+\delta$ and some $\delta > 0$. Since $N^{1/p-1/2} \to 0$ as $N \to \infty$ it follows that $|\phi_N| = o_p(1)$.

Similarly, we have

$$|\psi_N| = \left|N^{-1}\boldsymbol{\Delta}'_N\mathbf{D}'_N\mathbf{A}_N\mathbf{D}_N\boldsymbol{\Delta}_N\right| = N^{-1}\left|\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{\Delta}'_N\mathbf{d}'_{i.,N}\mathbf{d}'_{i.,N}a_{ij,N}\mathbf{d}_{j.,N}\boldsymbol{\Delta}_N\right| \tag{C.9}$$

$$\le N^{-1}\|\boldsymbol{\Delta}_N\|^2\sum_{i=1}^{N}\|\mathbf{d}'_{i.,N}\|\left\|\sum_{j=1}^{N}\|\mathbf{d}_{j.,N}\|\,|a_{ij,N}|\right.$$

$$\le N^{-1}\|\boldsymbol{\Delta}_N\|^2\sum_{i=1}^{N}\|\mathbf{d}'_{i.,N}\|\left(\sum_{j=1}^{N}\|\mathbf{d}_{j.,N}\|^p\right)^{1/p}\left(\sum_{j=1}^{N}|a_{ij,N}|^q\right)^{1/q}$$

$$\le N^{1/p}\overline{K}\|\boldsymbol{\Delta}_N\|^2\left(N^{-1}\sum_{i=1}^{N}\|\mathbf{d}'_{i.,N}\|\right)\left(N^{-1}\sum_{j=1}^{N}\|\mathbf{d}_{j.,N}\|^p\right)^{1/p}$$

$$= N^{1/p-1/2}N^{-1/2}\overline{K}\left(N^{1/2}\|\boldsymbol{\Delta}_N\|\right)^2\left(N^{-1}\sum_{j=1}^{N}\|\mathbf{d}_{j.,N}\|^p\right)^{2/p} = o_p(1).$$

Summing up, we have proved that $\widetilde{\vartheta}_N - \vartheta_N = \phi_N + \psi_N = o_p(1)$.

Proof of part (b)

Denote by $\vartheta^*_{s,N}$ the s-th element of $N^{-1}\mathbf{D}'_N\mathbf{A}_N\mathbf{u}_N$. In light of the discussion after Assumption 3 and given Assumption 4 there exists a constant $\overline{\overline{K}} < \infty$ such that $E(u^2_{i,N}) \le \overline{\overline{K}}$ and $E|d_{ij,N}|^p \le \overline{\overline{K}}$ with $p = 2+\delta$ for some $\delta > 0$. Without loss of generality we assume that the row and column sums of the matrices $\mathbf{A}_N$ are uniformly bounded by $\overline{\overline{K}} < \infty$. Notice first that, using the Cauchy-Schwarz and Lyapunov inequalities, we have

$$E|u_{i,N}\|d_{js,N}| \le \left(Eu^2_{i,N}\right)^{1/2}\left(Ed^2_{js,N}\right)^{1/2}$$

$$\leq \left(Eu_{i,N}^2\right)^{1/2}\left(E\left|d_{js,N}\right|^p\right)^{1/p}$$

$$\leq \overline{\overline{K}}^{1/2}\overline{\overline{K}}^{1/p} = \overline{\overline{K}}^{1/2+1/p} \text{ with } p \text{ as before.}$$

It follows that

$$E\left|\vartheta_{s,N}^*\right| = N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|a_{ij,N}\right| E\left|u_{i,N}\right|\left|d_{js,N}\right| \tag{C.10}$$

$$\leq \overline{\overline{K}}^{1/2+1/p} N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|a_{ij,N}\right| \leq \overline{\overline{K}}^{1/2+1/p} N^{-1} N\overline{\overline{K}} = \overline{\overline{K}}^{3/2+1/p} < \infty,$$

which shows that indeed $E\left|N^{-1}\mathbf{d}'_{.s,N}\mathbf{A}_N\mathbf{u}_N\right| = O(1)$. Of course, the argument also shows that

$$\boldsymbol{\alpha}_N = N^{-1}E[\mathbf{D}'_N(\mathbf{A}_N + \mathbf{A}'_N)\mathbf{u}_N] = O(1).$$

Next, observe that

$$N^{-1}\mathbf{D}'_N\mathbf{A}_N\widetilde{\mathbf{u}}_N = N^{-1}\mathbf{D}'_N\mathbf{A}_N\mathbf{u}_N + \phi_N^*, \tag{C.11}$$

where $\phi_N^* = N^{-1}\mathbf{D}'_N\mathbf{A}_N\mathbf{D}_N\boldsymbol{\Delta}_N$. It now follows from the demonstration of $\phi_N = N^{-1}[\boldsymbol{\Delta}'_N\mathbf{D}'_N(\mathbf{A}_N + \mathbf{A}'_N)\mathbf{u}_N] = o_p(1)$ that also $\phi_N^* = o_p(1)$.

<u>Proof of part (c)</u>
In light of the proof of part (a) and using $\widetilde{\mathbf{u}}_N = \mathbf{u}_N + \mathbf{D}_N\boldsymbol{\Delta}_N$,

$$N^{-1/2}\widetilde{\mathbf{u}}'_N\mathbf{A}_N\widetilde{\mathbf{u}}_N = N^{-1/2}(\mathbf{u}'_N + \boldsymbol{\Delta}'_N\mathbf{D}'_N)\mathbf{A}_N(\mathbf{u}_N + \mathbf{D}_N\boldsymbol{\Delta}_N) \tag{C.12}$$

$$= N^{-1/2}\mathbf{u}'_N\mathbf{A}_N\mathbf{u}_N + N^{1/2}\boldsymbol{\Delta}'_N N^{-1}E[\mathbf{D}'_N(\mathbf{A}_N + \mathbf{A}'_N)\mathbf{u}_N] + N^{1/2}\psi_N + o_p(1),$$

where $N^{1/2}\psi_N = o_p(1)$, compare (C.9).

In light of (b) and since $N^{1/2}\boldsymbol{\Delta}'_N = O_p(1)$ by Assumption 4, we have

$$N^{-1/2}\widetilde{\mathbf{u}}'_N\mathbf{A}_N\widetilde{\mathbf{u}}_N = N^{-1/2}\mathbf{u}'_N\mathbf{A}_N\mathbf{u}_N + N^{1/2}\boldsymbol{\alpha}'_N\boldsymbol{\Delta}_N + o_p(1). \tag{C.13}$$

Remark C.2.[18]

In light of Remark A.1, the constant $\overline{K}$ used in the proof of Lemma C.1, part (a), can be chosen such that $\overline{K} = 2c_p c_A$, where c_p and c_A are the bounds for the row and column sums of the matrices $\mathbf{P}_N = (\mathbf{I}_N - \sum_{m=1}^{S} \rho_m \mathbf{M}_N)^{-1}$ and $\mathbf{A}_N$ respectively. Furthermore, notice that $\left|\widetilde{\vartheta}_N - \vartheta_N\right| \leq 2c_A c_P \varsigma_N$ with $\varsigma_N = o_p(1)$.

Proof of Theorem 1.

The objective function of the weighted nonlinear least squares estimator defined by (10) and its nonstochastic counterpart are given by

$$R_N(\omega, \boldsymbol{\rho}) = (\widetilde{\boldsymbol{\gamma}}_N - \widetilde{\boldsymbol{\Gamma}}_N \mathbf{b})' \widetilde{\boldsymbol{\Theta}}_N (\widetilde{\boldsymbol{\gamma}}_N - \widetilde{\boldsymbol{\Gamma}}_N \mathbf{b}) \text{ and} \tag{C.14}$$

$$\overline{R}_N(\boldsymbol{\rho}) = (\boldsymbol{\gamma}_N - \boldsymbol{\Gamma}_N \mathbf{b})' \boldsymbol{\Theta}_N (\boldsymbol{\gamma}_N - \boldsymbol{\Gamma}_N \mathbf{b}). \tag{C.15}$$

In general, $\overline{R}_N(\boldsymbol{\rho}) \geq 0$ and in light of (8), $\overline{R}_N(\boldsymbol{\rho}) = 0$ for $\boldsymbol{\rho} = \boldsymbol{\rho}_N$, i.e., the objective function is zero when evaluated at the true parameter value.

Using $\boldsymbol{\gamma}_N = \boldsymbol{\Gamma}_N \mathbf{b}_N$, we obtain

$$\begin{aligned} &\overline{R}_N(\boldsymbol{\rho}) - \overline{R}_N(\boldsymbol{\rho}_N) = \overline{R}_N(\boldsymbol{\rho}) \\ &= (\boldsymbol{\Gamma}_N \mathbf{b}_N - \boldsymbol{\Gamma}_N \mathbf{b})' \boldsymbol{\Theta}_N (\boldsymbol{\Gamma}_N \mathbf{b}_N - \boldsymbol{\Gamma}_N \mathbf{b}). \\ &= (\mathbf{b}_N - \mathbf{b})' \boldsymbol{\Gamma}'_N \boldsymbol{\Theta}_N \boldsymbol{\Gamma}_N (\mathbf{b}_N - \mathbf{b}) \end{aligned} \tag{C.16}$$

In light of Rao (1973, p. 62) and Mittelhammer (1996, p. 254)

$$\begin{aligned} &\overline{R}_N(\boldsymbol{\rho}) - \overline{R}_N(\boldsymbol{\rho}_N) \geq \lambda_{\min}(\boldsymbol{\Gamma}'_N \boldsymbol{\Theta}_N \boldsymbol{\Gamma}_N)(\mathbf{b}_N - \mathbf{b})'(\mathbf{b}_N - \mathbf{b}) \text{ and} \\ &\geq \lambda_{\min}(\boldsymbol{\Gamma}'_N \boldsymbol{\Gamma}_N) \lambda_{\min}(\boldsymbol{\Theta}_N)(\mathbf{b}_N - \mathbf{b})'(\mathbf{b}_N - \mathbf{b}) \\ &\geq \lambda_* \left\|\boldsymbol{\rho}_N - \boldsymbol{\rho}\right\|^2 \end{aligned} \tag{C.17}$$

for some $\lambda_* > 0$ by Assumption 5.

Hence, for every $\varepsilon > 0$ and every N we have

$$\inf_{-\mathbf{a}^\rho \leq \boldsymbol{\rho} \leq \mathbf{a}^\rho, \left\|\boldsymbol{\rho}_N - \boldsymbol{\rho}\right\| \geq \varepsilon} [\overline{R}_N(\boldsymbol{\rho}) - \overline{R}_N(\boldsymbol{\rho}_N)] \geq \inf_{-\mathbf{a}^\rho \leq \boldsymbol{\rho} \leq \mathbf{a}^\rho, \left\|\boldsymbol{\rho}_N - \boldsymbol{\rho}\right\| \geq \varepsilon} \lambda_* \left\|\boldsymbol{\rho}_N - \boldsymbol{\rho}\right\|^2 = \lambda_* \varepsilon^2 > 0, \tag{C.18}$$

[18] Compare (Kelejian and Prucha, 2008, p. 35).

which proves that the true parameter vector $\boldsymbol{\rho}_N$ is identifiable unique (compare Lemma Lemma 4.1 in Pötscher and Prucha, 1997).

Next, let $\boldsymbol{\Phi}_N = (\boldsymbol{\gamma}_N, -\boldsymbol{\Gamma}_N)$ and $\widetilde{\boldsymbol{\Phi}}_N = (\widetilde{\boldsymbol{\gamma}}_N, -\widetilde{\boldsymbol{\Gamma}}_N)$, then the objective function and its nonstochastic counterpart can be written as

$$R_N(\omega, \boldsymbol{\rho}) = (1, \mathbf{b}')\widetilde{\boldsymbol{\Phi}}'_N \widetilde{\boldsymbol{\Theta}}_N \widetilde{\boldsymbol{\Phi}}_N (1, \mathbf{b}')' \text{ and} \tag{C.19}$$

$$\overline{R}_N(\boldsymbol{\rho}) = (1, \mathbf{b}')\boldsymbol{\Phi}'_N \boldsymbol{\Theta}_N \boldsymbol{\Phi}_N (1, \mathbf{b}')'. \tag{C.20}$$

It follows that

$$\begin{aligned} \left|R_N(\omega, \boldsymbol{\rho}) - \overline{R}_N(\boldsymbol{\rho})\right| &= \left|(1, \mathbf{b}')(\widetilde{\boldsymbol{\Phi}}'_N \widetilde{\boldsymbol{\Theta}}_N \widetilde{\boldsymbol{\Phi}}_N - \boldsymbol{\Phi}'_N \boldsymbol{\Theta}_N \boldsymbol{\Phi}_N)(1, \mathbf{b}')'\right| \\ &\leq \left\|(1, \mathbf{b}')\right\| \left\|\widetilde{\boldsymbol{\Phi}}'_N \widetilde{\boldsymbol{\Theta}}_N \widetilde{\boldsymbol{\Phi}}_N - \boldsymbol{\Phi}'_N \boldsymbol{\Theta}_N \boldsymbol{\Phi}_N\right\| \left\|(1, \mathbf{b}')'\right\| \\ &\leq \left\|\widetilde{\boldsymbol{\Phi}}'_N \widetilde{\boldsymbol{\Theta}}_N \widetilde{\boldsymbol{\Phi}}_N - \boldsymbol{\Phi}'_N \boldsymbol{\Theta}_N \boldsymbol{\Phi}_N\right\| \left\|(1, \mathbf{b}')'\right\|^2 \\ &\leq \left\|\widetilde{\boldsymbol{\Phi}}'_N \widetilde{\boldsymbol{\Theta}}_N \widetilde{\boldsymbol{\Phi}}_N - \boldsymbol{\Phi}'_N \boldsymbol{\Theta}_N \boldsymbol{\Phi}_N\right\| [1 + S(a^\rho)^2 + \frac{2S + S(S-1)}{2}(a^\rho)^4]. \end{aligned} \tag{C.21}$$

As can be seen from the right-hand side of (11), the elements of $\boldsymbol{\Phi}_N = [\boldsymbol{\gamma}_N, -\boldsymbol{\Gamma}_N]$ are all of the form $N^{-1}E(\mathbf{u}'_N \mathbf{A}_N \mathbf{u}_N)$ and $N^{-1}(\widetilde{\mathbf{u}}'_N \mathbf{A}_N \widetilde{\mathbf{u}}_N)$, where the row and column sums of the matrices $\mathbf{A}_N$ are bounded uniformly in absolute value (see Remark A.1). It now follows from Lemma C.1 that $\boldsymbol{\Phi}_N - \widetilde{\boldsymbol{\Phi}}_N \xrightarrow{p} \mathbf{0}$ as $N \to \infty$, and that the elements of $\boldsymbol{\Phi}_N$ and $\widetilde{\boldsymbol{\Phi}}_N$ are $O(1)$ and $O_p(1)$ respectively. The analogous properties are seen to hold for $\boldsymbol{\Theta}_N$ and $\widetilde{\boldsymbol{\Theta}}_N$, i.e., the elements of $\boldsymbol{\Theta}_N$ and $\widetilde{\boldsymbol{\Theta}}_N$ are $O(1)$ and $O_p(1)$ respectively and $\boldsymbol{\Theta}_N - \widetilde{\boldsymbol{\Theta}}_N \xrightarrow{p} \mathbf{0}$ by Assumption A.5. It follows from the above inequality that $\left|R_N(\omega, \boldsymbol{\rho}) - \overline{R}_N(\boldsymbol{\rho})\right|$ converges to zero uniformly over the optimization space $-\mathbf{a}^\rho \leq \boldsymbol{\rho} \leq \mathbf{a}^\rho$, i.e.,

$$\begin{aligned} &\sup_{-\mathbf{a}^\rho \leq \boldsymbol{\rho} \leq \mathbf{a}^\rho} \left|R_N(\omega, \boldsymbol{\rho}) - \overline{R}_N(\boldsymbol{\rho})\right| \\ &\leq \left\|\widetilde{\boldsymbol{\Phi}}'_N \widetilde{\boldsymbol{\Theta}}_N \widetilde{\boldsymbol{\Phi}}_N - \boldsymbol{\Phi}'_N \boldsymbol{\Theta}_N \boldsymbol{\Phi}_N\right\| [1 + S(a^\rho)^2 + \frac{2S + S(S-1)}{2}(a^\rho)^4] \xrightarrow{p} 0 \text{ as } N \to \infty, \end{aligned} \tag{C.22}$$

The consistency of $\widetilde{\boldsymbol{\rho}}_N = (\widetilde{\rho}_{1,N}, ..., \widetilde{\rho}_{S,N})'$ now follows directly from Lemma 3.1 in Pötscher and Prucha (1997).

II. Proof of Theorem 2 (Asymptotic Normality of $\widetilde{\boldsymbol{\rho}}_N$)

The limiting distribution of the GM estimator of $\boldsymbol{\rho}_N$ will be seen to depend on (the inverse of) $\mathbf{J}_N'\boldsymbol{\Theta}_N\mathbf{J}_N$ and the variance-covariance matrix of the vector of quadratic forms $\mathbf{v}_N$ as defined by (12).

We first consider $\mathbf{J}_N$, the $2S \times S$ matrix of derivatives of the $2S \times 1$ vector of moment conditions given by (8a):

$$\mathbf{J}_N(\boldsymbol{\rho}) = \frac{\partial(\boldsymbol{\gamma}_N - \boldsymbol{\Gamma}_N\mathbf{b}_N)}{\partial\boldsymbol{\rho}'} = [\frac{\partial(\boldsymbol{\gamma}_{i.,N} - \boldsymbol{\Gamma}_{i.,N}\mathbf{b}_N)}{\partial\rho_1},...,\frac{\partial(\boldsymbol{\gamma}_{i.,N} - \boldsymbol{\Gamma}_{i.,N}\mathbf{b}_N)}{\partial\rho_S}], \tag{C.23}$$

$$= (j_{is,N}) = [\frac{\partial(\boldsymbol{\gamma}_{i.,N} - \boldsymbol{\Gamma}_{i.,N}\mathbf{b}_N)}{\partial\rho_s}],\ i = 1,...,2S,\ s = 1,...,S,$$

where $\boldsymbol{\gamma}_{i.,N}$ and $\boldsymbol{\Gamma}_{i.,N}$ denote the *i*-th row of $\boldsymbol{\gamma}_N$ and $\boldsymbol{\Gamma}_N$.

Accounting for the fact that $\frac{\partial(\boldsymbol{\gamma}_N)}{\partial\boldsymbol{\rho}'} = \mathbf{0}$ (and ignoring the negative sign), we have

$$\mathbf{J}_N(\boldsymbol{\rho}) = \frac{\partial}{\partial\boldsymbol{\rho}'}\boldsymbol{\Gamma}_N\mathbf{b}_N = \boldsymbol{\Gamma}_N\mathbf{B}_N, \tag{C.24}$$

where $\mathbf{B}_N$ is a $[2S + S(S-1)/2] \times S$ matrix, which is defined as follows:

$$\mathbf{B}_N = (\mathbf{B}'_{1,N}, \mathbf{B}'_{2,N}, \mathbf{B}'_{3,N})', \tag{C.25}$$

with

$$\mathbf{B}_{1N} = \mathbf{I}_S, \tag{C.26}$$

$$\mathbf{B}_{2N} = diag_{s=1}^{S}(2\rho_{s,N}), \tag{C.27}$$

and $\mathbf{B}_{3N} = (\mathbf{B}'_{3,1,N},...,\mathbf{B}'_{3,S-1,N})'$ is an $S(S-1)/2 \times S$ matrix, consisting of $(S-1)$ vertically arranged blocks $\mathbf{B}_{3,m,N}$, $m = 1,...,(S-1)$, which have the following structure:

$$\mathbf{B}_{3,m,N} = (\mathbf{C}_{m,N}, \mathbf{d}_{m,N}, \mathbf{E}_{m,N}), \text{ where} \tag{C.28}$$

$\mathbf{C}_{m,N}$ is a $(S-m)\times(m-1)$ matrix of zeros[19], $\mathbf{d}_{m,N}$ is a $(S-m)\times 1$ vector, defined as $\mathbf{d}_{m,N}=(\rho_{m+1},...,\rho_S)$, and $\mathbf{E}_{m,N}=\rho_m\mathbf{I}_{S-m}$.

For later reference, note that $\mathbf{B}_N$ has full column rank (*S*); as a consequence, the $S\times S$ matrix $\mathbf{B}'_N\mathbf{B}_N$ is positive definite (see, e.g., Greene, 2003, p. 835).

Next, define

$$\mathbf{q}_N(\boldsymbol{\rho}_N,\boldsymbol{\Delta}_N)=\tilde{\boldsymbol{\gamma}}_N-\tilde{\boldsymbol{\Gamma}}_N\mathbf{b}_N=\begin{bmatrix} N^{-1}\tilde{\mathbf{u}}'_N\mathbf{C}_{11,N}\tilde{\mathbf{u}}_N \\ N^{-1}\tilde{\mathbf{u}}'_N\mathbf{C}_{21,N}\tilde{\mathbf{u}}_N \\ \cdot \\ \cdot \\ N^{-1}\tilde{\mathbf{u}}'_N\mathbf{C}_{1S,N}\tilde{\mathbf{u}}_N \\ N^{-1}\tilde{\mathbf{u}}'_N\mathbf{C}_{2S,N}\tilde{\mathbf{u}}_N \end{bmatrix}, \tag{C.29}$$

where the $N\times N$ matrices $\mathbf{C}_{1s,N}$ and $\mathbf{C}_{2s,N}$, $s=1,...,S$ are defined as follows:

$$\mathbf{C}_{1s,N}=1/2(\mathbf{I}_N-\sum_{m=1}^{S}\rho_{m,N}\mathbf{M}'_{m,N})(\mathbf{A}_{1s,N}+\mathbf{A}'_{1s,N})(\mathbf{I}_N-\sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N}), \tag{C.30}$$

$$\mathbf{C}_{2s,N}=1/2(\mathbf{I}_N-\sum_{m=1}^{S}\rho_{m,N}\mathbf{M}'_{m,N})(\mathbf{A}_{2s,N}+\mathbf{A}'_{2s,N})(\mathbf{I}_N-\sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N}). \tag{C.31}$$

By the properties of $\mathbf{M}_{s,N}$ and in light of Remark A.1 in Appendix A, the row and column sums of the matrices $\mathbf{C}_{1s,N}$ and $\mathbf{C}_{2s,N}$, $s=1,...,S$, are uniformly bounded in absolute value.

In Theorem 1 we showed that the GM estimator $\tilde{\boldsymbol{\rho}}_N$ defined by (9) is consistent. It follows that – apart from a set of the sample space whose probability tends to zero – the estimator satisfies the following first-order condition:

$$\frac{\partial}{\partial\boldsymbol{\rho}}\mathbf{q}_N(\tilde{\boldsymbol{\rho}}_N,\boldsymbol{\Delta}_N)'\tilde{\boldsymbol{\Theta}}_N\mathbf{q}_N(\tilde{\boldsymbol{\rho}}_N,\boldsymbol{\Delta}_N)=\frac{\partial\mathbf{q}_N(\tilde{\boldsymbol{\rho}}_N,\boldsymbol{\Delta}_N)}{\partial\boldsymbol{\rho}}\tilde{\boldsymbol{\Theta}}_N\mathbf{q}_N(\tilde{\boldsymbol{\rho}}_N,\boldsymbol{\Delta}_N)=\mathbf{0}, \tag{C.32}$$

which is a $S\times 1$ vector, each row *s* corresponding the partial derivative of the criterion function with respect to ρ_s.[20]

[19] I.e., there is no block $\mathbf{C}_{1,N}$ in the first line.

[20] The leading two and the negative sign are ignored without further consequences for the proof.

Substituting the mean value theorem expression

$$\mathbf{q}_N(\widetilde{\boldsymbol{\rho}}_N, \boldsymbol{\Delta}_N) = \mathbf{q}_N(\boldsymbol{\rho}_N, \boldsymbol{\Delta}_N) + \frac{\partial \mathbf{q}_N(\overline{\boldsymbol{\rho}}_N, \boldsymbol{\Delta}_N)}{\partial \boldsymbol{\rho}'}(\widetilde{\boldsymbol{\rho}}_N - \boldsymbol{\rho}_N), \tag{C.33}$$

where $\overline{\boldsymbol{\rho}}_N$ is some between value, into the first-order condition yields

$$\frac{\partial \mathbf{q}_N(\widetilde{\boldsymbol{\rho}}_N, \boldsymbol{\Delta}_N)}{\partial \boldsymbol{\rho}} \widetilde{\boldsymbol{\Theta}}_N \frac{\partial \mathbf{q}_N(\overline{\boldsymbol{\rho}}_N, \boldsymbol{\Delta}_N)}{\partial \boldsymbol{\rho}'} N^{1/2}(\widetilde{\boldsymbol{\rho}}_N - \boldsymbol{\rho}_N) = -\frac{\partial \mathbf{q}_N(\widetilde{\boldsymbol{\rho}}_N, \boldsymbol{\Delta}_N)}{\partial \boldsymbol{\rho}} \widetilde{\boldsymbol{\Theta}}_N N^{1/2} \mathbf{q}_N(\boldsymbol{\rho}_N, \boldsymbol{\Delta}_N). \tag{C.34}$$

Observe that $\frac{\partial \mathbf{q}_N(\boldsymbol{\rho}, \boldsymbol{\Delta}_N)}{\partial \boldsymbol{\rho}'} = \widetilde{\boldsymbol{\Gamma}}_N \mathbf{B}_N$ and consider the two $S \times S$ matrices

$$\widetilde{\boldsymbol{\Xi}}_N = \frac{\partial \mathbf{q}_N(\widetilde{\boldsymbol{\rho}}_N, \boldsymbol{\Delta}_N)}{\partial \boldsymbol{\rho}} \widetilde{\boldsymbol{\Theta}}_N \frac{\partial \mathbf{q}_N(\overline{\boldsymbol{\rho}}_N, \boldsymbol{\Delta}_N)}{\partial \boldsymbol{\rho}'} = \widetilde{\mathbf{B}}'_N \widetilde{\boldsymbol{\Gamma}}'_N \widetilde{\boldsymbol{\Theta}}_N \widetilde{\boldsymbol{\Gamma}}_N \overline{\mathbf{B}}_N, \text{ and} \tag{C.35}$$

$$\boldsymbol{\Xi}_N = \mathbf{B}'_N \boldsymbol{\Gamma}'_N \boldsymbol{\Theta}_N \boldsymbol{\Gamma}_N \mathbf{B}_N, \tag{C.36}$$

where $\widetilde{\mathbf{B}}_N$ and $\overline{\mathbf{B}}_N$ correspond to $\mathbf{B}_N$ as defined above with $\widetilde{\boldsymbol{\rho}}_N$ and $\overline{\boldsymbol{\rho}}_N$ substituted for $\boldsymbol{\rho}_N$. Notice that $\boldsymbol{\Xi}_N$ is positive definite, since $\boldsymbol{\Gamma}_N$ and $\boldsymbol{\Theta}_N$ are positive definite by assumption and the $[2S + S(S-1)/2] \times S$ matrix $\mathbf{B}_N$ has full column rank.

In the proof of Theorem 1 (and Lemma C.1) we have demonstrated that $\widetilde{\boldsymbol{\Gamma}}_N - \boldsymbol{\Gamma}_N \xrightarrow{p} \mathbf{0}$ and that the elements of $\boldsymbol{\Gamma}_N$ and $\widetilde{\boldsymbol{\Gamma}}_N$ are $O(1)$ and $O_p(1)$, respectively. By Assumption 5 we have $\widetilde{\boldsymbol{\Theta}}_N - \boldsymbol{\Theta}_N = o_p(1)$ and also that the elements of $\boldsymbol{\Theta}_N$ and $\widetilde{\boldsymbol{\Theta}}_N$ are $O(1)$ and $O_p(1)$. Since $\widetilde{\boldsymbol{\rho}}_N$ and $\overline{\boldsymbol{\rho}}_N$ (and thus also $\widetilde{\mathbf{B}}_N$ and $\overline{\mathbf{B}}_N$) are consistent and bounded in probability, it follows that $\widetilde{\boldsymbol{\Xi}}_N - \boldsymbol{\Xi}_N \xrightarrow{p} \mathbf{0}$ as $N \to \infty$ and furthermore $\widetilde{\boldsymbol{\Xi}}_N = O_p(1)$ and $\boldsymbol{\Xi}_N = O(1)$. Moreover, $\boldsymbol{\Xi}_N$ is positive definite and thus invertible, and its inverse $\boldsymbol{\Xi}_N^{-1}$ is also $O(1)$.

Denote $\widetilde{\boldsymbol{\Xi}}_N^+$ as the generalized inverse of $\widetilde{\boldsymbol{\Xi}}_N$. It then follows as a special case of Lemma F1 in Pötscher and Prucha (1997) (see Appendix B) that $\widetilde{\boldsymbol{\Xi}}_N$ is non-singular with probability approaching 1, that $\widetilde{\boldsymbol{\Xi}}_N^+$ is $O_p(1)$, and that $\widetilde{\boldsymbol{\Xi}}_N^+ - \boldsymbol{\Xi}_N^{-1} \xrightarrow{p} \mathbf{0}$ as $N \to \infty$.

Premultiplying (C.34) with $\widetilde{\boldsymbol{\Xi}}_N^+$ we obtain, after rearranging terms,

$$N^{1/2}(\widetilde{\boldsymbol{\rho}}_N - \boldsymbol{\rho}_N) = (\mathbf{I}_S - \widetilde{\boldsymbol{\Xi}}_N^{+}\widetilde{\boldsymbol{\Xi}}_N)N^{1/2}(\widetilde{\boldsymbol{\rho}}_N - \boldsymbol{\rho}_N) - N^{1/2}\widetilde{\boldsymbol{\Xi}}_N^{+}\frac{\partial \mathbf{q}_N(\widetilde{\boldsymbol{\rho}}_N, \boldsymbol{\Delta}_N)}{\partial \boldsymbol{\rho}}\widetilde{\boldsymbol{\Theta}}_N \mathbf{q}_N(\boldsymbol{\rho}_N, \boldsymbol{\Delta}_N) \quad \text{(C.37)}$$

In light of the discussion above the first term on the right-hand side is zero on ω-sets of probability approaching 1 (compare Pötscher and Prucha, 1997, p. 228ff.). This yields

$$N^{1/2}(\widetilde{\boldsymbol{\rho}}_N - \boldsymbol{\rho}_N) = -\widetilde{\boldsymbol{\Xi}}_N^{+}\frac{\partial \mathbf{q}_N(\widetilde{\boldsymbol{\rho}}_N, \boldsymbol{\Delta}_N)}{\partial \boldsymbol{\rho}}\widetilde{\boldsymbol{\Theta}}_N N^{1/2}\mathbf{q}_N(\boldsymbol{\rho}_N, \boldsymbol{\Delta}_N) + o_p(1)\,. \quad \text{(C.38)}$$

Next observe that

$$\widetilde{\boldsymbol{\Xi}}_N^{+}\frac{\partial \mathbf{q}_N(\widetilde{\boldsymbol{\rho}}_N, \boldsymbol{\Delta}_N)}{\partial \boldsymbol{\rho}}\widetilde{\boldsymbol{\Theta}}_N - \boldsymbol{\Xi}_N^{-1}\mathbf{B}'_N\boldsymbol{\Gamma}'_N\boldsymbol{\Theta}_N = o_p(1)\text{, since} \quad \text{(C.39)}$$

$$\widetilde{\boldsymbol{\Xi}}_N^{+} - \boldsymbol{\Xi}_N^{-1} = o_p(1) \text{ and } \frac{\partial \mathbf{q}_N(\widetilde{\boldsymbol{\rho}}_N, \boldsymbol{\Delta}_N)}{\partial \boldsymbol{\rho}} - \mathbf{B}'_N\boldsymbol{\Gamma}'_N = o_p(1)\,. \quad \text{(C.40)}$$

We next consider the distribution of the vector $N^{1/2}\mathbf{q}_N(\boldsymbol{\rho}_N, \boldsymbol{\Delta}_N)$. In light of (C.29) and Lemma C.1 the elements of $N^{1/2}\mathbf{q}_N(\boldsymbol{\rho}_N, \boldsymbol{\Delta}_N)$ can be expressed as

$$N^{1/2}\mathbf{q}_N(\boldsymbol{\rho}_N, \boldsymbol{\Delta}_N) = \begin{bmatrix} N^{-1/2}\widetilde{\mathbf{u}}'_N\mathbf{C}_{11,N}\widetilde{\mathbf{u}}_N \\ N^{-1/2}\widetilde{\mathbf{u}}'_N\mathbf{C}_{21,N}\widetilde{\mathbf{u}}_N \\ \cdot \\ \cdot \\ N^{-1/2}\widetilde{\mathbf{u}}'_N\mathbf{C}_{1S,N}\widetilde{\mathbf{u}}_N \\ N^{-1/2}\widetilde{\mathbf{u}}'_N\mathbf{C}_{2S,N}\widetilde{\mathbf{u}}_N \end{bmatrix} = \begin{bmatrix} N^{-1/2}\mathbf{u}'_N\mathbf{C}_{11N}\mathbf{u}_N + \boldsymbol{\alpha}'_{11,N}N^{1/2}\boldsymbol{\Delta}_N + o_p(1) \\ N^{-1/2}\mathbf{u}'_N\mathbf{C}_{21,N}\mathbf{u}_N + \boldsymbol{\alpha}'_{21,N}N^{1/2}\boldsymbol{\Delta}_N + o_p(1) \\ \cdot \\ \cdot \\ N^{-1/2}\mathbf{u}'_N\mathbf{C}_{1S,N}\mathbf{u}_N + \boldsymbol{\alpha}'_{1S,N}N^{1/2}\boldsymbol{\Delta}_N + o_p(1) \\ N^{-1/2}\mathbf{u}'_N\mathbf{C}_{2S,N}\mathbf{u}_N + \boldsymbol{\alpha}'_{2S,N}N^{1/2}\boldsymbol{\Delta}_N + o_p(1) \end{bmatrix}, \quad \text{(C.41)}$$

where

$$\boldsymbol{\alpha}_{1s,N} = 2N^{-1}E(\mathbf{D}'_N\mathbf{C}_{1s,N}\mathbf{u}_N) \text{ and } \boldsymbol{\alpha}_{2s,N} = 2N^{-1}E(\mathbf{D}'_N\mathbf{C}_{2s,N}\mathbf{u}_N)\,. \quad \text{(C.42)}$$

Furthermore, Lemma C.1 implies that the elements of $\boldsymbol{\alpha}_{1s,N}$ and $\boldsymbol{\alpha}_{2s,N}$ are uniformly bounded in absolute value for $s = 1,...,S$.

Using the definition of $\mathbf{C}_{1s,N}$ and $\mathbf{C}_{2s,N}$ and utilizing $\mathbf{u}_N = (\mathbf{I}_N - \sum_{m=1}^{M}\rho_{m,N}\mathbf{M}_{m,N})^{-1}\boldsymbol{\varepsilon}_N$, as well as Assumption 7, we have

$$N^{1/2}\mathbf{q}_N(\boldsymbol{\rho}_N,\boldsymbol{\Delta}_N) = N^{-1/2}\begin{bmatrix} \frac{1}{2}\boldsymbol{\varepsilon}'_N(\mathbf{A}_{11,N}+\mathbf{A}'_{11,N})\boldsymbol{\varepsilon}_N + \mathbf{a}'_{11,N}\boldsymbol{\varepsilon}_N \\ \frac{1}{2}\boldsymbol{\varepsilon}'_N(\mathbf{A}_{21,N}+\mathbf{A}'_{21,N})\boldsymbol{\varepsilon}_N + \mathbf{a}'_{21,N}\boldsymbol{\varepsilon}_N \\ \cdot \\ \cdot \\ \frac{1}{2}\boldsymbol{\varepsilon}'_N(\mathbf{A}_{1S,N}+\mathbf{A}'_{1S,N})\boldsymbol{\varepsilon}_N + \mathbf{a}'_{1S,N}\boldsymbol{\varepsilon}_N \\ \frac{1}{2}\boldsymbol{\varepsilon}'_N(\mathbf{A}_{2S,N}+\mathbf{A}'_{2S,N})\boldsymbol{\varepsilon}_N + \mathbf{a}'_{2S,N}\boldsymbol{\varepsilon}_N \end{bmatrix} + o_p(1) \qquad \text{(C.43)}$$

$$= N^{-1/2}\mathbf{v}^*_N + o_p(1) = \mathbf{v}_N + o_p(1),$$

where

$$\mathbf{a}_{1s,N} = \mathbf{T}_N\boldsymbol{\alpha}_{1s,N} \text{ and } \mathbf{a}_{2s,N} = \mathbf{T}_N\boldsymbol{\alpha}_{2s,N}. \qquad \text{(C.44)}$$

Observe that the elements of $\mathbf{a}_{1s,N}$ and $\mathbf{a}_{2s,N}$ are uniformly bounded in absolute value for $s = 1,...,S$ (by Assumption 7 and Lemma C.1). We define the $2S\times 1$ vector

$$\mathbf{v}_N = \begin{bmatrix} \mathbf{v}_{1,N} \\ . \\ \mathbf{v}_{S,N} \end{bmatrix} = (\mathbf{v}_{s,N}),\ s = 1,...,S, \text{ where} \qquad \text{(C.45)}$$

$$\mathbf{v}_{s,N} = \begin{bmatrix} v_{1s,N} \\ v_{2s,N} \end{bmatrix} = \begin{bmatrix} N^{-1/2}[\frac{1}{2}\boldsymbol{\varepsilon}'_N(\mathbf{A}_{1s,N}+\mathbf{A}'_{1s,N})\boldsymbol{\varepsilon}_N + \mathbf{a}'_{1s,N}\boldsymbol{\varepsilon}_N] \\ N^{-1/2}[\frac{1}{2}\boldsymbol{\varepsilon}'_N(\mathbf{A}_{2s,N}+\mathbf{A}'_{2s,N})\boldsymbol{\varepsilon}_N + \mathbf{a}'_{2s,N}\boldsymbol{\varepsilon}_N] \end{bmatrix}.$$

The $2S\times 2S$ (limiting) variance-covariance matrix of $\mathbf{v}_N$ takes the following form:

$$\boldsymbol{\Psi}_N = E(\mathbf{v}_N\mathbf{v}'_N) = E\begin{bmatrix} \mathbf{v}_{1,N}\mathbf{v}'_{1,N} & \mathbf{v}_{1,N}\mathbf{v}'_{S,N} \\ & \\ \mathbf{v}_{S,N}\mathbf{v}'_{1,N} & \mathbf{v}_{S,N}\mathbf{v}'_{S,N} \end{bmatrix} = E(v_{p,N}v'_{q,N}),\ p,q = 1,...,S. \qquad \text{(C.46)}$$

The elements of $\boldsymbol{\Psi}_N$ are defined in main text after (14). To derive the asymptotic distribution of $\mathbf{v}_N$ we invoke the central limit theorem for vectors of linear quadratic forms given by Kelejian and Prucha (2008, Theorem A.1; see Appendix B). In light of Assumptions 1, 2 and 7 (and Lemma C.1), the innovations $\boldsymbol{\varepsilon}_N$, the matrices $(\mathbf{A}_{1s,N}+\mathbf{A}'_{1s,N})$ and $(\mathbf{A}_{2s,N}+\mathbf{A}'_{2s,N})$, and the vectors $\mathbf{a}_{1s,N}$ and $\mathbf{a}_{2s,N}$, $s = 1,...,S$, satisfy the assumptions of Theorem A.1 in Kelejian and Prucha (2008), such that

$$-\boldsymbol{\Sigma}_{\mathbf{v}_N^*}^{-1/2}\mathbf{v}_n^* = -N^{-1/2}\boldsymbol{\Psi}_N^{-1/2}\mathbf{v}_N^* = -\boldsymbol{\Psi}_N^{-1/2}\mathbf{v}_N \xrightarrow{d}(\mathbf{0},\mathbf{I}_{2S}), \tag{C.47}$$

since $N^{-1}\lambda_{\min}(\boldsymbol{\Sigma}_{v_N^*}) = N^{-1}\lambda_{\min}(N\boldsymbol{\Psi}_N) = \lambda_{\min}(\boldsymbol{\Psi}_N) > 0$ as required in Theorem A.1.

Since the row and column sums of the matrices $\mathbf{A}_{1s,N}$ and $\mathbf{A}_{2s,N}$ are uniformly bounded in absolute value, and the elements of the vectors $\mathbf{a}_{1s,N}$ and $\mathbf{a}_{2s,N}$ and the variances are uniformly bounded in absolute value, it follows in light of (15a) that the elements of $\boldsymbol{\Psi}_N$ and also those of $\boldsymbol{\Psi}_N^{1/2}$ are uniformly bounded in absolute value.

It now follows from (C.38) and (C.39) and (C.43) that

$$N^{1/2}(\tilde{\boldsymbol{\rho}}_N - \boldsymbol{\rho}_N) = \boldsymbol{\Xi}_N^{-1}\mathbf{J}_N'\boldsymbol{\Theta}_N\boldsymbol{\Psi}_N^{1/2}(-\boldsymbol{\Psi}_N^{-1/2}\mathbf{v}_N) + o_p(1). \tag{C.48}$$

Since all nonstochastic terms on the right hand side from (C.48) are $O(1)$ it follows that $N^{1/2}(\tilde{\boldsymbol{\rho}}_N - \boldsymbol{\rho}_N)$ is $O_p(1)$. To derive the asymptotic distribution of $N^{1/2}(\tilde{\boldsymbol{\rho}}_N - \boldsymbol{\rho}_N)$, we invoke (part of) Corollary F4 (together with the Assumptions stated in Corollary F3) in Pötscher and Prucha (1997) (see Appendix B). In the present context we have

$\boldsymbol{\zeta}_N = -\boldsymbol{\Psi}_N^{-1/2}\mathbf{v}_N \xrightarrow{d}\boldsymbol{\zeta} \sim N(\mathbf{0},\mathbf{I}_{2S})$, and
$N^{1/2}(\tilde{\boldsymbol{\rho}}_N - \boldsymbol{\rho}_N) = \mathbf{A}_N\boldsymbol{\zeta}_N + o_p(1)$, where
$\mathbf{A}_N = \boldsymbol{\Xi}_N^{-1}\mathbf{J}_N'\boldsymbol{\Theta}_N\boldsymbol{\Psi}_N^{1/2}$.

As a final point we demonstrate that $\liminf_{N\to\infty}\lambda_{\min}(\mathbf{A}_N\mathbf{A}_N') > 0$ as required in Corollary F4 in Pötscher and Prucha (1997). Observe that

$$\begin{aligned}\lambda_{\min}(\mathbf{A}_N\mathbf{A}_N') &= \lambda_{\min}(\boldsymbol{\Xi}_N^{-1}\mathbf{J}_N'\boldsymbol{\Theta}_N\boldsymbol{\Psi}_N\boldsymbol{\Theta}_N'\mathbf{J}_N\boldsymbol{\Xi}_N^{-1}) \\ &\geq \lambda_{\min}(\boldsymbol{\Psi}_N)\lambda_{\min}(\boldsymbol{\Theta}_N\boldsymbol{\Theta}_N')\lambda_{\min}(\boldsymbol{\Xi}_N^{-1}\boldsymbol{\Xi}_N^{-1})\lambda_{\min}(\boldsymbol{\Gamma}_N'\boldsymbol{\Gamma}_N)\lambda_{\min}(\mathbf{B}_N'\mathbf{B}_N) > 0,\end{aligned} \tag{C.49}$$

since the matrices involved are all positive definite.

Hence, the expectation of $N^{1/2}(\tilde{\boldsymbol{\rho}}_N - \boldsymbol{\rho}_N)$ is zero and its limiting variance-covariance matrix is given by

$$\boldsymbol{\Omega}_{\tilde{\boldsymbol{\rho}}_N}(\boldsymbol{\Theta}_N) = (\mathbf{J}_N'\boldsymbol{\Theta}_N\mathbf{J}_N)^{-1}\mathbf{J}_N'\boldsymbol{\Theta}_N\boldsymbol{\Psi}_N\boldsymbol{\Theta}_N\mathbf{J}_N(\mathbf{J}_N'\boldsymbol{\Theta}_N\mathbf{J}_N)^{-1}, \tag{C.50}$$

where $\boldsymbol{\Omega}_{\tilde{\boldsymbol{\rho}}_N}$ is positive definite.

III. Proof of Theorem 3 (Variance-Covariance Estimation)

As part of proving Theorem 3 it has to be shown that $\widetilde{\boldsymbol{\Psi}}_N - \boldsymbol{\Psi}_N = o_p(1)$. Observe that in light of (15), $\widetilde{\boldsymbol{\Psi}}_N$ and $\boldsymbol{\Psi}_N$ are made of up S^2 blocks of dimension 2×2, $\widetilde{\psi}_{pq,N}$ and $\psi_{pq,N}$, whose elements can be written as $\widetilde{\psi}^{11}_{pq,N} = \widetilde{\psi}^{11,*}_{pq,N} + \widetilde{\psi}^{11,**}_{pq,N}$ and $\psi^{11}_{pq,N} = \psi^{11,*}_{pq,N} + \psi^{11,**}_{pq,N}$, where

$$\widetilde{\psi}^{11,*}_{pq,N} = \frac{1}{2} N^{-1} \sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij,1p,1q,N} \widetilde{\varepsilon}^2_{i,N} \widetilde{\varepsilon}^2_{j,N} \text{ , where } a_{ij,1p,1q,N} = (a_{ij,1p,N} + a_{ji,1p,N})(a_{ij,1q,N} + a_{ji,1q,N}), \quad \text{(C.51)}$$

$$\widetilde{\psi}^{11,**}_{pq,N} = N^{-1} \widetilde{\boldsymbol{\alpha}}'_{1p,N} \widetilde{\mathbf{P}}'_N \widetilde{\mathbf{F}}'_N \widetilde{\boldsymbol{\Sigma}}_N \widetilde{\mathbf{F}}_N \widetilde{\mathbf{P}}_N \widetilde{\boldsymbol{\alpha}}_{1q,N}, \quad \text{(C.52)}$$

$$\psi^{11,*}_{pq,N} = \frac{1}{2} N^{-1} \sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij,1p,1q,N} \sigma^2_{i,N} \sigma^2_{j,N} \text{ , where } a_{ij,1p,1q,N} = (a_{ij,1p,N} + a_{ji,1p,N})(a_{ij,1q,N} + a_{ji,1q,N}), \quad \text{(C.53)}$$

$$\psi^{11,**}_{pq,N} = = N^{-1} \boldsymbol{\alpha}'_{1p,N} \mathbf{P}'_N \mathbf{F}'_N \boldsymbol{\Sigma}_N \mathbf{F}_N \mathbf{P}_N \boldsymbol{\alpha}_{1q,N} \quad \text{(C.54)}$$

Analogous definitions apply to the other three elements $\psi^{12}_{pq,N}, \psi^{21}_{pq,N}$, and $\psi^{22}_{pq,N}$. In the subsequent proof, we consider element $\psi^{11}_{pq,N}$, but it is readily observed that the same reasoning applies to $\psi^{12}_{pq,N}, \psi^{21}_{pq,N}$, and $\psi^{22}_{pq,N}$ as well.

Two Lemmata (C.3 and C.4) will be used to show that $\widetilde{\psi}^{11,*}_{pq,N} - \psi^{11,*}_{pq,N} = o_p(1)$. Two Lemmata (C.5 and C.6) will be used to show that $\widetilde{\psi}^{11,**}_{pq,N} - \psi^{11,**}_{pq,N} = o_p(1)$.

Lemma C.3[21]

Suppose Assumptions 1-3 hold. Let $\Lambda_N = N^{-1} \boldsymbol{\sigma}^{2\prime}_N \mathbf{A}_N \boldsymbol{\sigma}^2_N$ and $\overline{\Lambda}_N = N^{-1} \boldsymbol{\varepsilon}^{2\prime}_N \mathbf{A}_N \boldsymbol{\varepsilon}^2_N$ with $\boldsymbol{\sigma}^2_N = (\sigma^2_{1,N}, ..., \sigma^2_{N,N})'$ and $\boldsymbol{\varepsilon}^2_N = (\varepsilon^2_{1,N}, ..., \varepsilon^2_{N,N})'$ and where the $N \times N$ matrices $\mathbf{A}_N$ are real, nonstochastic, and symmetric. Suppose further that the diagonal elements of the matrices $\mathbf{A}_N$ are zero and that their row and column sums are uniformly bounded in absolute value. Then $E\overline{\Lambda}_N = \Lambda_N = O(1)$ and $Var(\overline{\Lambda}_N) = o(1)$, and hence $\overline{\Lambda}_N - \Lambda_N \xrightarrow{p} 0$ as $N \to \infty$, and $\overline{\Lambda}_N = O_p(1)$.

Lemma C.3 as used here is exactly equal to Lemma C.3 in Kelejian and Prucha (2008, p. 40) in the first-order case, where the reader is referred to for a proof.

[21] Compare Kelejian and Prucha (2008, p. 40ff.).

Lemma C.4[22]

Suppose Assumptions 1-4 hold. Let $\boldsymbol{\varepsilon}_N = (\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N}\mathbf{M}_{m,N})\mathbf{u}_N$ and let $\widetilde{\boldsymbol{\varepsilon}}_N = (\mathbf{I}_N - \sum_{m=1}^{S} \widetilde{\rho}_m\mathbf{M}_{m,N})\widetilde{\mathbf{u}}_N$ with $\widetilde{\mathbf{u}}_N = \mathbf{u}_N + \mathbf{D}_N\boldsymbol{\Delta}_N$ and $\mathbf{D}_N = (\mathbf{d}'_{1.,N},...,\mathbf{d}'_{N.,N})'$, and where the $S \times 1$ vector $\widetilde{\boldsymbol{\rho}}_N$ can be any estimator that satisfies $N^{1/2}(\widetilde{\boldsymbol{\rho}}_N - \boldsymbol{\rho}_N) = O_p(1)$. Define $\widetilde{\Lambda}_N = N^{-1}(\widetilde{\boldsymbol{\varepsilon}}_N^2)'\mathbf{A}_N(\widetilde{\boldsymbol{\varepsilon}}_N^2)$, $\overline{\Lambda}_N = N^{-1}(\boldsymbol{\varepsilon}_N^2)'\mathbf{A}_N(\boldsymbol{\varepsilon}_N^2)$ with $\widetilde{\boldsymbol{\varepsilon}}_N^2 = (\widetilde{\varepsilon}_{1,N}^2,...,\widetilde{\varepsilon}_{N,N}^2)'$ and $\boldsymbol{\varepsilon}_N^2 = (\varepsilon_{1,N}^2,...,\varepsilon_{N,N}^2)'$, and where the $N \times N$ matrices $\mathbf{A}_N$ are real, nonstochastic, and symmetric. Suppose further that the diagonal elements of the matrices $\mathbf{A}_N$ are zero and that their row and column sums are uniformly bounded in absolute value, and that $Ed_{ij,N}^4 \leq K_d < \infty$. Then, $\widetilde{\Lambda}_N - \overline{\Lambda}_N \xrightarrow{p} 0$ as $N \to \infty$, and $\widetilde{\boldsymbol{\Lambda}}_N = O_p(1)$.

Proof.
Observe that

$$\widetilde{\Lambda}_N - \overline{\Lambda}_N = N^{-1}[\sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij,N}(\widetilde{\varepsilon}_{i,N}^2\widetilde{\varepsilon}_{j,N}^2 - \varepsilon_{i,N}^2\varepsilon_{j,N}^2)], \tag{C.55}$$

which can be written as $\widetilde{\boldsymbol{\Lambda}}_N - \overline{\boldsymbol{\Lambda}}_N = \varphi_{1,N} + \varphi_{2,N} + \varphi_{3,N}$, where

$$\varphi_{1,N} = N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij,N}(\widetilde{\varepsilon}_{i,N}^2 - \varepsilon_{i,N}^2)\varepsilon_{j,N}^2, \tag{C.56}$$

$$\varphi_{2,N} = N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij,N}\varepsilon_{i,N}^2(\widetilde{\varepsilon}_{j,N}^2 - \varepsilon_{j,N}^2), \tag{C.57}$$

$$\varphi_{3,N} = N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij,N}(\widetilde{\varepsilon}_{j,N}^2 - \varepsilon_{j,N}^2)(\widetilde{\varepsilon}_{i,N}^2 - \varepsilon_{i,N}^2). \tag{C.58}$$

We next show that $\varphi_{1,N}$, $\varphi_{2,N}$, and $\varphi_{3,N}$ are all $o_p(1)$. Observe that

$$\widetilde{\boldsymbol{\varepsilon}}_N = (\mathbf{I}_N - \sum_{m=1}^{S} \widetilde{\rho}_{m,N}\mathbf{M}_{m,N})\widetilde{\mathbf{u}}_N = (\mathbf{I}_N - \sum_{m=1}^{S} \widetilde{\rho}_{m,N}\mathbf{M}_{m,N})(\mathbf{u}_N + \mathbf{D}_N\boldsymbol{\Delta}_N) \tag{C.59}$$

$$= [\mathbf{I}_N - \sum_{m=1}^{S}(\rho_{m,N} + \widetilde{\rho}_{m,N} - \rho_{m,N})\mathbf{M}_{m,N}](\mathbf{u}_N + \mathbf{D}_N\boldsymbol{\Delta}_N)$$

$$= (\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N}\mathbf{M}_{m,N})\mathbf{u}_N - [\sum_{m=1}^{S}(\widetilde{\rho}_{m,N} - \rho_{m,N})\mathbf{M}_{m,N}]\mathbf{u}_N$$

[22] Compare Kelejian and Prucha (2008, p. 41ff.).

$$+(\mathbf{I}_N-\sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N})\mathbf{D}_N\boldsymbol{\Delta}_N-[\sum_{m=1}^{S}(\tilde{\rho}_{m,N}-\rho_{m,N})\mathbf{M}_{m,N}]\mathbf{D}_N\boldsymbol{\Delta}_N$$
$$=\boldsymbol{\varepsilon}_N+\boldsymbol{\eta}_N,$$

where

$$\boldsymbol{\eta}_N=(\mathbf{I}_N-\sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N})\mathbf{D}_N\boldsymbol{\Delta}_N+[\sum_{m=1}^{S}(\rho_{m,N}-\tilde{\rho}_{m,N})\mathbf{M}_{m,N}(\mathbf{I}_N-\sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N})^{-1}\boldsymbol{\varepsilon}_N]$$
$$+[\sum_{m=1}^{S}(\rho_{m,N}-\tilde{\rho}_{m,N})\mathbf{M}_{m,N}]\mathbf{D}_N\boldsymbol{\Delta}_N. \qquad \text{(C.60)}$$

This can also be written as

$$\boldsymbol{\eta}_N=\mathbf{R}_N\mathbf{g}_N, \qquad \text{(C.61)}$$

where

$\mathbf{R}_N=[\mathbf{R}_{1,N},\mathbf{R}_{2,N},\mathbf{R}_{3,N}]$ with

$\mathbf{R}_{1,N}=(\mathbf{I}_N-\sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N})\mathbf{D}_N,$

$\mathbf{R}_{2,N}=[\mathbf{M}_{1,N}(\mathbf{I}_N-\sum_{m=1}^{S}\rho_m\mathbf{M}_{m,N})^{-1}\boldsymbol{\varepsilon}_N,...,\mathbf{M}_{S,N}(\mathbf{I}_N-\sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N})^{-1}\boldsymbol{\varepsilon}_N],$

$\mathbf{R}_{3,N}=[\mathbf{M}_{1,N}\mathbf{D}_N,...,\mathbf{M}_{S,N}\mathbf{D}_N]$, and

$$\mathbf{g}_N=\begin{bmatrix}\boldsymbol{\Delta}_N\\(\boldsymbol{\rho}_N-\tilde{\boldsymbol{\rho}}_N)\\(\boldsymbol{\rho}_N-\tilde{\boldsymbol{\rho}}_N)\otimes\boldsymbol{\Delta}_N\end{bmatrix}.$$

In light of Assumption 3 and since the elements of $\mathbf{D}_N=(\mathbf{d}'_{1.,N},...,\mathbf{d}'_{N.,N})'$ and $\boldsymbol{\varepsilon}_N$ have bounded fourth moments, each column of the matrix $\mathbf{R}_N$ is of the form $\underline{\boldsymbol{\pi}}_N+\underline{\boldsymbol{\Pi}}_N\boldsymbol{\xi}_N$, where the elements of the $N\times 1$ vector $\underline{\boldsymbol{\pi}}_N$ are uniformly bounded in absolute value by some finite constant, the row and column sums of the $N\times N$ matrix $\underline{\boldsymbol{\Pi}}_N$ are uniformly bounded in absolute value by some finite constant, and the fourth moments of the elements of $\boldsymbol{\xi}_N$ are bounded by some finite constant. It follows that the fourth moments of the elements of $\mathbf{R}_N$ are also bounded by some finite constant by Lemma C.2 in Kelejian and Prucha, 2008 (see also Remark A.1 in Appendix A). As a consequence,

$$\|\boldsymbol{\eta}_N\|\le\|\mathbf{R}_N\|\|\mathbf{g}_N\|, \qquad \text{(C.62)}$$

or for the *i*-th element of the $N \times 1$ vector $\boldsymbol{\eta}_N$:

$$\left|\eta_{i,N}\right| \le \left\|\mathbf{r}_{i.,N}\right\| \left\|\mathbf{g}_N\right\| = \left\|\mathbf{g}_N\right\| \left\|\mathbf{r}_{i.,N}\right\| = \alpha_N \beta_{i,N}, \tag{C.63}$$

where $\mathbf{r}_{i.,N}$ denotes the *i*-th row of $\mathbf{R}_N$, and $\alpha_N = \left\|\mathbf{g}_N\right\|$ and $\beta_{i,N} = \left\|\mathbf{r}_{i.,N}\right\|$ with $E(\beta_{i,N}^4) \le K_\beta < \infty$. Without loss of generality we can select K_β such that $E(\beta_{i,N}^\gamma) \le K_\beta$ for $\gamma \le 4$.

Note that $N^{1/2}\alpha_N = O_p(1)$. Given Assumption 2, we have $E\left|\varepsilon_{i,N}\right|^\gamma \le K_\varepsilon < \infty$, $\gamma \le 4+\eta$ (for some $\eta > 0$) and some K_ε. By the Assumption of Lemma C.4, the row and column sums of $\mathbf{A}_N$ are uniformly bounded in absolute value, i.e., $\sum_{i=1}^N \left|a_{ij,N}\right| \le K_a < \infty$. In the following let $K = \max(1, K_\beta, K_\varepsilon, K_a)$. In light of Remark C.1 in Kelejian and Prucha (2008) (see Appendix A) it follows that $\sum_{l=1}^N \left|a_{il,N}\right|^\gamma \le K^\gamma$.

From (C.59) and (C.63) we have:

$$\begin{aligned}
&\left|\widetilde{\varepsilon}_{i,N}^2 - \varepsilon_{i,N}^2\right| = \left|(\varepsilon_{i,N} + \eta_{i,N})^2 - \varepsilon_i^2\right| = \left|(\varepsilon_{i,N}^2 + 2\varepsilon_{i,N}\eta_{i,N} + \eta_{i,N}^2) - \varepsilon_i^2\right| \qquad \text{(C.64)}\\
&= \left|2\varepsilon_{i,N}\eta_{i,N} + \eta_{i,N}^2\right|\\
&\le \left|2\varepsilon_{i,N}\eta_{i,N}\right| + \left|\eta_{i,N}^2\right|\\
&= 2\left|\varepsilon_{i,N}\right|\left|\eta_{i,N}\right| + \left|\eta_{i,N}^2\right|\\
&\le 2\alpha_N\beta_{i,N}\left|\varepsilon_{i,N}\right| + \alpha_N^2\beta_{i,N}^2 \ .
\end{aligned}$$

Taking the norm of (C.56) we have

$$\begin{aligned}
&\left|\varphi_{1,N}\right| = \left|N^{-1}\sum_{i=1}^N\sum_{j=1}^N a_{ij,N}(\widetilde{\varepsilon}_{i,N}^2 - \varepsilon_{i,N}^2)\varepsilon_{j,N}^2\right| \qquad \text{(C.65)}\\
&\le N^{-1}\sum_{i=1}^N\sum_{j=1}^N \left|a_{ij,N}\right|\left|(\widetilde{\varepsilon}_{i,N}^2 - \varepsilon_{i,N}^2)\right|\left|\varepsilon_{j,N}^2\right|\\
&\le N^{-1}\sum_{i=1}^N\sum_{j=1}^N \left|a_{ij,N}\right|(2\alpha_N\beta_{i,N}\left|\varepsilon_{i,N}\right| + \alpha_N^2\beta_{i,N}^2)\left|\varepsilon_{j,N}^2\right|\\
&= N^{-1}\sum_{i=1}^N (2\alpha_N\beta_{i,N}\left|\varepsilon_{i,N}\right| + \alpha_N^2\beta_{i,N}^2)\sum_{j=1}^N \left|a_{ij,N}\right|\left|\varepsilon_{j,N}^2\right|\\
&= N^{-1}2\alpha_N\sum_{i=1}^N \beta_{i,N}\left|\varepsilon_{i,N}\right|\sum_{j=1}^N \left|a_{ij,N}\right|\left|\varepsilon_{j,N}^2\right| + N^{-1}\alpha_N^2\sum_{i=1}^N \beta_{i,N}^2\sum_{j=1}^N \left|a_{ij,N}\right|\left|\varepsilon_{j,N}^2\right|
\end{aligned}$$

$$\leq \delta_{1,N}^{*} + \delta_{1,N}^{**},$$

with

$$\delta_{1,N}^{*} = 2KN^{2/(4+\eta)-1/2}(N^{1/2}\alpha_N)\zeta_N\left(N^{-1}\sum_{i=1}^{N}\beta_{i,N}\left|\varepsilon_{i,N}\right|\right), \tag{C.66}$$

$$\delta_{1,N}^{**} = KN^{2/(4+\eta)-1}(N^{1/2}\alpha_N)\zeta_N\left(N^{-1}\sum_{i=1}^{N}\beta_{i,N}^{2}\right), \tag{C.67}$$

where $\zeta_N = \left(N^{-1}\sum_{j=1}^{N}\left|\varepsilon_{j,N}\right|^{4+\eta}\right)^{2/(4+\eta)}$.

Next note that $E\left|\varepsilon_{j,N}\right|^{4+\eta} \leq K$ and thus $N^{-1}\sum_{j=1}^{N}\left|\varepsilon_{j,N}\right|^{4+\eta} = O_p(1)$ and also $\zeta_N = O_p(1)$. Moreover, $N^{1/2}\alpha_N = O_p(1)$.

Since $E(\beta_{i,N}\left|\varepsilon_{i,N}\right|) = \left(E\left|\beta_{i,N}\right|^2\right)^{1/2}\left(E\left|\varepsilon_{i,N}\right|^2\right)^{1/2} \leq K$ it follows that $N^{-1}\sum_{i=1}^{N}\beta_{i,N}\left|\varepsilon_{i,N}\right| = O_p(1)$. Since $E(\beta_{i,N}^2) \leq K$ it follows that $N^{-1}\sum_{i=1}^{N}\beta_{i,N} = O_p(1)$. Finally, $N^{2/(4+\eta)-1/2} = o(1)$ and $N^{2/(4+\eta)-1} = o(1)$. It follows that $\delta_{1,N}^{*} = \delta_{1,N}^{**} = o_p(1)$ and thus $\varphi_{1,N} = o_p(1)$.

Because of symmetry $\varphi_{1,N} = o_p(1)$ implies that $\varphi_{2,N} = o_p(1)$.

Now consider $\varphi_{3,N}$:

$$\begin{aligned}\left|\varphi_{3,N}\right| &= \left|N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}a_{ij,N}(\tilde{\varepsilon}_{j,N}^2 - \varepsilon_{j,N}^2)(\tilde{\varepsilon}_{i,N}^2 - \varepsilon_{i,N}^2)\right| \\ &= 4\alpha_N^2 N^{-1}\sum_{i=1}^{N}\beta_{i,N}\left|\varepsilon_{i,N}\right|\sum_{j=1}^{N}\left|a_{ij,N}\right|\beta_{j,N}\left|\varepsilon_{j,N}\right| \\ &+ 2\alpha_N^3 N^{-1}\sum_{i=1}^{N}\beta_{i,N}^2\sum_{j=1}^{N}\left|a_{ij,N}\right|\beta_{j,N}\left|\varepsilon_{j,N}\right| + 2\alpha_N^3 N^{-1}\sum_{i=1}^{N}\beta_{i,N}\left|\varepsilon_{i,N}\right|\sum_{j=1}^{N}\left|a_{ij,N}\right|\beta_{j,N}^2 \\ &+ \alpha_N^4 N^{-1}\sum_{i=1}^{N}\beta_{i,N}^2\sum_{j=1}^{N}\left|a_{ij,N}\right|\beta_{j,N}^2 \leq \delta_{3,N}^{*} + \delta_{3,N}^{**} + \delta_{3,N}^{***},\end{aligned} \tag{C.68}$$

with

$$\delta_{3,N}^{*} = 4KN^{-1/2}(N^{1/2}\alpha_N)^2\bar{\zeta}_N\left(N^{-1}\sum_{i=1}^{N}\beta_{i,N}\left|\varepsilon_{i,N}\right|\right), \tag{C.69}$$

$$\delta_{3,N}^{**} = N^{-1}4K(N^{1/2}\alpha_N)^3\bar{\bar{\zeta}}_N(N^{-1}\sum_{i=1}^{N}\beta_{i,N}\left|\varepsilon_{i,N}\right|), \tag{C.70}$$

$$\delta_{3,N}^{***} = N^{-3/2}(N^{1/2}\alpha_N)^4K\bar{\bar{\zeta}}_N\left(N^{-1}\sum_{i=1}^{N}\beta_{i,N}^2\right), \tag{C.71}$$

where $\bar{\zeta}_N = \left(N^{-1}\sum_{j=1}^{N}\beta_{j,N}^2\left|\varepsilon_{j,N}\right|^2\right)^{1/2}$ and $\bar{\bar{\zeta}}_N = \left(N^{-1}\sum_{j=1}^{N}\beta_{j,N}^4\right)^{1/2}$.

Next note that $E(\beta_{j,N}^2\left|\varepsilon_{j,N}\right|^2) \le \left(E\beta_{j,N}^4\right)^{1/2}\left(E\left|\varepsilon_{j,N}\right|^4\right)^{1/2} \le K$ and thus $N^{-1}\sum_{j=1}^{N}\beta_{j,N}^2\left|\varepsilon_{j,N}\right|^2 = O_p(1)$ and also $\bar{\zeta}_N = O_p(1)$. Then $E(\beta_{j,N}^4) \le K$ and thus $\left(N^{-1}\sum_{j=1}^{N}\beta_{j,N}^4\right)^{1/2} = O_p(1)$ and also $\bar{\bar{\zeta}}_N = O_p(1)$.

Moreover, $E(\beta_{j,N}\left|\varepsilon_{j,N}\right|) \le \left(E\beta_{j,N}^2\right)^{1/2}\left(E\left|\varepsilon_{j,N}\right|^2\right)^{1/2} \le K$ and thus $N^{-1}\sum_{j=1}^{N}\beta_{j,N}\left|\varepsilon_{j,N}\right| = O_p(1)$. We also have that $E(\beta_{j,N}^2) \le K$ and thus $N^{-1}\sum_{j=1}^{N}\beta_{j,N}^2 = O_p(1)$ and $N^{1/2}\alpha_N = O_p(1)$.

Since $N^{-\delta} = o(1)$ for any $\delta > 0$ it follows that $\delta_{3,N}^{*} = \delta_{3,N}^{**} = \delta_{3,N}^{***} = o_p(1)$ and thus $\varphi_{3,N} = o_p(1)$. Taking stock, we have proofed that $\varphi_{1,N} = \varphi_{2,N} = \varphi_{3,N} = o_p(1)$.

Lemma C.5[23]

Suppose Assumptions 1-4 hold. Let $\boldsymbol{\varepsilon}_N = (\mathbf{I}_N - \sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N})\mathbf{u}_N$, and let $\tilde{\boldsymbol{\varepsilon}}_N = (\mathbf{I}_N - \sum_{m=1}^{S}\tilde{\rho}_{m,N}\mathbf{M}_{m,N})\tilde{\mathbf{u}}_N$ with $\tilde{\mathbf{u}}_N = \mathbf{u}_N\mathbf{D}_N\boldsymbol{\Delta}_N$ and $\mathbf{D}_N = (\mathbf{d}_{1.,N}',...,\mathbf{d}_{N.,N}')'$, and where the $S\times 1$ vector $\tilde{\boldsymbol{\rho}}_N$ can be any estimator that satisfies $(\tilde{\boldsymbol{\rho}}_N - \boldsymbol{\rho}_N) = o_p(1)$. Let $\mathbf{a}_N$ and $\mathbf{b}_N$ be $N\times 1$ vectors, whose elements are uniformly bounded in absolute value by $c < \infty$ and let $\boldsymbol{\Sigma}_N = diag_{i=1,...,N}(\sigma_i^2)$ and $\tilde{\boldsymbol{\Sigma}}_N = diag_{i=1,...,N}(\tilde{\varepsilon}_i^2)$. Then:

(a) $N^{-1}\mathbf{a}_N'\tilde{\boldsymbol{\Sigma}}_N\mathbf{b}_N - N^{-1}\mathbf{a}_N'\boldsymbol{\Sigma}_N\mathbf{b}_N = o_p(1)$ and $N^{-1}\mathbf{a}_N'\boldsymbol{\Sigma}_N\mathbf{b}_N = O(1)$.

[23] Compare Kelejian and Prucha (2008, p. 43ff.)

(b) There exist random variables ς_N that do not depend on $\mathbf{a}_N$ and $\mathbf{b}_N$ such that $\left|N^{-1}\mathbf{a}_N'\widetilde{\mathbf{\Sigma}}_N\mathbf{b}_N - N^{-1}\mathbf{a}_N'\mathbf{\Sigma}_N\mathbf{b}_N\right| \leq K(c)(1+\varsigma_N)$

with $\varsigma_N = o_p(1)$ and where $K(c) < \infty$ is a constant that depends monotonically on c (as well as on some other bounds maintained in the assumptions).

Proof.

Let $\widetilde{\tau}_N = N^{-1}\mathbf{a}_N'\widetilde{\mathbf{\Sigma}}_N\mathbf{b}_N$, $\overline{\tau}_N = N^{-1}\mathbf{a}_N'\overline{\mathbf{\Sigma}}_N\mathbf{b}_N$, and $\tau_N = N^{-1}\mathbf{a}_N'\mathbf{\Sigma}_N\mathbf{b}_N$, where $\overline{\mathbf{\Sigma}}_N = diag_{i=1,\ldots,N}(\varepsilon_{i,N}^2)$. It follows from the triangle inequality that

$$\left|\widetilde{\tau}_N - \tau_N\right| \leq \left|\overline{\tau}_N - \tau_N\right| + \left|\widetilde{\tau}_N - \overline{\tau}_N\right|. \tag{C.72}$$

By the weak law of large numbers for i.d. variables (see, e.g., White, 2001, p. 35), observing that the fourth moments of ε_i are uniformly bounded by Assumption 2, we have

$$\overline{\tau}_N - \tau_N = N^{-1}(\mathbf{a}_N'\overline{\mathbf{\Sigma}}_N\mathbf{b}_N - \mathbf{a}_N'\mathbf{\Sigma}_N\mathbf{b}_N) = N^{-1}\mathbf{a}_N'(\overline{\mathbf{\Sigma}}_N - \mathbf{\Sigma}_N)\mathbf{b}_N \tag{C.73}$$
$$= N^{-1}\sum_{i=1}^{N} a_{i,N}(\varepsilon_{i,N}^2 - \sigma_{i,N}^2)b_{i,N} = o_p(1).$$

We thus also have $\left|\overline{\tau}_N - \tau_N\right| = o_p(1)$. Next let

$$\zeta_{1,N} = N^{-1}\sum_{i=1}^{N}\left|\,(\varepsilon_{i,N}^2 - \sigma_{i,N}^2) - E\left|\varepsilon_{i,N}^2 - \sigma_{i,N}^2\right|\,\right| \tag{C.74}$$

and c_* be such that $\sigma_{i,N}^2 \leq c_*$.

Note that $\zeta_{1,N} = N^{-1}\sum_{i=1}^{N}(\left|\varepsilon_{i,N}^2 - \sigma_{i,N}^2\right| - E\left|\varepsilon_{i,N}^2 - \sigma_{i,N}^2\right|) = o_p(1)$. It then follows that

$$\left|\overline{\tau}_N - \tau_N\right| = N^{-1}\left|\sum_{i=1}^{N} a_{i,N}(\varepsilon_{i,N}^2 - \sigma_{i,N}^2)b_{i,N}\right| \tag{C.75}$$
$$= N^{-1}\left|\sum_{i=1}^{N}[a_{i,N}b_{i,N}(\varepsilon_{i,N}^2 - \sigma_{i,N}^2) - E\left|\varepsilon_{i,N}^2 - \sigma_{i,N}^2\right| + E\left|\varepsilon_{i,N}^2 - \sigma_{i,N}^2\right|]\right|$$
$$\leq c^2\zeta_{1,N} + c^2N^{-1}\left|\sum_{i=1}^{N}2c_*\right| = c^2\zeta_{1,N} + c^2N^{-1}\sum_{i=1}^{N}2c_* = c^2\zeta_{1,N} + 2c^2c_*.$$

Next rewrite

$$\widetilde{\tau}_N = N^{-1}\mathbf{a}_N'\widetilde{\boldsymbol{\Sigma}}_N\mathbf{b}_N = N^{-1}\sum_{i=1}^{N} a_{i,N}\widetilde{\varepsilon}_{i,N}^2 b_{i,N} = N^{-1}\sum_{i=1}^{N}\widetilde{\varepsilon}_{i,N}a_{i,N}b_{i,N}\widetilde{\varepsilon}_{i,N} \tag{C.76}$$
$$= N^{-1}\widetilde{\boldsymbol{\varepsilon}}_N'\mathbf{C}_N\widetilde{\boldsymbol{\varepsilon}}_N,$$

where $\mathbf{C}_N = diag_{i=1,...,N}(a_{i,N}b_{i,N})$ and $\bar{\tau}_N = N^{-1}\boldsymbol{\varepsilon}_N'\mathbf{C}_N\boldsymbol{\varepsilon}_N$.

Using $\boldsymbol{\varepsilon}_N = (\mathbf{I}_N - \sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N})\mathbf{u}_N$, we have

$$\widetilde{\tau}_N = N^{-1}\widetilde{\boldsymbol{\varepsilon}}_N'\mathbf{C}_N\widetilde{\boldsymbol{\varepsilon}}_N = N^{-1}\widetilde{\mathbf{u}}_N'(\mathbf{I}_N - \sum_{m=1}^{S}\widetilde{\rho}_{m,N}\mathbf{M}_{m,N}')\mathbf{C}_N(\mathbf{I}_N - \sum_{m=1}^{S}\widetilde{\rho}_{m,N}\mathbf{M}_{m,N})\widetilde{\mathbf{u}}_N \tag{C.77}$$
$$= N^{-1}\widetilde{\mathbf{u}}_N'\mathbf{C}_N\widetilde{\mathbf{u}}_N - 2\widetilde{\mathbf{u}}_N'\left(\sum_{m=1}^{S}\widetilde{\rho}_{m,N}\mathbf{M}_{m,N}'\mathbf{C}_N\right)\widetilde{\mathbf{u}}_N + \widetilde{\mathbf{u}}_N'\left(\sum_{m=1}^{S}\sum_{l=1}^{S}\widetilde{\rho}_{m,N}\widetilde{\rho}_{l,N}\mathbf{M}_{m,N}'\mathbf{C}_N\mathbf{M}_{l,N}\right)\widetilde{\mathbf{u}}_N,$$

and

$$\bar{\tau}_N = N^{-1}\boldsymbol{\varepsilon}_N'\mathbf{C}_N\boldsymbol{\varepsilon}_N \tag{C.78}$$
$$= N^{-1}\mathbf{u}_N'\mathbf{C}_N\mathbf{u}_N - 2\mathbf{u}_N'\left(\sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N}'\mathbf{C}_N\right)\mathbf{u}_N + \mathbf{u}_N'\left(\sum_{m=1}^{S}\sum_{l=1}^{S}\rho_{m,N}\rho_{l,N}\mathbf{M}_{m,N}'\mathbf{C}_N\mathbf{M}_{l,N}\right)\mathbf{u}_N.$$

By the properties of the matrices $\mathbf{C}_N$ and $\mathbf{M}_{m,N}$, $m = 1,...,S$ and in light of Remark A.1 these are all quadratic forms in matrices whose row and column sums are uniformly bounded in absolute value by some constants that depend monotonically on *c* as well as on other bounds maintained in the assumptions.

Using the triangle inequality, it follows that

$$\left|\widetilde{\tau}_N - \bar{\tau}_N\right| \le \left|N^{-1}\widetilde{\mathbf{u}}_N'\mathbf{C}_N\widetilde{\mathbf{u}}_N - N^{-1}\mathbf{u}_N'\mathbf{C}_N\mathbf{u}_N\right| \tag{C.79}$$
$$+\left|2N^{-1}(\widetilde{\rho}_{1,N} - \rho_{1,N})(\widetilde{\mathbf{u}}_N'\mathbf{M}_{1,N}'\mathbf{C}_N\widetilde{\mathbf{u}}_N - \mathbf{u}_N'\mathbf{M}_{1,N}'\mathbf{C}_N\mathbf{u}_N)\right| + ...$$
$$+\left|N^{-1}(\widetilde{\rho}_{1,N}^2 - \rho_{1,N}^2)(\widetilde{\mathbf{u}}_N'\mathbf{M}_{1,N}'\mathbf{C}_N\mathbf{M}_{S,N}\widetilde{\mathbf{u}}_N - \mathbf{u}_N'\mathbf{M}_{1,N}'\mathbf{C}_N\mathbf{M}_{S,N}\mathbf{u}_N)\right| + ...$$
$$+\left|N^{-1}(\widetilde{\rho}_{1,N}\widetilde{\rho}_{2,N} - \rho_{1,N}\rho_{2,N})(\widetilde{\mathbf{u}}_N'\mathbf{M}_{1,N}'\mathbf{C}_N\mathbf{M}_{S,N}\widetilde{\mathbf{u}}_N - \mathbf{u}_N'\mathbf{M}_{1,N}'\mathbf{C}_N\mathbf{M}_{S,N}\mathbf{u}_N)\right| + ...$$
$$\le \left|N^{-1}\widetilde{\mathbf{u}}_N'\mathbf{C}_N\widetilde{\mathbf{u}}_N - N^{-1}\mathbf{u}_N'\mathbf{C}_N\mathbf{u}_N\right| +$$
$$+\left|\widetilde{\rho}_{1,N} - \rho_{1,N}\right|\left|2N^{-1}(\widetilde{\mathbf{u}}_N'\mathbf{M}_{1,N}'\mathbf{C}_N\widetilde{\mathbf{u}}_N - \mathbf{u}_N'\mathbf{M}_{1,N}'\mathbf{C}_N\mathbf{u}_N)\right| + ...$$
$$+\left|\widetilde{\rho}_{1,N}^2 - \rho_{1,N}^2\right|\left|N^{-1}(\widetilde{\mathbf{u}}_N'\mathbf{M}_{1,N}'\mathbf{C}_N\mathbf{M}_{S,N}\widetilde{\mathbf{u}}_N - \mathbf{u}_N'\mathbf{M}_{1,N}'\mathbf{C}_N\mathbf{M}_{S,N}\mathbf{u}_N)\right| + ...$$

$$+\left|\widetilde{\rho}_{1,N}\widetilde{\rho}_{2,N}-\rho_{1,N}\rho_{2,N}\right|\left|N^{-1}(\widetilde{\mathbf{u}}_N'\mathbf{M}_{1,N}'\mathbf{C}_N\mathbf{M}_{S,N}\widetilde{\mathbf{u}}_N-\mathbf{u}_N'\mathbf{M}_{1,N}'\mathbf{C}_N\mathbf{M}_{S,N}\mathbf{u}_N)\right|+\ldots$$

Then Remark C.2 in Kelejian and Prucha (2008) (see Appendix C in the present paper for the higher-order case) can be applied, observing that $(\widetilde{\rho}_{s,N}-\rho_{s,N})=o_p(1)$, $s=1,...,S$. Factoring out the $o_p(1)$ terms, we obtain

$$\left|\widetilde{\tau}_N-\bar{\tau}_N\right|\le k(c)\zeta_{2,N}, \tag{C.80}$$

where $\zeta_{2,N}=o_p(1)$ and does not depend on $\mathbf{a}_N$ and $\mathbf{b}_N$ and the constant $k(c)$ depends monotonically on c and other bounds maintained in the assumptions.

The first claim of part (a) of Lemma C.5 now follows from (C.73) and (C.80). The second claim follows from

$$\tau_N=N^{-1}\mathbf{a}_N'\boldsymbol{\Sigma}_N\mathbf{b}_N=N^{-1}\sum_{i=1}^{N}\left|a_{i,N}\right|\left|\sigma_{i,N}^2\right|\left|b_{i,N}\right|=N^{-1}\sum_{i=1}^{N}\left|a_{i,N}\right|\left|b_{i,N}\right|\left|\sigma_{i,N}^2\right|\le c^2c_*<\infty. \tag{C.81}$$

Part (b) of the Lemma follows from (C.72), (C.75), and (C.80), noting that $\zeta_{1,N}$, $\zeta_{2,N}$ are nonnegative:

$\left|\widetilde{\tau}_N-\tau_N\right|\le[c^2+2c^2c_*+k(c)]\varsigma_N$ where $\varsigma_N=(1+\zeta_N)$.

Lemma C.6[24]

Suppose Assumptions 1-4 hold. Furthermore, assume that $\sup_N\sum_{m=1}^{S}\left|\rho_{m,N}\right|<1$, and that the row and column sums of $\mathbf{M}_{m,N}$, $m=1,...,S$ are uniformly bounded in absolute value by 1 and some finite constant respectively. Let $\boldsymbol{\varepsilon}_N=(\mathbf{I}_N-\sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N})\mathbf{u}_N$, and let $\widetilde{\boldsymbol{\varepsilon}}_N=(\mathbf{I}_N-\sum_{m=1}^{S}\widetilde{\rho}_{m,N}\mathbf{M}_{m,N})\widetilde{\mathbf{u}}_N$ with $\widetilde{\mathbf{u}}_N=\mathbf{u}_N\mathbf{D}_N\boldsymbol{\Delta}_N$ and $\mathbf{D}_N=(\mathbf{d}_{1.,N}',...,\mathbf{d}_{N.,N}')'$. The $S\times1$ vector $\widetilde{\boldsymbol{\rho}}_N$ can be any estimator that satisfies $(\widetilde{\boldsymbol{\rho}}_N-\boldsymbol{\rho}_N)=o_p(1)$.

Let $\mathbf{F}_N=(\mathbf{I}_N-\sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N})$, $\widetilde{\mathbf{F}}_N=(\mathbf{I}_N-\sum_{m=1}^{S}\widetilde{\rho}_{m,N}\mathbf{M}_{m,N})\mathbf{H}_N$ or $\mathbf{F}_N=(\mathbf{I}_N-\sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N})^{-1}\mathbf{H}_N$, $\widetilde{\mathbf{F}}_N=(\mathbf{I}_N-\sum_{m=1}^{S}\widetilde{\rho}_{m,N}\mathbf{M}_{m,N})^{+}\mathbf{H}_N$, where $\mathbf{H}_N$ is an $N\times P_*$ matrix whose elements are uniformly

[24] Compare Kelejian and Prucha (2008, p. 45ff.)

bounded in absolute value by some constant $c < \infty$, let $\boldsymbol{\Sigma}_N = diag_{i=1,...,N}(\sigma_i^2)$ and $\widetilde{\boldsymbol{\Sigma}}_N = diag_{i=1,...,N}(\widetilde{\varepsilon}_i^2)$. Then, $N^{-1}\widetilde{\mathbf{F}}_N'\widetilde{\boldsymbol{\Sigma}}_N\widetilde{\mathbf{F}}_N - N^{-1}\mathbf{F}_N'\boldsymbol{\Sigma}_N\mathbf{F}_N = o_p(1)$ and $N^{-1}\mathbf{F}_N'\boldsymbol{\Sigma}_N\mathbf{F}_N = O(1)$.

Proof.

The subsequent proof will focus on the more general case, where $\mathbf{F}_N = (\mathbf{I}_N - \sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N})^{-1}\mathbf{H}_N$ and $\widetilde{\mathbf{F}}_N = (\mathbf{I}_N - \sum_{m=1}^{S}\widetilde{\rho}_{m,N}\mathbf{M}_{m,N})^{+}\mathbf{H}_N$; it is readily observed from the proof that this also covers the case where $\mathbf{F}_N = (\mathbf{I}_N - \sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N})\mathbf{H}_N$ and $\widetilde{\mathbf{F}}_N = (\mathbf{I}_N - \sum_{m=1}^{S}\widetilde{\rho}_{m,N}\mathbf{M}_{m,N})\mathbf{H}_N$.

Under the maintained assumptions there exists a ρ_* with $\sup\sum_{m=1}^{S}|\rho_{m,N}| < \rho_* < 1$. It follows immediately by the properties of the matrices $\mathbf{M}_{m,N}$ that the row and column sums of $\rho_*\mathbf{M}_{m,N}$, $m = 1,...,S$ are uniformly bounded in absolute value by 1 and some finite constant respectively. For later reference, also note that the elements of the vector $\rho_*^k\mathbf{M}_N^k\mathbf{h}_{.s,N}$ are also uniformly bounded in absolute value by *c*.

Denote the (*r*,*s*)-th element of the difference $N^{-1}\widetilde{\mathbf{F}}_N'\widetilde{\boldsymbol{\Sigma}}_N\widetilde{\mathbf{F}}_N - N^{-1}\mathbf{F}_N\boldsymbol{\Sigma}_N\mathbf{F}_N$ as ν_N, which is given by

$$\nu_N = N^{-1}(\widetilde{\mathbf{f}}_{.r,N}'\widetilde{\boldsymbol{\Sigma}}_N\widetilde{\mathbf{f}}_{.s,N} - \mathbf{f}_{.r,N}'\boldsymbol{\Sigma}_N\mathbf{f}_{.s,N}),\ r,s = 1,...,P, \tag{C.83}$$

which can be written as $\nu_N = \sum_{i=1}^{7}\nu_{i,N}$, where

$$\begin{aligned}
\nu_{1,N} &= N^{-1}(\widetilde{\mathbf{f}}_{.r,N} - \mathbf{f}_{.r,N})'(\widetilde{\boldsymbol{\Sigma}}_N - \boldsymbol{\Sigma}_N)(\widetilde{\mathbf{f}}_{.s,N} - \mathbf{f}_{.s,N}) \qquad \text{(C.84)}\\
\nu_{2,N} &= N^{-1}(\widetilde{\mathbf{f}}_{.r,N} - \mathbf{f}_{.r,N})'(\widetilde{\boldsymbol{\Sigma}}_N - \boldsymbol{\Sigma}_N)\mathbf{f}_{.s,N}\\
\nu_{3,N} &= N^{-1}\mathbf{f}_{.r,N}'(\widetilde{\boldsymbol{\Sigma}}_N - \boldsymbol{\Sigma}_N)(\widetilde{\mathbf{f}}_{.s,N} - \mathbf{f}_{.s,N})\\
\nu_{4,N} &= N^{-1}\mathbf{f}_{.r,N}'(\widetilde{\boldsymbol{\Sigma}}_N - \boldsymbol{\Sigma}_N)\mathbf{f}_{.s,N}\\
\nu_{5,N} &= N^{-1}(\widetilde{\mathbf{f}}_{.r,N} - \mathbf{f}_{.r,N})'\boldsymbol{\Sigma}_N(\widetilde{\mathbf{f}}_{.s,N} - \mathbf{f}_{.s,N})\\
\nu_{6,N} &= N^{-1}(\widetilde{\mathbf{f}}_{.r,N} - \mathbf{f}_{.r,N})'\boldsymbol{\Sigma}_N\mathbf{f}_{.s,N}\\
\nu_{7,N} &= N^{-1}\mathbf{f}_{.r,N}'\boldsymbol{\Sigma}_N(\widetilde{\mathbf{f}}_{.s,N} - \mathbf{f}_{.s,N}).
\end{aligned}$$

Next note that $\mathbf{f}_{.s,N} = (\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N}\mathbf{M}_{m,N})^{-1}\mathbf{h}_{.s,N}$ and thus

$$\widetilde{\mathbf{f}}_{.s,N} - \mathbf{f}_{.s,N} = [(\mathbf{I}_N - \sum_{m=1}^{S} \widetilde{\rho}_{m,N}\mathbf{M}_{m,N})^{+} - (\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N}\mathbf{M}_{m,N})^{-1}]\mathbf{h}_{.s,N} \tag{C.85}$$

We next demonstrate that $\widetilde{\mathbf{f}}_{.s,N} - \mathbf{f}_{.s,N} = o_p(1)$ by showing that each summand $\nu_{i,N} = o_p(1)$, $i = 1,...,7$.

To do so we invoke the following theorem (see, e.g., Resnik, 1999, p. 171): Let $(X, X_N, N \geq 1)$ be real valued random variables. Then, $X_N \xrightarrow{p} X$ if and only if each subsequence X_{Na} contains a further subsequence $X_{N'a}$ that converges almost surely to X.

As we show below we will be confronted with terms of the form:

$$\aleph_N^{(k,l)} = N'^{-1} p_*^{l+k} \mathbf{h}'_{.r,N'} \overline{\mathbf{M}}'^{l}_{N} \widetilde{\mathbf{\Sigma}}_N \mathbf{M}^{k}_{N} \mathbf{h}_{.s,N} - N^{-1} p_*^{l+k} \mathbf{h}'_{.r,N} \overline{\mathbf{M}}'^{l}_{N} \mathbf{\Sigma}_N \overline{\overline{\mathbf{M}}}^{k}_{N} \mathbf{h}_{.s,N}. \tag{C.86}$$

where $\overline{\mathbf{M}}^{l}_{N}$ is a matrix, whose row and column sums are uniformly bounded in absolute value by some constant $c_{\overline{M}}$. It follows that the absolute values of the elements of the vector $\overline{\mathbf{M}}^{k}_{N}\mathbf{h}_{.s,N}$ (and also that of $\rho_*^k \overline{\mathbf{M}}^{k}_{N}\mathbf{h}_{.s,N}$) are uniformly bounded in absolute value by some finite constant $c_* = \bar{c}c$ (and $c_{**} = \rho_*^k \bar{c}c$). (See Remark A.1 in Appendix A.)

Hence, Lemma C.5 applies and it follows that $\aleph_N^{(k,l)} = o_p(1)$ and that there exist random variables $\varsigma_N = o_p(1)$ such that $\left|\aleph_N^{(k,l)}\right| \leq K(c_*)(1 + \varsigma_N)$.

Now, let the index N_a denote some subsequence. In light of the aforementioned equivalence, there exists a subsequence of this subsequence ($N_{a'}$) such that for events $\omega \in A$, with $P(A^C) = 0$, it holds that

$$\left|\aleph_{N'_a}^{(k,l)}\right| \to 0,\ \left|\varsigma_{N'_a}\right| \to 0,\ \left|\widetilde{\rho}_{m,N'_a} - \rho_{m,N'_a}\right| \to 0,\ m = 1,...,S \tag{C.87}$$

and that for some $N'_a \geq \overline{N}_\omega$, $\left|\varsigma_{N'_a}(\omega)\right| \leq 1$, and thus

$$\left|\aleph_{N'_a}^{(k,l)}(\omega)\right| \leq K(c_*)(1 + \left|\varsigma_{N'_a}\right|) \leq 2K(c), \tag{C.88}$$

and finally

$$\sum_{m=1}^{S}\left|\widetilde{\rho}_{m,N_a'}(\omega)\right| \le p_{**}, \text{ where } p_{**} = \frac{\sup_N \sum_{m=1}^{S}\left|\rho_{m,N}\right| + p_*}{2} < 1. \tag{C.89}$$

In the following, assume that $N_a' \ge \overline{N}_\omega$. Since $\sum_{m=1}^{S}\left|\widetilde{\rho}_{m,N_a'}(\omega)\right| < 1$, it follows from Horn and Johnson (1985, p. 301) that $(\mathbf{I}_N - \sum_{m=1}^{S}\widetilde{\rho}_{m,N_a'}(\omega)\mathbf{M}_{m,N_a'})^+$ is invertible and that

$$\begin{aligned}\widetilde{\mathbf{f}}_{.s,N_a'} - \mathbf{f}_{.s,N_a'} &= [(\mathbf{I}_N - \sum_{m=1}^{S}\widetilde{\rho}_{m,N_a'}(\omega)\mathbf{M}_{m,N_a'})^{-1} - (\mathbf{I}_N - \sum_{m=1}^{S}\rho_{m,N_a'}\mathbf{M}_{m,N_a'})^{-1}]\mathbf{h}_{.s,N_a'} \\ &= \sum_{l=1}^{\infty}\left[\left(\sum_{m=1}^{S}\widetilde{\rho}_{m,N_a'}(\omega)\mathbf{M}_{m,N_a'}\right)^l - \left(\sum_{m=1}^{S}\rho_{m,N_a'}\mathbf{M}_{m,N_a'}\right)^l\right]\mathbf{h}_{.s,N_a'}.\end{aligned} \tag{C.90}$$

Substituting into the expression for $\nu_{1,N_a'}$ given by (C.84) yields

$$\begin{aligned}\nu_{1,N_a'} &= N_a'^{-1}(\widetilde{\mathbf{f}}_{.r,N_a'} - \mathbf{f}_{.r,N_a'})'(\widetilde{\boldsymbol{\Sigma}}_{N_a'} - \boldsymbol{\Sigma}_{N_a'})(\widetilde{\mathbf{f}}_{.s,N_a'} - \mathbf{f}_{.s,N_a'}) \\ &= N_a'^{-1}\mathbf{h}'_{.r,N_a'}\sum_{l=1}^{\infty}\left[\left(\sum_{m=1}^{S}\widetilde{\rho}_{m,N_a'}\mathbf{M}_{m,N_a'}\right)^l - \left(\sum_{m=1}^{S}\rho_{m,N_a'}\mathbf{M}_{m,N_a'}\right)^l\right]' \\ &\times(\widetilde{\boldsymbol{\Sigma}}_{N_a'} - \boldsymbol{\Sigma}_{N_a'})\sum_{k=1}^{\infty}\left[\left(\sum_{m=1}^{S}\widetilde{\rho}_{m,N_a'}\mathbf{M}_{m,N_a'}\right)^k - \left(\sum_{m=1}^{S}\rho_{m,N_a'}\mathbf{M}_{m,N_a'}\right)^k\right]\mathbf{h}_{.s,N_a'} \\ &= N_a'^{-1}\sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\mathbf{h}'_{.r,N_a'}\left[\left(\sum_{m=1}^{S}\widetilde{\rho}_{m,N_a'}\mathbf{M}_{m,N_a'}\right)^l - \left(\sum_{m=1}^{S}\rho_{m,N_a'}\mathbf{M}_{m,N_a'}\right)^l\right]'\widetilde{\boldsymbol{\Sigma}}_{N_a'}\left[\left(\sum_{m=1}^{S}\widetilde{\rho}_{m,N_a'}\mathbf{M}_{m,N_a'}\right)^k - \left(\sum_{m=1}^{S}\rho_{m,N_a'}\mathbf{M}_{m,N_a'}\right)^k\right]\mathbf{h}_{.s,N_a'} \\ &- N_a'^{-1}\mathbf{h}'_{.r,N_a'}\sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\left[\left(\sum_{m=1}^{S}\widetilde{\rho}_{m,N_a'}\mathbf{M}_{m,N_a'}\right)^l - \left(\sum_{m=1}^{S}\rho_{m,N_a'}\mathbf{M}_{m,N_a'}\right)^l\right]'\boldsymbol{\Sigma}_{N_a'}\left[\left(\sum_{m=1}^{S}\widetilde{\rho}_{m,N_a'}\mathbf{M}_{m,N_a'}\right)^k - \left(\sum_{m=1}^{S}\rho_{m,N_a'}\mathbf{M}_{m,N_a'}\right)^k\right]\mathbf{h}_{.s,N_a'}\end{aligned} \tag{C.91}$$

A single element with index (*k*,*l*) of this infinite double sum over *k* and *l* is given by

$$\begin{aligned}&N_a'^{-1}\mathbf{h}'_{.r,N_a'}\left[\left(\sum_{m=1}^{S}\widetilde{\rho}_{m,N_a'}\mathbf{M}_{m,N_a'}\right)^l - \left(\sum_{m=1}^{S}\rho_{m,N_a'}\mathbf{M}_{m,N_a'}\right)^l\right]'\widetilde{\boldsymbol{\Sigma}}_{N_a'}\left[\left(\sum_{m=1}^{S}\widetilde{\rho}_{m,N_a'}\mathbf{M}_{m,N_a'}\right)^k - \left(\sum_{m=1}^{S}\rho_{m,N_a'}\mathbf{M}_{m,N_a'}\right)^k\right]\mathbf{h}_{.s,N_a'} \\ &- N_a'^{-1}\mathbf{h}'_{.r,N_a'}\left[\left(\sum_{m=1}^{S}\widetilde{\rho}_{m,N_a'}\mathbf{M}_{m,N_a'}\right)^l - \left(\sum_{m=1}^{S}\rho_{m,N_a'}\mathbf{M}_{m,N_a'}\right)^l\right]'\boldsymbol{\Sigma}_{N_a'}\left[\left(\sum_{m=1}^{S}\widetilde{\rho}_{m,N_a'}\mathbf{M}_{m,N_a'}\right)^k - \left(\sum_{m=1}^{S}\rho_{m,N_a'}\mathbf{M}_{m,N_a'}\right)^k\right]\mathbf{h}_{.s,N_a'}.\end{aligned} \tag{C.92}$$

Next note that for any valued of $\boldsymbol{\rho}_{N_a'}$ and any $\widetilde{\boldsymbol{\rho}}_{N_a'}(\omega)$ there exist matrices $\widehat{\mathbf{M}}_{N_a'}$ and $\breve{\mathbf{M}}_{N_a'}$, whose row and column sums are uniformly bounded in absolute value, such that:

$$\sum_{m=1}^{S}\rho_{m,N_a'}\mathbf{M}_{m,N_a'}=\sum_{m=1}^{S}\rho_{m,N_a'}\overline{\mathbf{M}}_{N_a'} \text{ and } \sum_{m=1}^{S}\widetilde{\rho}_{m,N_a'}\mathbf{M}_{m,N_a'}=\sum_{m=1}^{S}\widetilde{\rho}_{m,N_a'}\overline{\overline{\mathbf{M}}}_{N_a'} . \tag{C.93}$$

$\overline{\mathbf{M}}_{N_a'}$ and $\overline{\overline{\mathbf{M}}}_{N_a'}$ can thus be factored out of the sum, yielding

$$\left[\left(\sum_{m=1}^{S}\widetilde{\rho}_{m,N_a'}(\omega)\right)^{l}\widehat{\mathbf{M}}_{N_a'}^{l}-\left(\sum_{m=1}^{S}\rho_{m,N_a'}\right)^{l}\breve{\mathbf{M}}_{N_a'}^{l}\right]. \tag{C.94}$$

By the same reasoning, for any values of $\left(\sum_{m=1}^{S}\widetilde{\rho}_{m,N_a'}(\omega)\right)$ and $\left(\sum_{m=1}^{S}\rho_{m,N_a'}\right)$, there exists a matrix $\overline{\mathbf{M}}_{N_a'}$, whose row and column sums are uniformly bounded in absolute value, such that:

$$\left[\left(\sum_{m=1}^{S}\widetilde{\rho}_{m,N_a'}(\omega)\right)^{l}\widehat{\mathbf{M}}_{N_a'}^{l}-\left(\sum_{m=1}^{S}\rho_{m,N_a'}\right)^{l}\breve{\mathbf{M}}_{N_a'}^{l}\right]=\left[\left(\sum_{m=1}^{S}\widetilde{\rho}_{m,N_a'}(\omega)\right)^{l}-\left(\sum_{m=1}^{S}\rho_{m,N_a'}\right)^{l}\right]\overline{\mathbf{M}}_{N_a'}^{l}. \tag{C.95}$$

Substituting $\overline{\mathbf{M}}_{N_a'}$ into the expression for $\nu_{1,N_a'}$, we obtain

(C.127)

$$\nu_{1,N_a'}=N_a'^{-1}\sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\mathbf{h}'_{.r,N_a'}\left[\left(\sum_{m=1}^{S}\widetilde{\rho}_{m,N_a'}(\omega)\right)^{l}-\left(\sum_{m=1}^{S}\rho_{m,N_a'}\right)^{l}\right]'\overline{\mathbf{M}}_{N_a'}'^{l}\widetilde{\boldsymbol{\Sigma}}_{N_a'}\left[\left(\sum_{m=1}^{S}\widetilde{\rho}_{m,N_a'}(\omega)\right)^{k}-\left(\sum_{m=1}^{S}\rho_{m,N_a'}\right)^{k}\right]\overline{\mathbf{M}}_{N_a'}^{k}\mathbf{h}_{.s,N_a'}$$
$$-N_a'^{-1}\sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\mathbf{h}'_{.r,N_a'}\left[\left(\sum_{m=1}^{S}\widetilde{\rho}_{m,N_a'}(\omega)\right)^{l}-\left(\sum_{m=1}^{S}\rho_{m,N_a'}\right)^{l}\right]'\overline{\mathbf{M}}_{N_a'}'^{l}\boldsymbol{\Sigma}_{N_a'}\left[\left(\sum_{m=1}^{S}\widetilde{\rho}_{m,N_a'}(\omega)\right)^{k}-\left(\sum_{m=1}^{S}\rho_{m,N_a'}\right)^{k}\right]\overline{\mathbf{M}}_{N_a'}^{k}\mathbf{h}_{.s,N_a'}.$$

Hence, we can then write

$$\nu_{1,N_a'}=\sum_{k=1}^{\infty}\sum_{l=1}^{\infty}X_{N_a'}^{(k,l)}, \tag{C.96}$$

where $X_{N_a'}^{(k,l)}=a_{N_a'}^{(k,l)}\aleph_{N_a'}^{(k,l)}$ with

$$a_{N_a'}^{(k,l)}=\frac{\left[\left(\sum_{m=1}^{S}\widetilde{\rho}_{m,N_a'}\right)^{l}-\left(\sum_{m=1}^{S}\rho_{m,N_a'}\right)^{l}\right]}{\rho_*^{l}}\frac{\left[\left(\sum_{m=1}^{S}\widetilde{\rho}_{m,N_a'}\right)^{l}-\left(\sum_{m=1}^{S}\rho_{m,N_a'}\right)^{k}\right]}{\rho_*^{k}} \text{ and} \tag{C.97}$$

$$\aleph_{N_a'}^{(k,l)} = N'^{-1} p_*^{l+k} \mathbf{h}'_{.r,N_a'} \overline{\mathbf{M}}'^{l}_{N_a'} \widetilde{\mathbf{\Sigma}}_{N_a'} \overline{\mathbf{M}}^{k}_{N_a'} \mathbf{h}_{.s,N_a'} - N^{-1} p_*^{l+k} \mathbf{h}'_{.r,N_a'} \overline{\mathbf{M}}'^{l}_{N_a'} \mathbf{\Sigma}_{N_a'} \overline{\mathbf{M}}^{k}_{N} \mathbf{h}_{.s,N_a'} . \tag{C.98}$$

Note that $\left| a_{N_a'}^{(k,l)} \right| \to 0$ in light of the aforementioned results. It follows that $\left| X_{N_a'}^{(k,l)} \right| \to 0$. Moreover,

$$\begin{aligned} \left| a_{N_a'} \right| &= \left| \frac{\left(\sum_{m=1}^{S} \widetilde{\rho}_{m,N_a'} \right)^{l} - \left(\sum_{m=1}^{S} \rho_{m,N_a'} \right)^{l}}{\rho_*^{l}} \frac{\left(\sum_{m=1}^{S} \widetilde{\rho}_{m,N_a'} \right)^{l} - \left(\sum_{m=1}^{S} \rho_{m,N_a'} \right)^{k}}{\rho_*^{k}} \right| \\ &\leq 2 \left(\frac{\rho_{**}}{\rho_*} \right)^{l} 2 \left(\frac{\rho_{**}}{\rho_*} \right)^{k} = 4 \left(\frac{\rho_{**}}{\rho_*} \right)^{l+k} \end{aligned} \tag{C.99}$$

for $N \geq \overline{N}_{\omega}$, $\left| \aleph_{N_a'}^{(k,l)}(\omega) \right| \leq 2K(c)$, such that we have

$$\left| X_{N_a'}^{(k,l)} \right| \leq B^{(l,k)} = 2K(c) 4 \left(\frac{\rho_{**}}{\rho_*} \right)^{l+k} = 8K(c) \left(\frac{\rho_{**}}{\rho_*} \right)^{l+k} . \tag{C.100}$$

Hence, there exists a dominating function $B^{(l,k)}$ for all values of *k*,*l*. Moreover, since $\left(\frac{\rho_{**}}{\rho_*} \right) < 1$ by construction, we also have that

$$\sum_{k=1}^{\infty} \sum_{l=1}^{\infty} \left| B^{(l,k)} \right| = \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} B^{(l,k)} < \infty \ , \tag{C.101}$$

i.e., the dominating function is integrable (summable).

It follows from dominated convergence that

$$\lim_{N_a' \to \infty} \nu_{1,N_a'} = 0 . \tag{C.102}$$

The same holds for the $\nu_{i,N_a'}$, $i = 2,...,7$. It follows that $\nu_{i,N_a'} \to 0$ as N_a' and in light of Resnik (1999, p. 171) that $\nu_N \to o_p(1)$.

Thus, $N^{-1} \widetilde{\mathbf{F}}'_N \widetilde{\mathbf{\Sigma}}_N \widetilde{\mathbf{F}}_N - N^{-1} \mathbf{F}'_N \mathbf{\Sigma}_N \mathbf{F}_N = o_p(1)$. That $N^{-1} \mathbf{F}'_N \mathbf{\Sigma}_N \mathbf{F}_N = O(1)$ follows from the properties maintained for the row and column sums of $(\mathbf{I}_N - \sum_{m=1}^{S} \rho_m \mathbf{M}_{m,N})^{-1}$ and the elements of $\mathbf{\Sigma}_N$ and $\mathbf{H}_N$.

Proof of Theorem 3.

To show that $\widetilde{\mathbf{\Omega}}_{\widetilde{\rho}_N} - \mathbf{\Omega}_{\widetilde{\rho}_N} = o_p(1)$, we first prove that $\widetilde{\mathbf{\Psi}}_N - \mathbf{\Psi}_N = o_p(1)$, using the expressions for $\widetilde{\mathbf{\Psi}}_N$ and $\mathbf{\Psi}_N$ as given by (C.51)–(C.54). By assumption the row and columns sums of the matrices $\mathbf{A}_{1s,N}$ and $\mathbf{A}_{2s,N}$, $s = 1,...,S$, are uniformly bounded in absolute value, and hence so are the row and column sums of the matrices $(\mathbf{A}_{1s,N} + \mathbf{A}'_{1s,N})$ and $(\mathbf{A}_{2s,N} + \mathbf{A}'_{2s,N})$ as well as products of them (see Remark A.1 in Appendix A). It follows from Lemma C.3 and C.4 that $\widetilde{\psi}^{11,*}_{pq,N} - \psi^{11,*}_{pq,N} = o_p(1)$, $\psi^{11,*}_{pq,N} = O(1)$, $\widetilde{\psi}^{11,*}_{pq,N} = O_p(1)$. It is readily observed that the same holds true for $\widetilde{\psi}^{21,*}_{pq,N} = o_p(1)$, $\widetilde{\psi}^{12,*}_{pq,N}$, $\widetilde{\psi}^{22,*}_{pq,N}$.

Next observe that the row and columns sums of the matrices

$$(\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N} \mathbf{M}'_{m,N})(\mathbf{A}_{1s,N} + \mathbf{A}'_{1s,N})(\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N} \mathbf{M}_{m,N}) \text{ and} \tag{C.103}$$

$$(\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N} \mathbf{M}'_{m,N})(\mathbf{A}_{2s,N} + \mathbf{A}'_{2s,N})(\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N} \mathbf{M}_{m,N}) \tag{C.104}$$

are bounded uniformly in absolute value. It follows from Lemma C.1 that $\widetilde{\boldsymbol{\alpha}}_{1s,N} - \boldsymbol{\alpha}_{1s,N} = o_p(1)$ and $\boldsymbol{\alpha}_{2s,N} - \boldsymbol{\alpha}_{2s,N} = o_p(1)$, $\boldsymbol{\alpha}_{1s,N} = O(1)$, $\boldsymbol{\alpha}_{2s,N} = O(1)$, and thus $\widetilde{\boldsymbol{\alpha}}_{1s,N} = O_p(1)$, and $\widetilde{\boldsymbol{\alpha}}_{2s,N} = O_p(1)$.

By assumption $\widetilde{\mathbf{P}}_N - \mathbf{P}_N = o_p(1)$, $\mathbf{P}_N = O(1)$, and $\widetilde{\mathbf{P}}_N = O_p(1)$. By Lemma C.6 we have $N^{-1}\widetilde{\mathbf{F}}'_N \widetilde{\mathbf{\Sigma}}_N \widetilde{\mathbf{F}}_N - N^{-1}\mathbf{F}'_N \mathbf{\Sigma}_N \mathbf{F}_N = o_p(1)$, $N^{-1}\mathbf{F}'_N \mathbf{\Sigma}_N \mathbf{F}_N = O(1)$, and $N^{-1}\widetilde{\mathbf{F}}'_N \widetilde{\mathbf{\Sigma}}_N \widetilde{\mathbf{F}}_N = O_p(1)$. It follows that $\widetilde{\psi}^{11,**}_{pq,N} - \psi^{11,**}_{pq,N} = o_p(1)$, $\psi^{11,**}_{pq,N} = O(1)$, $\widetilde{\psi}^{11,**}_{pq,N} = O_p(1)$. It is readily observed that the same holds true for $\widetilde{\psi}^{21,**}_{pq,N}$, $\widetilde{\psi}^{12,**}_{pq,N}$, $\widetilde{\psi}^{22,**}_{pq,N}$. Hence, $\widetilde{\mathbf{\Psi}}_N - \mathbf{\Psi}_N = o_p(1)$, $\mathbf{\Psi}_N = O(1)$, and $\widetilde{\mathbf{\Psi}}_N = O_p(1)$.

By Assumption 5, we have $\widetilde{\mathbf{\Theta}}_N - \mathbf{\Theta}_N = o_p(1)$, $\mathbf{\Theta}_N = O(1)$ and $\widetilde{\mathbf{\Theta}}_N = O_p(1)$.

Let

$$\mathbf{\Xi}_N = \mathbf{J}'_N \mathbf{\Theta} \mathbf{J}_N = \mathbf{B}'_N \mathbf{\Gamma}'_N \mathbf{\Theta}_N \mathbf{\Gamma}_N \mathbf{B}_N \text{ and} \tag{C.105}$$

$$\widetilde{\mathbf{\Xi}}_N = \widetilde{\mathbf{J}}'_N \widetilde{\mathbf{\Theta}} \widetilde{\mathbf{J}}_N = \widetilde{\mathbf{B}}'_N \widetilde{\mathbf{\Gamma}}'_N \widetilde{\mathbf{\Theta}}_N \widetilde{\mathbf{\Gamma}}_N \widetilde{\mathbf{B}}_N .^{25} \tag{C.106}$$

[25] There is a slight discrepancy to the definition of $\widetilde{\mathbf{\Xi}}_N$ in Theorem 2: Here $\widetilde{\mathbf{B}}_N$ is used rather than $\overline{\mathbf{B}}_N$, which does not affect the proof, however, noting that both $\widetilde{\boldsymbol{\rho}}_N$ and $\overline{\boldsymbol{\rho}}_N$ are consistent.

In Theorem 2, we showed that $\widetilde{\mathbf{J}}_N = O_p(1)$, $\mathbf{J}_N = O(1)$, and $\widetilde{\mathbf{J}}_N - \mathbf{J}_N = o_p(1)$ and that $\widetilde{\boldsymbol{\Xi}}_N^{+} = O_p(1)$, $\boldsymbol{\Xi}_N^{-1} = O_p(1)$ and that $\widetilde{\boldsymbol{\Xi}}_N^{+} - \boldsymbol{\Xi}_N^{-1} = o_p(1)$. It now follows that $\widetilde{\boldsymbol{\Omega}}_N - \boldsymbol{\Omega}_N = o_p(1)$.

IV. Proof of Theorem 4 (Joint Distribution of $\widetilde{\boldsymbol{\rho}}_N$ and Other Model Parameters)

The subsequent proof will focus on the case $\mathbf{F}_N = (\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N} \mathbf{M}'_{m,N})^{-1} \mathbf{H}_N$, since this covers also the case where $\mathbf{F}_N = (\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N} \mathbf{M}'_{m,N}) \mathbf{H}_N$. The first line in (26) holds in light of Assumption 7 and Theorem 2 .We next prove that $\boldsymbol{\xi}_{o,N} \xrightarrow{d} N(0, \mathbf{I}_{P^*+2S})$ by verifying that the assumptions of the central limit theorem A.1 by Kelejian and Prucha (2008) are fulfilled. Note that $\lambda_{\min}(\boldsymbol{\Psi}_{o,N}) \geq c^*_{\boldsymbol{\Psi}_o} > 0$ by assumption. In Theorem 2, we verified that the innovations $\boldsymbol{\varepsilon}_n$ and the elements of $\mathbf{a}_{1s,N}$, $\mathbf{a}_{2s,N}$ and $\mathbf{A}_{1s,N}$, $\mathbf{A}_{2s,N}$ appearing in $\mathbf{v}_N$ satisfy all assumptions stated in Theorem A.1.

Next, consider

$$\mathbf{F}_N = (\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N} \mathbf{M}'_{m,N})^{-1} \mathbf{H}_N . \tag{C.107}$$

Since the row and columns sums of $(\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N} \mathbf{M}'_{m,N})^{-1}$ are uniformly bounded in absolute value and since the elements of the matrix $\mathbf{H}_N$ are uniformly bounded in absolute value, it follows that the elements of $\mathbf{F}_N$ are also uniformly bounded in absolute value. Hence, the linear form $\mathbf{F}'_N \boldsymbol{\varepsilon}_N$ also fulfils the assumptions of Theorem A.1. As a consequence, $\boldsymbol{\xi}_{o,N} \xrightarrow{d} N(0, \mathbf{I}_{P^*+2S})$.

In the proofs of Theorems 2 and 3, we showed that $\widetilde{\boldsymbol{\Psi}}_N - \boldsymbol{\Psi}_N = o_p(1)$, $\boldsymbol{\Psi}_N = O(1)$, and $\widetilde{\boldsymbol{\Psi}}_N = O_p(1)$. By analogous arguments, this also holds for the submatrices $\boldsymbol{\Psi}_{\Delta\Delta,N}$ and $\boldsymbol{\Psi}_{\Delta\rho,N}$. Hence, $\widetilde{\boldsymbol{\Psi}}_{o,N} - \boldsymbol{\Psi}_{o,N} = o_p(1)$, $\boldsymbol{\Psi}_{o,N} = O(1)$ and thus $\widetilde{\boldsymbol{\Psi}}_{o,N} = O_p(1)$.

By assumption $\widetilde{\mathbf{P}}_N - \mathbf{P}_N = o_p(1)$, $\mathbf{P}_N = O(1)$, and $\widetilde{\mathbf{P}}_N = O_p(1)$ as well as $\widetilde{\boldsymbol{\Theta}}_N - \boldsymbol{\Theta}_N = o_p(1)$, $\boldsymbol{\Theta}_N = O(1)$ and $\widetilde{\boldsymbol{\Theta}}_N = O_p(1)$. In the proof of Theorem 2 we showed that $\widetilde{\mathbf{J}}_N - \mathbf{J}_N = o_p(1)$, $\mathbf{J}_N = O(1)$, and $\widetilde{\mathbf{J}}_N = O_p(1)$, and that $(\widetilde{\mathbf{J}}'_N \widetilde{\boldsymbol{\Theta}}_N \widetilde{\mathbf{J}}_N)^{+} - (\mathbf{J}'_N \boldsymbol{\Theta}_N \mathbf{J}_N)^{-1} = o_p(1)$, $(\mathbf{J}'_N \boldsymbol{\Theta}_N \mathbf{J}_N)^{-1} = O(1)$,

and $(\widetilde{\mathbf{J}}_N'\widetilde{\mathbf{\Theta}}_N\widetilde{\mathbf{J}}_N)^+ = O_p(1)$. It now follows that $\widetilde{\mathbf{\Omega}}_{o,N} - \mathbf{\Omega}_{o,N} = o_p(1)$ and $\mathbf{\Omega}_{o,N} = O(1)$ and thus $\widetilde{\mathbf{\Omega}}_{o,N} = O_p(1)$.

Appendix D.

Proof of Lemma 1.

In light of equation (2a) and (2b), Assumptions 3 and 8, as well as $\sup_N \|\boldsymbol{\beta}_N\| \leq b < \infty$ it follows that all columns of $\mathbf{Z}_N = (\mathbf{X}_N, \overline{\mathbf{Y}}_N)$ are of the form $\vartheta_N = \boldsymbol{\pi}_N + \mathbf{\Pi}_N\boldsymbol{\varepsilon}_N$, where the elements of the vector $\boldsymbol{\pi}_N$ and the row and column sums of the matrix $\mathbf{\Pi}_N$ are uniformly bounded in absolute value (see Remark A.1 in Appendix A). It follows from Lemma C.2 in Kelejian and Prucha (2008) that the fourth moments of the elements of the matrix $\mathbf{D}_N = -\mathbf{Z}_N$ are uniformly bounded by some finite constant and that Assumption 6 holds.

Next, note that

$$N^{1/2}(\widetilde{\boldsymbol{\delta}}_N - \boldsymbol{\delta}_N) = \widetilde{\mathbf{P}}_N' N^{-1/2}\mathbf{F}_N'\boldsymbol{\varepsilon}_N, \tag{D.1}$$

where $\widetilde{\mathbf{P}}_N$ is defined in the Lemma and $\mathbf{F}_N = (\mathbf{I}_N - \sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N}')^{-1}\mathbf{H}_N$. In light of Assumption 10, $\widetilde{\mathbf{P}}_N - \mathbf{P}_N = o_p(1)$ and $\mathbf{P}_N = O(1)$, with $\mathbf{P}_N$ as defined in the Lemma. By Assumption 3 and Assumption 9, the elements of $\mathbf{F}_N$ are uniformly bounded in absolute value. By Assumption 2, $E(\boldsymbol{\varepsilon}_N) = \mathbf{0}$ and its diagonal variance-covariance matrix has uniformly bounded elements. Thus, $E(N^{-1/2}\mathbf{F}_N'\boldsymbol{\varepsilon}_N) = \mathbf{0}$ and the elements of the variance-covariance matrix of $N^{-1/2}\mathbf{F}_N'\boldsymbol{\varepsilon}_N$, i.e., $N^{-1}\mathbf{F}_N'\mathbf{\Sigma}_N\mathbf{F}_N$ are uniformly bounded in absolute value (see Remark A.1 in Appendix A). It follows from Chebychev's inequality that $N^{-1/2}\mathbf{F}_N'\boldsymbol{\varepsilon}_N = O_p(1)$, and consequently $N^{1/2}(\widetilde{\boldsymbol{\delta}}_N - \boldsymbol{\delta}_N) = \mathbf{P}_N N^{-1/2}\mathbf{F}_N'\boldsymbol{\varepsilon}_N + o_p(1)$ and $\mathbf{P}_N N^{-1/2}\mathbf{F}_N'\boldsymbol{\varepsilon}_N = O_p(1)$. This completes the proof, recalling that $\mathbf{T}_N = \mathbf{F}_N\mathbf{P}_N$.

Proof of Lemma 2.

Note from (1b) and (1c) that

$$\breve{\mathbf{y}}_N^* = \breve{\mathbf{Z}}_N^*\boldsymbol{\delta}_N + \boldsymbol{\varepsilon}_N - \sum_{m=1}^{S}(\breve{\boldsymbol{\rho}}_N - \boldsymbol{\rho}_N)\mathbf{M}_{m,N}\mathbf{u}_N. \tag{D.2}$$

It follows that

$$N^{1/2}(\check{\hat{\boldsymbol{\delta}}}_N - \boldsymbol{\delta}_N) = (N^{-1}\hat{\check{\mathbf{Z}}}_N^{*\prime}\check{\mathbf{Z}}_N^{*})^{-1}N^{-1/2}\hat{\check{\mathbf{Z}}}_N^{*\prime}\,[\boldsymbol{\varepsilon}_N - \sum_{m=1}^{S}(\check{\boldsymbol{\rho}}_N - \boldsymbol{\rho}_N)\mathbf{M}_{m,N}\mathbf{u}_N]$$

$$= \check{\mathbf{P}}_N^{*\prime}[N^{-1/2}\check{\mathbf{F}}_N^{*\prime}\boldsymbol{\varepsilon}_N - \sum_{m=1}^{S}(\hat{\rho}_{m,N} - \rho_{m,N})N^{-1/2}\check{\mathbf{F}}_N^{**\prime}\boldsymbol{\varepsilon}_N],$$

where $\check{\mathbf{P}}_N^{*}$ is defined as in the Lemma,

$$\check{\mathbf{F}}_N^{*} = \mathbf{F}_N^{*} + \sum_{m=1}^{S}(\check{\rho}_{m,N} - \rho_{m,N})\overline{\mathbf{M}}_N\mathbf{H}_N\,,$$

and $\overline{\mathbf{M}}_N$ is a matrix, whose row and column sums are uniformly bounded in absolute value, satisfying $\sum_{m=1}^{S}(\check{\rho}_{m,N} - \rho_{m,N})\mathbf{M}_{m,N} = \sum_{m=1}^{S}(\check{\rho}_{m,N} - \rho_{m,N})\overline{\mathbf{M}}_N$. Moreover,

$$\check{\mathbf{F}}_N^{**} = \mathbf{F}_N^{**} + \sum_{m=1}^{S}(\check{\rho}_{m,N} - \rho_{m,N})(\mathbf{I}_N - \sum_{s=1}^{S}\rho_{m,N}\mathbf{M}_{m,N}^{\prime})^{-1}\overline{\mathbf{M}}_N^{\prime}\overline{\mathbf{M}}_N\mathbf{H}_N\,.$$

Hence, we have

$$N^{1/2}(\check{\hat{\boldsymbol{\delta}}}_N - \boldsymbol{\delta}_N) = \check{\mathbf{P}}_N^{*\prime}\{N^{-1/2}\mathbf{F}_N^{*\prime}\boldsymbol{\varepsilon}_N + \sum_{m=1}^{S}(\check{\rho}_{m,N} - \rho_{m,N})N^{-1/2}\overline{\mathbf{M}}_N\mathbf{H}_N\boldsymbol{\varepsilon}_N +$$

$$-\sum_{m=1}^{S}(\hat{\rho}_{m,N} - \rho_{m,N})N^{-1/2}\mathbf{F}_N^{**\prime}\boldsymbol{\varepsilon}_N - [\sum_{m=1}^{S}(\check{\rho}_{m,N} - \rho_{m,N})]^2N^{-1/2}(\mathbf{I}_N - \sum_{s=1}^{S}\rho_{m,N}\mathbf{M}_{m,N}^{\prime})^{-1}\overline{\mathbf{M}}_N^{\prime}\overline{\mathbf{M}}_N\mathbf{H}_N\boldsymbol{\varepsilon}_N\},$$

with $\mathbf{F}_N^{*}$ as defined in Lemma 2 and $\mathbf{F}_N^{**} = (\mathbf{I} - \sum_{s=1}^{S}\rho_{m,N}\mathbf{M}_{m,N}^{\prime})^{-1}\overline{\mathbf{M}}_N^{\prime}[\mathbf{I}_N - \sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N}]\mathbf{H}_N$.

In light of Assumption 12 and since $\check{\boldsymbol{\rho}}_N$ is $N^{1/2}$-consistent it follows that

$$N^{-1}\hat{\check{\mathbf{Z}}}_N^{*\prime}\check{\mathbf{Z}}_N^{*} - \mathbf{Q}_{\mathbf{H}^*\mathbf{Z}^*}^{\prime}\mathbf{Q}_{\mathbf{H}^*\mathbf{H}^*}^{-1\prime}\mathbf{Q}_{\mathbf{H}^*\mathbf{Z}^*} = o_p(1)\,. \tag{D.4}$$

Assumption 12 implies we also have $\mathbf{Q}_{\mathbf{H}^*\mathbf{Z}^*}^{\prime}\mathbf{Q}_{\mathbf{H}^*\mathbf{H}^*}^{-1\prime}\mathbf{Q}_{\mathbf{H}^*\mathbf{Z}^*} = O(1)$ and thus $(\mathbf{Q}_{\mathbf{H}^*\mathbf{Z}^*}^{\prime}\mathbf{Q}_{\mathbf{H}^*\mathbf{H}^*}^{-1\prime}\mathbf{Q}_{\mathbf{H}^*\mathbf{Z}^*})^{-1} = O(1)$. It follows as a special case of Pötscher and Prucha (1997, Lemma F1) that

$$(N^{-1}\hat{\check{\mathbf{Z}}}_N^{*\prime}\check{\mathbf{Z}}_N^{*})^{-1} - (\mathbf{Q}_{\mathbf{H}^*\mathbf{Z}^*}^{\prime}\mathbf{Q}_{\mathbf{H}^*\mathbf{H}^*}^{-1\prime}\mathbf{Q}_{\mathbf{H}^*\mathbf{Z}^*})^{-1} = o_p(1)\,. \tag{D.5}$$

It follows further that $\breve{\mathbf{P}}_N^* - \mathbf{P}_N^* = o_p(1)$ and $\mathbf{P}_N^* = O(1)$ with $\mathbf{P}_N^*$ defined in the Lemma. By arguments analoguous to the proof of Lemma 1 it follows that $N^{-1/2}\mathbf{F}_N^{*\prime}\boldsymbol{\varepsilon}_N = O_p(1)$, $N^{-1/2}\mathbf{F}_N^{**\prime}\boldsymbol{\varepsilon}_N = O_p(1)$, and also that $N^{-1/2}\overline{\mathbf{M}}_N\mathbf{H}_N\boldsymbol{\varepsilon}_N = O_p(1)$ and $N^{-1/2}[(\mathbf{I}_N - \sum_{s=1}^{S}\rho_{m,N}\mathbf{M}'_{m,N})^{-1}]\overline{\mathbf{M}}'_N\overline{\mathbf{M}}_N\mathbf{H}_N\boldsymbol{\varepsilon}_N = O_p(1)$.

As a consequence, $N^{1/2}(\hat{\breve{\boldsymbol{\delta}}}_N - \boldsymbol{\delta}_N) = \mathbf{P}_N^{*\prime}N^{-1/2}\mathbf{F}_N^{*\prime}\boldsymbol{\varepsilon}_N + o_p(1)$ and $\mathbf{P}_N^{*\prime}N^{-1/2}\mathbf{F}_N^{*\prime}\boldsymbol{\varepsilon}_N = O_p(1)$, observing again that $(\breve{\boldsymbol{\rho}}_N - \boldsymbol{\rho}_N) = o_p(1)$. This completes the proof, recalling that $\mathbf{T}_N^* = \mathbf{F}_N^*\mathbf{P}_N^*$.

APPENDIX E

Tables A.1 and A.2 show the Monte Carlos results for sample size $N = 100$ and $N = 250$, when the untransformed instruments matrix $\mathbf{H}$ is used in the FGTSLS estimation.

Table A1. Monte Carlo Results, $N = 100$, 2000 draws, instrument matrix **H**

Constellation[1]	(1a)	(1b)	(1c)	(2a)	(2b)	(2c)	(3a)	(3b)	(3c)	(4)	(5a)	(5b)	average [2]
λ_1	0.5	0.5	0.5	0.5	0.5	0.5	0	0	0	0.2	0	0.2	0.2833
Bias	0.0027	0.0027	0.0032	0.0010	0.0011	0.0015	0.0158	0.0074	0.0039	0.0377	0.0013	0.0033	0.0068
RMSE	0.0278	0.0289	0.0330	0.0241	0.0234	0.0261	0.0677	0.0595	0.0491	0.0792	0.0358	0.0359	0.0409
Rej. Rate	0.0755	0.0800	0.0920	0.0810	0.0820	0.0845	0.1560	0.1280	0.1105	0.2365	0.0830	0.0880	0.1081
λ_2	0.3	0.3	0	0.3	0.3	0	0	0	0	0.1	0	0.2	0.1250
Bias	0.0003	0.0004	-0.0009	-0.0012	-0.0011	-0.0012	0.0135	0.0051	0.0006	0.0332	-0.0004	0.0018	0.0042
RMSE	0.0359	0.0367	0.0392	0.0359	0.0359	0.0357	0.0679	0.0589	0.0513	0.0756	0.0414	0.0418	0.0464
Rej. Rate	0.0740	0.0795	0.0875	0.0790	0.0795	0.0800	0.1500	0.1225	0.0985	0.2250	0.0855	0.0820	0.1036
λ_3	0.1	0	0	0.1	0	0	0	0	0	0.1	0	0.2	0.0417
Bias	0.0017	0.0015	0.0021	0.0006	0.0010	0.0016	0.0131	0.0048	0.0019	0.0279	0.0016	0.0039	0.0051
RMSE	0.0289	0.0300	0.0326	0.0249	0.0253	0.0269	0.0561	0.0485	0.0437	0.0621	0.0349	0.0356	0.0375
Rej. Rate	0.0695	0.0710	0.0805	0.0775	0.0780	0.0775	0.1310	0.1050	0.0880	0.1965	0.0790	0.0810	0.0945
β_1	1	1	1	1	1	1	1	1	1	1	1	1	1.0000
Bias	0.0004	0.0004	0.0004	0.0002	0.0002	0.0002	0.0032	0.0014	0.0008	0.0066	0.0004	0.0007	0.0013
RMSE	0.0204	0.0205	0.0206	0.0214	0.0214	0.0214	0.0233	0.0222	0.0214	0.0239	0.0211	0.0208	0.0215
Rej. Rate	0.0760	0.0780	0.0780	0.0775	0.0770	0.0795	0.0840	0.0730	0.0765	0.1000	0.0780	0.0740	0.0793
β_2	1	1	1	1	1	1	1	1	1	1	1	1	1.0000
Bias	0.0006	0.0005	0.0006	0.0004	0.0004	0.0004	0.0031	0.0015	0.0009	0.0064	0.0006	0.0008	0.0014
RMSE	0.0202	0.0203	0.0204	0.0212	0.0212	0.0212	0.0227	0.0220	0.0211	0.0233	0.0210	0.0207	0.0213
Rej. Rate	0.0700	0.0705	0.0710	0.0705	0.0730	0.0705	0.0850	0.0775	0.0730	0.0890	0.0745	0.0715	0.0747
ρ_1	0.4	0.4	0.4	0	0	0	0.4	0.4	0.4	0.5	0	0.2	0.2583
Bias	-0.0452	-0.0469	-0.0582	-0.0604	-0.0647	-0.0644	-0.0349	-0.0449	-0.0559	-0.0296	-0.0655	-0.0466	-0.0514
RMSE	0.1748	0.1897	0.2269	0.2871	0.2883	0.2840	0.1687	0.1821	0.2122	0.1496	0.2556	0.2005	0.2183
Rej. Rate	0.0745	0.0900	0.1105	0.0795	0.0840	0.0905	0.0630	0.0835	0.0995	0.0600	0.0755	0.0720	0.0819
ρ_2	0.2	0.2	0	0	0	0	0.2	0.2	0	0.3	0	0.2	0.1083
Bias	-0.0260	-0.0317	-0.0124	-0.0433	-0.0477	-0.0374	-0.0217	-0.0257	-0.0137	-0.0274	-0.0448	-0.0377	-0.0308
RMSE	0.1807	0.1948	0.2030	0.2662	0.2613	0.2567	0.1740	0.1915	0.2003	0.1692	0.2322	0.1969	0.2106
Rej. Rate	0.0775	0.0830	0.0945	0.0875	0.0975	0.0945	0.0700	0.0795	0.0920	0.0650	0.0780	0.0835	0.0835
ρ_3	0.1	0	0	0	0	0	0.2	0	0	0.1	0	0.2	0.0500
Bias	-0.0106	-0.0004	-0.0018	-0.0075	-0.0087	-0.0070	-0.0221	0.0008	-0.0009	-0.0294	-0.0107	-0.0255	-0.0103
RMSE	0.1778	0.1944	0.2073	0.2488	0.2401	0.2387	0.1781	0.1937	0.1997	0.1729	0.2199	0.2037	0.2063
Rej. Rate	0.0615	0.0645	0.0665	0.0635	0.0615	0.0590	0.0680	0.0650	0.0670	0.0780	0.0490	0.0605	0.0637
Joint Tests [3]													
Rej. Rate			0.1370	0.0990	0.1055	0.1080	0.1925	0.1755	0.1645		0.1625		0.1431

Note: [1] Each colmn corresponds to one parameter constellation (see Table 1). [2] Average of absolute row values. [3] Rejections rates for the following hypotheses. (1c): $H_0^{\lambda,\rho,*}: \lambda_2 = \lambda_3 = \rho_2 = \rho_3 = 0$, (2a), (2b), (2c): $H_0^{\rho}: \rho_1 = \rho_2 = \rho_3 = 0$, (3a), (3b), (3c): $H_0^{\lambda}: \lambda_1 = \lambda_2 = \lambda_3 = 0$, (5a): $H_0^{\lambda,\rho}: \lambda_1 = \lambda_2 = \lambda_3 = \rho_1 = \rho_2 = \rho_3 = 0$.

Table A2. Monte Carlo Results, $N = 250$, 2000 draws, instrument matrix **H**

Constellation[1)]	(1a)	(1b)	(1c)	(2a)	(2b)	(2c)	(3a)	(3b)	(3c)	(4)	(5a)	(5b)	average [2)]
λ_1	0.5	0.5	0.5	0.5	0.5	0.5	0	0	0	0.2	0	0.2	0.2833
Bias	0.0017	0.0021	0.0022	0.0009	0.0010	0.0009	0.0084	0.0048	0.0025	0.0188	0.0008	0.0022	0.0039
RMSE	0.0179	0.0184	0.0204	0.0155	0.0150	0.0160	0.0420	0.0366	0.0301	0.0497	0.0222	0.0224	0.0255
Rej. Rate	0.0645	0.0640	0.0620	0.0620	0.0640	0.0620	0.0845	0.0770	0.0585	0.1310	0.0620	0.0645	0.0713
λ_2	0.3	0.3	0	0.3	0.3	0	0	0	0	0.1	0	0.2	0.1250
Bias	0.0005	0.0005	0.0003	-0.0001	-0.0003	0.0000	0.0064	0.0028	0.0007	0.0146	0.0001	0.0014	0.0023
RMSE	0.0205	0.0209	0.0220	0.0204	0.0205	0.0198	0.0379	0.0329	0.0291	0.0428	0.0235	0.0235	0.0261
Rej. Rate	0.0625	0.0645	0.0685	0.0590	0.0605	0.0700	0.0840	0.0725	0.0670	0.1375	0.0680	0.0680	0.0735
λ_3	0.1	0	0	0.1	0	0	0	0	0	0.1	0	0.2	0.0417
Bias	-0.0008	-0.0008	-0.0006	-0.0008	-0.0007	-0.0007	0.0039	0.0011	-0.0006	0.0095	-0.0005	0.0004	0.0017
RMSE	0.0198	0.0205	0.0219	0.0165	0.0168	0.0183	0.0313	0.0284	0.0273	0.0342	0.0232	0.0234	0.0235
Rej. Rate	0.0655	0.0665	0.0660	0.0645	0.0655	0.0680	0.0780	0.0730	0.0630	0.1090	0.0645	0.0705	0.0712
β_1	1	1	1	1	1	1	1	1	1	1	1	1	1.0000
Bias	0.0000	0.0001	-0.0001	-0.0001	-0.0001	-0.0001	0.0013	0.0006	0.0002	0.0028	-0.0001	0.0001	0.0005
RMSE	0.0124	0.0125	0.0126	0.0130	0.0129	0.0129	0.0139	0.0135	0.0129	0.0145	0.0127	0.0127	0.0130
Rej. Rate	0.0585	0.0590	0.0575	0.0620	0.0600	0.0565	0.0550	0.0585	0.0555	0.0630	0.0585	0.0545	0.0582
β_2	1	1	1	1	1	1	1	1	1	1	1	1	1.0000
Bias	0.0007	0.0005	0.0004	0.0004	0.0003	0.0000	0.0015	0.0008	0.0005	0.0031	0.0005	0.0007	0.0008
RMSE	0.0127	0.0128	0.0128	0.0131	0.0131	0.0131	0.0140	0.0136	0.0132	0.0147	0.0129	0.0130	0.0133
Rej. Rate	0.0665	0.0675	0.0665	0.0660	0.0650	0.0665	0.0645	0.0615	0.0645	0.0725	0.0650	0.0720	0.0665
ρ_1	0.4	0.4	0.4	0	0	0	0.4	0.4	0.4	0.5	0	0.2	0.2583
Bias	-0.0198	-0.0227	-0.0300	-0.0303	-0.0309	-0.0306	-0.0143	-0.0211	-0.0271	-0.0127	-0.0263	-0.0179	0.0236
RMSE	0.0907	0.0899	0.0983	0.1244	0.1242	0.1237	0.0903	0.0896	0.0970	0.0824	0.1222	0.1020	0.1029
Rej. Rate	0.0580	0.0580	0.0615	0.0605	0.0580	0.0570	0.0565	0.0610	0.0595	0.0580	0.0590	0.0540	0.0584
ρ_2	0.2	0.2	0	0	0	0	0.2	0.2	0	0.3	0	0.2	0.1083
Bias	-0.0078	-0.0103	-0.0030	-0.0178	-0.0172	-0.0131	-0.0050	-0.0073	-0.0029	-0.0100	-0.0123	-0.0111	0.0098
RMSE	0.1008	0.1018	0.1001	0.1164	0.1160	0.1158	0.0996	0.1014	0.1002	0.1012	0.1152	0.1039	0.1060
Rej. Rate	0.0560	0.0580	0.0555	0.0505	0.0470	0.0485	0.0620	0.0560	0.0540	0.0585	0.0490	0.0560	0.0542
ρ_3	0.1	0	0	0	0	0	0.2	0	0	0.1	0	0.2	0.0500
Bias	-0.0112	-0.0068	-0.0070	-0.0127	-0.0106	-0.0095	-0.0158	-0.0064	-0.0067	-0.0173	-0.0094	-0.0173	0.0109
RMSE	0.0962	0.0962	0.0972	0.1142	0.1143	0.1138	0.0941	0.0947	0.0969	0.0911	0.1136	0.1048	0.1023
Rej. Rate	0.0455	0.0415	0.0455	0.0410	0.0415	0.0415	0.0465	0.0400	0.0480	0.0620	0.0400	0.0420	0.0446
Joint Tests [3)]													
Rej. Rate			0.0640	0.0585	0.0570	0.0575	0.1035	0.0965	0.0930		0.0855		0.0769

Note: [1)] Each column corresponds to one parameter constellation (see Table 1). [2)] Average of absolute row values. [3)] Rejections rates for the following hypotheses. (1c): $H_0^{\lambda,\rho,*}: \lambda_2 = \lambda_3 = \rho_2 = \rho_3 = 0$, (2a), (2b), (2c): $H_0^{\rho}: \rho_1 = \rho_2 = \rho_3 = 0$, (3a), (3b), (3c): $H_0^{\lambda}: \lambda_1 = \lambda_2 = \lambda_3 = 0$, (5a): $H_0^{\lambda,\rho}: \lambda_1 = \lambda_2 = \lambda_3 = \rho_1 = \rho_2 = \rho_3 = 0$.

CESifo Working Paper Series

for full list see www.cesifo-group.org/wp
(address: Poschingerstr. 5, 81679 Munich, Germany, office@cesifo.de)

2293 M. Hashem Pesaran and Andreas Pick, Forecasting Random Walks Under Drift Instability, April 2008

2294 Steven Brakman, Gus Garita, Harry Garretsen and Charles van Marrewijk, Unlocking the Value of Cross-Border Mergers and Acquisitions, May 2008

2295 Eric O'N. Fisher and Kathryn G. Marshall, The Structure of the American Economy, May 2008

2296 Claudia M. Buch and Martin Schlotter, Regional Origins of Employment Volatility: Evidence from German States, May 2008

2297 Helmuth Cremer, Philippe De Donder, Dario Maldonado and Pierre Pestieau, Taxing Sin Goods and Subsidizing Health Care, May 2008

2298 Reinhilde Veugelers and Frederick van der Ploeg, Reforming European Universities: Scope for an Evidence-Based Process, May 2008

2299 Jon H. Fiva and Lars J. Kirkebøen, Does the Housing Market React to New Information on School Quality?, May 2008

2300 Tina Klautke and Alfons J. Weichenrieder, Interest Income Tax Evasion, the EU Savings Directive, and Capital Market Effects, May 2008

2301 Harald Badinger and Peter Egger, GM Estimation of Higher Order Spatial Autoregressive Processes in Panel Data Error Component Models, May 2008

2302 Jan K. Brueckner, Slot-Based Approaches to Airport Congestion Management, May 2008

2303 Sören Blomquist, Vidar Christiansen and Luca Micheletto, Public Provision of Private Goods and Nondistortionary Marginal Tax Rates, May 2008

2304 Dan Anderberg and Alessandro Balestrino, The Political Economy of Post-Compulsory Education Policy with Endogenous Credit Constraints, May 2008

2305 Tomer Blumkin, Yoram Margalioth and Efraim Sadka, The Role of Stigma in the Design of Welfare Programs, May 2008

2306 Vesa Kanniainen and Paolo M. Panteghini, Tax Neutrality: Illusion or Reality? The Case of Entrepreneurship, May 2008

2307 Thomas Dohmen, Armin Falk, David Huffman and Uwe Sunde, The Intergenerational Transmission of Risk and Trust Attitudes, May 2008

2308 Guglielmo Maria Caporale and Mario Cerrato, Using Chebyshev Polynomials to Approximate Partial Differential Equations, May 2008

2309 Peter Egger and Doina Maria Radulescu, Labour Taxation and Foreign Direct Investment, May 2008

2310 Laurent Linnemer, Dissipative Advertising Signals Quality even without Repeat Purchases, May 2008

2311 Jordi Jofre-Monseny and Albert Solé-Ollé, Which Communities should be afraid of Mobility? The Effects of Agglomeration Economies on the Sensitivity of Firm Location to Local Taxes, May 2008

2312 Andreas Haufler and Ferdinand Mittermaier, Unionisation Triggers Tax Incentives to Attract Foreign Direct Investment, May 2008

2313 Ronel Elul and Piero Gottardi, Bankruptcy: Is it enough to Forgive or must we also Forget?, May 2008

2314 Andreas Irmen and Johanna Kuehnel, Productive Government Expenditure and Economic Growth, May 2008

2315 Beate Henschel, Carsten Pohl and Marcel Thum, Demographic Change and Regional Labour Markets: The Case of Eastern Germany, May 2008

2316 Gabriel Felbermayr, Wido Geis and Wilhelm Kohler, Restrictive Immigration Policy in Germany: Pains and Gains Foregone?, May 2008

2317 Michael Hofmann, Gerhard Kempkes and Helmut Seitz, Demographic Change and Public Sector Budgets in a Federal System, May 2008

2318 Paul De Grauwe, Macroeconomic Modeling when Agents are Imperfectly Informed, June 2008

2319 Johann K. Brunner and Susanne Pech, Optimum Taxation of Inheritances, June 2008

2320 Thomas Eichner and Marco Runkel, Corporate Income Taxation of Multinationals in a General Equilibrium Model, June 2008

2321 Rainald Borck and Matthias Wrede, Subsidies for Intracity and Intercity Commuting, June 2008

2322 Patricia Apps and Ray Rees, Testing the Pareto Efficiency of Household Resource Allocations, June 2008

2323 Amihai Glazer, Vesa Kanniainen and Panu Poutvaara, Firms' Ethics, Consumer Boycotts, and Signalling, June 2008

2324 Claudia M. Buch, Jörg Döpke and Kerstin Stahn, Great Moderation at the Firm Level? Unconditional vs. Conditional Output Volatility, June 2008

2325 Helmuth Cremer, Philippe De Donder, Dario Maldonado and Pierre Pestieau, Forced Saving, Redistribution and Nonlinear Social Security Schemes, June 2008

2326 M. Hashem Pesaran and Paolo Zaffaroni, Optimal Asset Allocation with Factor Models for Large Portfolios, June 2008

2327 Harald Badinger and Peter Egger, Horizontal versus Vertical Interdependence in Multinational Activity, June 2008

2328 Jan K. Brueckner and Harris Selod, A Theory of Urban Squatting and Land-Tenure Formalization in Developing Countries, June 2008

2329 Paolo M. Panteghini, Corporate Debt, Hybrid Securities and the Effective Tax Rate, June 2008

2330 Guglielmo Maria Caporale, Juncal Cuñado and Luis A. Gil-Alana, Modelling Long-Run Trends and Cycles in Financial Time Series Data, June 2008

2331 Avi Ben-Bassat and Momi Dahan, Social Identity and Voter Turnout, June 2008

2332 Martin R. West and Ludger Wößmann, "Every Catholic Child in a Catholic School": Historical Resistance to State Schooling, Contemporary Private Competition, and Student Achievement across Countries, June 2008

2333 Erkki Koskela and Panu Poutvaara, Outsourcing and Labor Taxation in Dual Labor Markets, June 2008

2334 Philippe Choné and Laurent Linnemer, Optimal Litigation Strategies with Signaling and Screening, June 2008

2335 Albert Solé-Ollé and Pilar Sorribas-Navarro, Does Partisan Alignment Affect the Electoral Reward of Intergovernmental Transfers?, June 2008

2336 Antonio Cabrales and Piero Gottardi, Markets for Information: Of Inefficient Firewalls and Efficient Monopolies, June 2008

2337 Sumon Majumdar and Sharun W. Mukand, The Leader as Catalyst – on Leadership and the Mechanics of Institutional Change, June 2008

2338 Ulrich Hange, Tax Competition, Elastic Labor Supply, and Growth, June 2008

2339 Guy Laroque and Bernard Salanié, Does Fertility Respond to Financial Incentives?, June 2008

2340 Adriano Paggiaro, Enrico Rettore and Ugo Trivellato, The Effect of Extending the Duration of Eligibility in an Italian Labour Market Programme for Dismissed Workers, June 2008

2341 Helmut Seitz, Minimum Standards, Fixed Costs and Taxing Autonomy of Subnational Governments, June 2008

2342 Robert S. Chirinko, Leo de Haan and Elmer Sterken, Asset Price Shocks, Real Expenditures, and Financial Structure: A Multi-Country Analysis, July 2008

2343 Wolfgang Leininger, Evolutionarily Stable Preferences in Contests, July 2008

2344 Hartmut Egger and Udo Kreickemeier, Fairness, Trade, and Inequality, July 2008

2345 Ngo Van Long and Bodhisattva Sengupta, Yardstick Competition, Corruption, and Electoral Incentives, July 2008

2346 Florian Baumann, Employment Protection: The Case of Limited Enforceability, July 2008

2347 Alessandro Balestrino, Cinzia Ciardi and Claudio Mammini, On the Causes and Consequences of Divorce, July 2008

2348 Dirk Schindler and Benjamin Weigert, Insuring Educational Risk: Opportunities versus Income, July 2008

2349 Lammertjan Dam and Ben J. Heijdra, The Environmental and Macroeconomic Effects of Socially Responsible Investment, July 2008

2350 Avner Greif, Contract Enforcement and Institutions among the Maghribi Traders: Refuting Edwards and Ogilvie, July 2008

2351 Helmuth Cremer, Philippe De Donder, Dario Maldonado and Pierre Pestieau, Habit Formation and Labor Supply, July 2008

2352 Francesco Menoncin and Paolo M. Panteghini, The Johansson-Samuelson Theorem in General Equilibrium: A Rebuttal, July 2008

2353 Michael Kaganovich and Itzhak Zilcha, Alternative Social Security Systems and Growth, July 2008

2354 Keith Blackburn, Kyriakos C. Neanidis and M. Emranul Haque, Corruption, Seigniorage and Growth: Theory and Evidence, July 2008

2355 Edward Castronova, A Test of the Law of Demand in a Virtual World: Exploring the Petri Dish Approach to Social Science, July 2008

2356 Harald Badinger and Peter Egger, GM Estimation of Higher-Order Spatial Autoregressive Processes in Cross-Section Models with Heteroskedastic Disturbances, July 2008