Harald Badinger and Peter Egger

Estimation and Testing of Higher-Order Spatial Autoregressive Panel Data Error Component Models

Estimation and Testing of Higher-Order Spatial Autoregressive Panel Data Error Component Models

Harald Badinger
Department of Economics, Vienna University of Economics and Business

Peter Egger
Department of Management, Technology, and Economics at ETH Zürich; CEPR

May 2012

Abstract: This paper develops an estimator for higher-order spatial autoregressive panel data error component models with spatial autoregressive disturbances, SARAR(R,S). We derive the moment conditions and optimal weighting matrix without distributional assumptions for a generalized moments (GM) estimation procedure of the spatial autoregressive parameters of the disturbance process and define a generalized two-stage 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: C13, C21, C23

Keywords: Higher-order spatial dependence; Generalized moments estimation; Two-stage least squares; Asymptotic statistics

I. Introduction

This paper considers the estimation of panel data models with higher-order spatially autocorrelated error components and spatially autocorrelated dependent variables (SARAR). Spatial interactions in data may originate from various sources such as strategic interaction between jurisdictions (to attract firms or other mobile agents) and firms (in their price, quantity, or quality setting) or general equilibrium effects which disseminate with spatial decay due to their transmission through trade flows, migration, or input-output relationships.[1]

Data sets used in empirical studies often share two features: first, they are available in the form of panel data, with a large cross-sectional and a small time series dimension. Second, spatial interactions of various kinds co-exist – such as geography-related, trade-related, migration-related interactions – or the decay function of a single spatial interaction is unknown. The estimator proposed here addresses these two features in a unified framework and provides a flexible setup for applied work, allowing specification tests, estimation, and inference in higher-order random effects panel data models.

There are two main motivations for the use of higher-order models. First, the distance between two cross-sectional units is not necessarily (only) geographical in nature, a point prominently made in the political science literature by Beck, Gleditsch, and Beardsley (2006). Possible candidates for channels of spatial dependence beyond adjacency or geographical distance between units (such as countries) relate to i) economic distance (e.g., trade, cross-border lending, migration, input-output relationships, profit shifting of multinational firms), ii) socio-economic distance (e.g., differences in per capita income, age structure, ethnic composition of population), iii) cultural distance (language, index of individualism, religion), or iv) political/institutional distance (electoral systems, degree of federalism, voting in international organisations). A higher-order approach allows including several weights matrices that are based on alternative concepts of distance, whose relative importance to each other is unknown. Apart from the fact that assessing the relevance of alternative transmission channels of spillovers is of interest in itself, wrongly imposing a first-order spatial regressive process may misattribute part of the spatial dependence in the data (due to omitted transmission channels) to the single transmission channel included in the model. As a consequence, the estimates of the spatial regressive parameters and their standard errors will then be biased and inconsistent.

Second, even with only one channel of interdependence (e.g., related to geographical distance), the functional form of the true distance decay function, reflected in the elements of the weights matrix, is typically unknown. The standard approach to assume a known spatial

[1] See Anselin (2007) and Pinkse and Slade (2010) for surveys on the past, present, and future of spatial econometrics. A recent textbook on the matter is Le Sage and Pace (2009).

weights matrix with binary elements (e.g., for nearest neighbours) or elements that are specified as a decreasing function of distance (with known functional form and known distance decay parameter) seems highly restrictive. In fact, the distance decay function may exhibit discontinuities (e.g., border effects for interactions between units of different jurisdictions or countries) or a decay which is different from the one that is assumed by the researcher. Then, allowing subsets of the elements of the weights matrix to bear different spatial regressive parameters may significantly reduce the bias rooting in the assumption of an inadequate, preimposed decay function in a single weights matrix

The specification of higher-order models introduces non-trivial issues regarding the specification, interpretation, and estimation of these models. See Elhorst, Lacombe, and Piras (2012) and LeSage and Pace (2012) for a discussion of potential pitfalls in the use of higher-order models. However, these complications can be dealt with, while the alternative to stick with a misspecified first-order model for the sake of simplicity may result in inferior estimates. At the very least, the robustness of the results from a first order specification should be thoroughly explored, not only against variations in the specification of a single weights matrix (as is common in applied work), but also against the inclusion of further weights matrices, whenever economic theory suggests various channels of interdependence.

Estimation and testing of both random and fixed effects spatial regressive panel data models has been considered in the recent literature by Baltagi, Song, and Koh (2003) and Lee and Yu (2010) in a maximum likelihood framework and by Kapoor, Kelejain and Prucha (2007) as well as Mutl and Pfaffermayr (2011) using the generalized moments (GM) approach introduced by Kelejian and Prucha (1999). Obvious advantages of GM over ML estimation are that it does not rely on distributional assumptions and its computational simplicity. Moreover, comprehensive (cross-sectional) Monte Carlo Evidence by Arraiz, Drukker, Kelejian and Prucha (2010) shows that the large sample distribution provides a good approximation to the actual small sample distribution of the GM estimators of spatial regressive models.

The present paper builds on Kapoor, Kelejian, and Prucha (2007). They propose a GM estimator for the parameters of the spatial regressive error process in a random effects panel data model without endogenous explanatory variables (such as spatial lags of the dependent variable), derive a simplified weighting matrix for the moment conditions under the assumption of normally distributed error components, and prove consistency of the GM estimates. They also establish the asymptotic distribution of the regression parameters of the feasible generalized least squares (FGLS) estimates of the parameters of the main equation.

The present paper extends the estimation framework in Kapoor, Kelejian, and Prucha (2007) in several respects. In particular, it makes the following contributions:

- First, we do not only prove consistency of the proposed estimators but also derive the joint asymptotic distribution of the feasible generalized (two-stage) least squares estimates of the regression parameters and the GM estimates of the parameters of the spatial regressive disturbance process.

- Second, we allow for endogenous variables, including spatial lags of the dependent variable in the main equation, which is shown to affect the optimal weighting matrix for the moment conditions and the distribution of the GM estimates.
- Third, we dispense with the assumption of normally distributed error components, used by Kapoor, Kelejian, and Prucha (2007) to derive a simplified weighting matrix of the moments, retaining one of the main advantages of the GM approach of maximum-likelihood estimation.
- Fourth, we allow for higher-order rather than only first-order spatial regressive processes in both the dependent variable and the error process. This enables a more flexible design of the 'spatial' interdependence decay function and allows for the co-existence of more than one mode of interdependence as often suggested by economic theory (see Lee and Liu, 2010; and Badinger and Egger, 2011; for a treatment of higher-order spatial models with cross-section data).
- Finally, we provide some Monte Carlo evidence on the small sample performance of the proposed estimation procedure.

The remainder of the paper is organized as follows. Section II introduces the basic model specification. Section III proposes GM estimators of the parameters of spatial dependence in the error components. Section IV derives a two-stage least-squares (TSLS) routine to estimate the regression parameters of the model and derives the asymptotic distribution of all model parameters. Section V presents the results of a Monte Carlo simulation and section VI concludes. The detailed proofs are relegated to a technical appendix.

II. Basic Model Specification and Notation

The basic specification is a generalization of Kapoor, Kelejian, and Prucha (2007), who consider a panel data error components model with nonstochastic explanatory variables and first-order spatial autoregressive disturbances, i.e., a SAR(1) model. The present paper allows for an *R*-th order spatial autoregressive process in the dependent variable and an *S*-th order spatial process in the disturbances, i.e., we consider a SARAR(*R*,*S*) panel data error components model with $i=1,...,N$ cross-sectional units and $t=1,...,T$ time periods.[2] For time period *t*, the model reads

[2] Except for the ones on error components, the catalogue of assumptions in this paper extends to the case of fixed effects estimation (see Mundlak, 1978, for an early treatment of "within" parameter estimation in the context of an error components model).

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

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

where $\mathbf{y}_N(t)$ is an $N \times 1$ vector with cross-sectional observations of the dependent variable in year *t*, $\mathbf{X}_N(t)$ is an $N \times K$ matrix of observations on K non-stochastic explanatory variables, i.e., $\mathbf{X}_N(t) = [\mathbf{x}_{1,N}(t),...,\mathbf{x}_{K,N}(t)]$ with each $N \times 1$ vector $\mathbf{x}_{k,N}(t)$ denoting the observations on the *k*-th explanatory variable. The structure of spatial dependence in $\mathbf{y}_N(t)$ is determined by the time-invariant $N \times N$ matrices $\mathbf{W}_{r,N}$, $r = 1,...,R$, whose elements $w_{ij,r,N}$ are assumed to be known (and often specified as decreasing function of geographical distance). The expression $\bar{\mathbf{y}}_{r,N}(t) = \mathbf{W}_{r,N}\mathbf{y}_N(t)$ is referred to as the *r*-th spatial lag of $\mathbf{y}_N$. The specification of a higher-order process allows the strength of spatial interdependence in the dependent variable (reflected in the spatial autoregressive parameters $\lambda_{r,N}$, $r = 1,...,R$) to vary across a fixed number of R subsets of relations between cross-sectional units.

In Equ. (1b), the $N \times (K + R)$ design matrix is given by $\mathbf{Z}_N(t) = [\mathbf{X}_N(t), \overline{\mathbf{Y}}_N(t)]$, with $\overline{\mathbf{Y}}_N(t) = [\bar{\mathbf{y}}_{1,N}(t),...,\bar{\mathbf{y}}_{R,N}(t)]$, 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 autoregressive 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 $\mathbf{u}_N(t) = [u_{1,N}(t),...,u_{N,N}(t)]'$ is assumed to follow a spatial autoregressive process given by

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

$$\boldsymbol{\varepsilon}_N(t) = \boldsymbol{\mu}_N + \mathbf{v}_N(t), \tag{1d}$$

where $\rho_{m,N}$ and $\mathbf{M}_{m,N}$ denote the time-invariant, unknown parameters and the known $N \times N$ matrix of spatial interdependence, respectively. The structure of spatial correlation in the disturbances is determined by the S different, time-invariant $N \times N$ matrices $\mathbf{M}_{m,N}$. The expression $\bar{\mathbf{u}}_{m,N}(t) = \mathbf{M}_{m,N}\mathbf{u}_N(t)$ is referred to as the *m*-th spatial lag of $\mathbf{u}_N$. The $S \times 1$ vector of the spatial autoregressive parameters of $\mathbf{u}_N(t)$ is defined as $\boldsymbol{\rho}_N = (\rho_{1,N},...,\rho_{S,N})'$.

Finally, the $N \times 1$ vector of error terms $\boldsymbol{\varepsilon}_N(t)$ consists of two components, $\boldsymbol{\mu}_N$ and $\mathbf{v}_N(t)$. As indicated by the notation, $\boldsymbol{\mu}_N$ is time-invariant while $\mathbf{v}_N(t)$ is not. The typical elements of

$\boldsymbol{\varepsilon}_N(t)$ and $\mathbf{v}_N(t)$ are the scalars $\varepsilon_{it,N}$ and $v_{it,N}$, respectively, and the $N \times 1$ vector of unit-specific error components is given by $\boldsymbol{\mu}_N = (\mu_{1,N},...,\mu_{N,N})'$.

Stacking observations for all time periods such that t is the slow index and i is the fast index with all vectors and matrices, the model reads

$$\mathbf{y}_N = \mathbf{X}_N \boldsymbol{\beta}_N + \overline{\mathbf{Y}}_N \boldsymbol{\lambda}_N + \mathbf{u}_N \text{, or} \tag{2a}$$

$$\mathbf{y}_N = \mathbf{Z}_N \boldsymbol{\delta}_N + \mathbf{u}_N, \tag{2b}$$

with the $NT \times K$ regressor matrix $\mathbf{X}_N = [\mathbf{X}'_N(1),...,\mathbf{X}'_N(T)]'$, and $\overline{\mathbf{Y}}_N = (\overline{\mathbf{y}}_{1,N},...,\overline{\mathbf{y}}_{R,N})$, where $\mathbf{y}_{r,N} = [\mathbf{y}'_{r,N}(1),...,\mathbf{y}'_{r,N}(T)]'$ is the $NT \times 1$ vector of observations on the *r*-th spatial lag of the dependent variable $\overline{\mathbf{y}}_{r,N}$. The $NT \times 1$ vector of disturbances $\mathbf{u}_N = [\mathbf{u}'_N(1),...,\mathbf{u}'_N(T)]$ for the spatial autoregressive process of order S is given by

$$\mathbf{u}_N = \sum_{m=1}^{S} \rho_{m,N} (\mathbf{I}_T \otimes \mathbf{M}_{m,N}) \mathbf{u}_N + \boldsymbol{\varepsilon}_N, \tag{2c}$$

where $\mathbf{I}_T$ is an identity matrix of dimension $T \times T$. The $NT \times 1$ vector $\boldsymbol{\varepsilon}_N = [\boldsymbol{\varepsilon}'_N(1),...,\boldsymbol{\varepsilon}'_N(T)]'$ is specified as

$$\boldsymbol{\varepsilon}_N = (\mathbf{e}_T \otimes \mathbf{I}_N) \boldsymbol{\mu}_N + \mathbf{v}_N, \tag{3a}$$

where $\mathbf{e}_T$ is a unit vector of dimension $T \times 1$ and $\mathbf{I}_N$ is an identity matrix of dimension $N \times N$. In light of (2c), the error term can also be written as

$$\boldsymbol{\varepsilon}_N = \mathbf{u}_N - \sum_{m=1}^{S} \rho_{m,N} (\mathbf{I}_T \otimes \mathbf{M}_{m,N}) \mathbf{u}_N = \mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N} \mathbf{M}_{m,N}) \mathbf{u}_N . \tag{3b}$$

It follows that

$$\mathbf{u}_N = [\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N} \mathbf{M}_{m,N})^{-1}] \boldsymbol{\varepsilon}_N \text{, and} \tag{4a}$$

$$\mathbf{y}_N = [\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{r=1}^{R} \lambda_{r,N} \mathbf{W}_{r,N})^{-1}] \mathbf{X}_N(t) \boldsymbol{\beta}_N + [\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{r=1}^{R} \lambda_{r,N} \mathbf{W}_{r,N})^{-1}] \mathbf{u}_N, \tag{4b}$$

The following assumptions are maintained throughout this paper.
Assumption 1.

Let T be a fixed positive integer. (a) For all $1 \leq t \leq T$ and $1 \leq i \leq N, N \geq 1$, the error components $v_{it,N}$ are identically and (mutually) independently distributed with $E(v_{it,N}) = 0$, $E(v_{it,N}^2) = \sigma_v^2$, where $0 < \sigma_v^2 < b_v < \infty$, and $E\left|v_{it,N}\right|^{4+\eta} < \infty$ for some $\eta > 0$. (b) For all $1 \leq i \leq N, N \geq 1$, the unit-specific error components $\mu_{i,N}$ are identically and (mutually) independently distributed with $E(\mu_{i,N}) = 0$, $E(\mu_{i,N}^2) = \sigma_\mu^2$, where $0 < \sigma_\mu^2 < b_\mu < \infty$, and $E\left|\mu_{i,N}\right|^{4+\eta} < \infty$ for some $\eta > 0$. (c) The processes $\{v_{it,N}\}$ and $\{\mu_{i,N}\}$ are independent of each other. Assumption 1 is slightly stronger than that in Kapoor, Kelejian, and Prucha (2007), since it requires not only the fourth but also the $(4+\eta)$-th moments of the error components to be finite for some $\eta > 0$. This is required to invoke the central limit theorem of Kelejian and Prucha (2010) in the derivation of the asymptotic distribution in section III.

Assumption 1 implies that

$$E(\varepsilon_{it,N}\varepsilon_{js,N}) = \sigma_\mu^2 + \sigma_v^2 \text{ for } i = j \text{ and } t = s, \tag{5a}$$

$$E(\varepsilon_{it,N}\varepsilon_{js,N}) = \sigma_\mu^2 \text{ for } i = j \text{ and } t \neq s, \tag{5b}$$

$$E(\varepsilon_{it,N}\varepsilon_{js,N}) = 0, \text{ otherwise.} \tag{5c}$$

As a consequence, the variance-covariance matrix of the stacked error term $\boldsymbol{\varepsilon}_N$ reads

$$\boldsymbol{\Omega}_{\boldsymbol{\varepsilon},N} = E(\boldsymbol{\varepsilon}_N\boldsymbol{\varepsilon}_N') = \sigma_\mu^2(\mathbf{J}_T \otimes \mathbf{I}_N) + \sigma_v^2\mathbf{I}_{NT}, \tag{6a}$$

where $\mathbf{J}_T = \mathbf{e}_T\mathbf{e}_T'$ is a $T \times T$ matrix with unitary elements and $\mathbf{I}_{NT}$ is an identity matrix of dimension $NT \times NT$. Eq. (6a) can also be written as

$$\boldsymbol{\Omega}_{\boldsymbol{\varepsilon},N} = \sigma_v^2\mathbf{Q}_{0,N} + \sigma_1^2\mathbf{Q}_{1,N}, \tag{6b}$$

where $\sigma_1^2 = \sigma_v^2 + T\sigma_\mu^2$. The two matrices $\mathbf{Q}_{0,N}$ and $\mathbf{Q}_{1,N}$, which are central to the estimation of error component models and the moment conditions of the GM estimator, are defined as

$$\mathbf{Q}_{0,N} = (\mathbf{I}_T - \frac{\mathbf{J}_T}{T}) \otimes \mathbf{I}_N, \tag{7}$$

$$\mathbf{Q}_{1,N} = \frac{\mathbf{J}_T}{T} \otimes \mathbf{I}_N. \tag{8}$$

Notice that $\mathbf{Q}_{0,N}$ and $\mathbf{Q}_{1,N}$ are both of order *NT* × *NT*, symmetric, idempotent, orthogonal to each other, and sum up to $\mathbf{I}_{NT}$.

Assumption 2.

(a) All diagonal elements of $\mathbf{W}_{r,N}$, $r=1,...,R$, and $\mathbf{M}_{s,N}$, $s=1,...,S$, are zero. Without loss of generality, we assume that they are row-normalized in the following. b) The parameters $\lambda_{r,N}$, $r = 1, \ldots, R$, and $\rho_{s,N}$, $s = 1, \ldots, S$, are finite and contained in the admissible parameter spaces $\lambda_{r,N} \in (-\underline{a}_N^{\lambda_r}, \overline{a}_N^{\lambda_r})$ and $\rho_{s,N} \in (-\underline{a}_N^{\rho_s}, \overline{a}_N^{\rho_s})$; with row-normalized matrices, we have $\sum_{r=1}^{R} |\lambda_{r,N}| < 1$ and $\sum_{s=1}^{S} |\rho_{s,N}| < 1$.[3]

Assumption 2 ensures invertibility of $(\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N} \mathbf{M}_{m,N})$ and $(\mathbf{I}_N - \sum_{r=1}^{R} \lambda_{r,N} \mathbf{W}_{r,N})$ and thus that $\mathbf{y}_N$ and $\mathbf{u}_N$ are uniquely identified by (4a) and (4b).

We emphasize that all results of the present paper hold under alternative normalizations of the weights matrices with corresponding modifications of the admissible parameter space (Lee and Liu, 2010). Notice further that the assumptions regarding the admissible parameters space given here and in Lee and Liu (2010) are sufficient but not necessary and might be overly restrictive. A detailed discussion of a possible relaxation of the constraints on the admissible parameter space is provided by Koch (2011a,b), Elhorst, Lacome, and Piras (2012), and LeSage and Pace (2012).

Assumption 3.

The row and column sums of $\mathbf{W}_{r,N}$, $r=1,...,R$, $\mathbf{M}_{s,N}$, $s=1,...,S$, $(\mathbf{I}_N - \sum_{r=1}^{R} \lambda_{r,N} \mathbf{W}_{r,N})^{-1}$, and $(\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N} \mathbf{M}_{m,N})^{-1}$ are bounded uniformly in absolute value.

By Assumptions 1-3 and Remark A.1 in the Appendix, it follows that $E(\mathbf{u}_N) = \mathbf{0}$ and the variance-covariance matrix of $\mathbf{u}_N$ is given by

$$\mathbf{\Omega}_{\mathbf{u},N} = E(\mathbf{u}_N \mathbf{u}_N') = [\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N} \mathbf{M}_{m,N})^{-1}] \mathbf{\Omega}_{\varepsilon,N} [\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N} \mathbf{M}_{m,N}')^{-1}], \text{ and} \tag{9a}$$

[3] All results of the present paper hold under alternative normalizations of the weights matrices with corresponding modifications of the admissible parameter space (Lee and Liu, 2010).

$$E[\mathbf{u}_N(t)\mathbf{u}'_N(t)] = (\sigma_\mu^2 + \sigma_v^2)(\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N}\mathbf{M}_{m,N})^{-1}(\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N}\mathbf{M}'_{m,N})^{-1} . \tag{9b}$$

All variables (including $\mathbf{X}_N$) and parameters except for the variances of the error components are allowed to depend on sample size *N*. As a result, the model specification in Eqs. (1a)-(1c) allows for higher-order spatial dependence in the dependent variable, the explanatory variables, and the disturbances.

III. GM Estimation of a SAR(*S*) Model

Below, we derive GM estimators for the spatial autoregressive parameters of the disturbance process (1c) and the asymptotic joint distribution of all model parameters.

1. Moment Conditions

With an *S*-th order process (SAR(*S*), with $S > 1$), GM estimators of $\rho_{1,N},...,\rho_{S,N}$, σ_v^2, and σ_1^2 are obtained by recognizing that – under Assumptions 1 and 2 – the moment conditions used by Kapoor, Kelejian, and Prucha (2007) hold for each matrix $\mathbf{M}_{s,N}$, $s = 1,...,S$. Define for each $\mathbf{M}_{s,N}$, $s = 1,...,S$

$$\bar{\boldsymbol{\varepsilon}}_{s,N} = (\mathbf{I}_T \otimes \mathbf{M}_{s,N})\boldsymbol{\varepsilon}_N = (\mathbf{I}_T \otimes \mathbf{M}_{s,N})[\mathbf{u}_N - \sum_{m=1}^{S} \rho_{m,N}(\mathbf{I}_T \otimes \mathbf{M}_{m,N})\mathbf{u}_N] . \tag{10}$$

The moment conditions are then given by

$$\text{M}_\text{a} \quad E[\frac{1}{N(T-1)}\boldsymbol{\varepsilon}'_N\mathbf{Q}_{0,N}\boldsymbol{\varepsilon}_N] = E[\frac{1}{N(T-1)}\mathbf{v}'_N\mathbf{Q}_{0,N}\mathbf{v}_N] = \sigma_v^2 , \tag{11}$$

$$\text{M}_{1,s} \quad E[\frac{1}{N(T-1)}\bar{\boldsymbol{\varepsilon}}'_{s,N}\mathbf{Q}_{0,N}\bar{\boldsymbol{\varepsilon}}_{s,N}] = E[\frac{1}{N(T-1)}\mathbf{v}'_N\mathbf{Q}_{0,N}(\mathbf{I}_T \otimes \mathbf{M}'_{s,N}\mathbf{M}_{s,N})\mathbf{Q}_{0,N}\mathbf{v}_N] = \sigma_v^2 \frac{1}{N} tr(\mathbf{M}'_{s,N}\mathbf{M}_{s,N}) ,$$

$$\text{M}_{2,s} \quad E[\frac{1}{N(T-1)}\bar{\boldsymbol{\varepsilon}}'_{s,N}\mathbf{Q}_{0,N}\boldsymbol{\varepsilon}_N] = E[\frac{1}{N(T-1)}\mathbf{v}'_N\mathbf{Q}_{0,N}(\mathbf{I}_T \otimes \mathbf{M}'_{s,N})\mathbf{Q}_{0,N}\mathbf{v}_N] = 0 ,$$

$$\text{M}_\text{b} \quad E(\frac{1}{N}\boldsymbol{\varepsilon}'_N\mathbf{Q}_{1,N}\boldsymbol{\varepsilon}_N) = E[\frac{1}{N}\boldsymbol{\mu}'_N(\mathbf{e}'_T\mathbf{e}_T \otimes \mathbf{I}_N)\boldsymbol{\mu}_N] + E(\frac{1}{N}\mathbf{v}'_N\mathbf{Q}_{1,N}\mathbf{v}_N) = \sigma_1^2 ,$$

$$\text{M}_{3,s} \quad E(\frac{1}{N}\bar{\boldsymbol{\varepsilon}}'_{s,N}\mathbf{Q}_{1,N}\bar{\boldsymbol{\varepsilon}}_{s,N}) = E[\frac{1}{N}\boldsymbol{\mu}'_N(\mathbf{e}'_T\mathbf{e}_T \otimes \mathbf{M}'_{s,N}\mathbf{M}_{s,N})\boldsymbol{\mu}_N] + E[\frac{1}{N}\mathbf{v}'_N\mathbf{Q}_{1,N}(\mathbf{I}_T \otimes \mathbf{M}'_{s,N}\mathbf{M}_{s,N})\mathbf{Q}_{1,N}\mathbf{v}_N]$$
$$= \sigma_1^2 \frac{1}{N} tr(\mathbf{M}'_{s,N}\mathbf{M}_{s,N}) ,$$

$$\text{M}_{4,s} \quad E(\frac{1}{N}\bar{\boldsymbol{\varepsilon}}'_{s,N}\mathbf{Q}_{1,N}\boldsymbol{\varepsilon}_N) = E[\frac{1}{N}\boldsymbol{\mu}'_N(\mathbf{e}'_T\mathbf{e}_T \otimes \mathbf{M}'_{s,N})\boldsymbol{\mu}_N] + E[\frac{1}{N}\mathbf{v}'_N\mathbf{Q}_{1,N}(\mathbf{I}_T \otimes \mathbf{M}'_{s,N})\mathbf{Q}_{1,N}\mathbf{v}_N] = 0 ,$$

where $\sigma_1^2 = \sigma_v^2 + T\sigma_\mu^2$. The moment conditions associated with matrices $\mathbf{M}_{s,N}$, $s = 1,...,S$, through (10), are indexed with subscripts 1 to 4. The remaining two moment conditions are independent of s and denoted as M_a and M_b. For an *S*-th order process as in (2c), we thus have ($4S + 2$) moment conditions.

It is apparent that under Assumptions 1-3 from $M_{1,s}$ and $M_{3,s}$ that there are potentially $S(S-1)$ further moment conditions, namely

$E[\frac{1}{N(T-1)}\bar{\boldsymbol{\varepsilon}}'_{s,N}\mathbf{Q}_{0,N}\bar{\boldsymbol{\varepsilon}}_{s',N}] = \sigma_v^2\frac{1}{N}tr(\mathbf{M}'_{s,N}\mathbf{M}_{s',N}), s \neq s'$ from $M_{1,s}$ and

$E(\frac{1}{N}\bar{\boldsymbol{\varepsilon}}'_{s,N}\mathbf{Q}_{1,N}\bar{\boldsymbol{\varepsilon}}_{s',N}) = \sigma_1^2\frac{1}{N}tr(\mathbf{M}'_{s,N}\mathbf{M}_{s',N}), s \neq s'$ from $M_{3,s}$[4]. For simplicity, we use only the moment conditions in (11) below, which are always available for weights matrices satisfying Assumptions 1 and 2. However, the results carry over to the more general estimator using all $4S + S(S-1) + 2$ moment conditions.

Substituting (3b), (10), and (1c) into the $4S + 2$ moment conditions (11) yields a ($4S + 2$) equation system in $(\rho_{1,N},...,\rho_{S,N},\sigma_v^2,\sigma_1^2)$, which can be written as

$$\boldsymbol{\gamma}_N - \boldsymbol{\Gamma}_N\mathfrak{b}_N = \mathbf{0}, \tag{12}$$

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

$$\mathfrak{b}_N = (\rho_{1,N},...,\rho_{S,N},\rho_{1,N}^2,...,\rho_{S,N}^2,\rho_{1,N}\rho_{2,N},...,\rho_{1,N}\rho_{S,N},...,\rho_{S-1,N}\rho_{S,N},\sigma_v^2,\sigma_1^2)',$$

i.e., $\mathfrak{b}_N$ contains S linear terms $\rho_{m,N}$, $m = 1,...,S$, S quadratic terms $\rho_{m,N}^2$, $m = 1,...,S$, $S(S-1)/2$ cross products $\rho_{m,N}\rho_{l,N}$, $m = 1,...,S-1, l = m+1,...,S$, as well as σ_v^2 and σ_1^2. For later reference, we define the $(S+2)\times 1$ vector of all parameters as $\boldsymbol{\theta}_N = (\boldsymbol{\rho}'_N,\sigma_v^2,\sigma_1^2)' = (\rho_{1,N},...,\rho_{S,N},\sigma_v^2,\sigma_1^2)'$.

$\boldsymbol{\gamma}_N$ is a $(4S+2)\times 1$ vector with elements $[\gamma_{i,N}]$, $i = 1,...,(4S+2)$, and $\boldsymbol{\Gamma}_N$ is a $(4S+2)\times[2S + S(S-1)/2 + 2]$ matrix with elements $[\gamma_{i,j,N}]$, $i = 1,...,(4S+2)$,

[4] The efficiency gain from using these additional moment conditions depends on the properties of the weights matrices. If two weights matrices are orthogonal, i.e., $\mathbf{M}'_{s,N}\mathbf{M}_{s',N} = \mathbf{0}$, the corresponding moment condition is trivially satisfied for any set of (finite) parameter values and does not add any information.

$j = 1,...,[2S + S(S-1)/2 + 2]$. The elements $\gamma_{i,N}$ and $\gamma_{i,j,N}$ 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 (12) has the following order. The first four rows correspond to moment restrictions $M_{1,1}$ to $M_{4,1}$ associated with matrix $\mathbf{M}_{1,N}$ through (10); rows five to eight correspond to $M_{1,2}$ to $M_{4,2}$ associated with matrix $\mathbf{M}_{2,N}$, and so forth; rows $(S-4)$ to $4S$ correspond to the $M_{1,S}$ to $M_{4,S}$ associated with matrix $\mathbf{M}_{S,N}$. Finally, rows $(4S+1)$ and $(4S+2)$ correspond to moment conditions M_a and M_b, respectively, which are independent of *s*.

The sample analogue to (12) is given by

$$\tilde{\boldsymbol{\gamma}}_N - \tilde{\boldsymbol{\Gamma}}_N \mathfrak{b}_N = \vartheta_N(\boldsymbol{\theta}_N), \tag{13}$$

where the elements of $\tilde{\boldsymbol{\gamma}}_N$ and $\tilde{\boldsymbol{\Gamma}}_N$ are equal to those of $\boldsymbol{\gamma}_N$ and $\boldsymbol{\Gamma}_N$ with the expectations operator suppressed and the disturbances $\mathbf{u}_N$ replaced by (consistent) estimates $\tilde{\mathbf{u}}_N$.

GM estimates of parameters $\rho_{1,N},...,\rho_{S,N}$, σ_v^2 and σ_1^2 are then obtained as the solution to

$$\underset{\rho_1,\rho_2,...,\rho_S,\sigma_v^2,\sigma_1^2}{\arg\min} [(\tilde{\boldsymbol{\gamma}}_N - \tilde{\boldsymbol{\Gamma}}_N \mathfrak{b}_N)' \tilde{\boldsymbol{\Theta}}_N (\tilde{\boldsymbol{\gamma}}_N - \tilde{\boldsymbol{\Gamma}}_N \mathfrak{b}_N)] = [\vartheta_N(\boldsymbol{\theta}_N)' \tilde{\boldsymbol{\Theta}}_N \vartheta_N(\boldsymbol{\theta}_N)], \tag{14}$$

i.e., the parameter estimates can be obtained from a (weighted) non-linear least squares regression of $\tilde{\boldsymbol{\gamma}}_N$ on the columns of $\tilde{\boldsymbol{\Gamma}}_N$. The optimal choice of the $(4S+2)\times(4S+2)$ weighting matrix $\boldsymbol{\Theta}_N$ will be discussed below.

Below, we define the elements of $\boldsymbol{\gamma}_N$ and $\boldsymbol{\Gamma}_N$, grouped by the corresponding moment conditions, using

$$\bar{\mathbf{u}}_{s,N} = (\mathbf{I}_T \otimes \mathbf{M}_{s,N})\mathbf{u}_N, \; s = 1,...,S, \text{ and} \tag{15a}$$

$$\bar{\bar{\mathbf{u}}}_{sm,N} = (\mathbf{I}_T \otimes \mathbf{M}_{s,N})(\mathbf{I}_T \otimes \mathbf{M}_{m,N})\mathbf{u}_N = (\mathbf{I}_T \otimes \mathbf{M}_{s,N}\mathbf{M}_{m,N})\mathbf{u}_N, \; s = 1,...,S, \; m = 1,...,S. \tag{15b}$$

$M_{1,s}$ delivers $s = 1,...,S$ rows $4(s-1)+1$ in (12):

$$\gamma_{4(s-1)+1,N} = \frac{1}{N(T-1)} E(\bar{\mathbf{u}}'_{s,N}\mathbf{Q}_{0,N}\bar{\mathbf{u}}_{s,N}), \tag{16a}$$

$$\gamma_{4(s-1)+1,m,N} = \frac{2}{N(T-1)} E(\bar{\mathbf{u}}'_{s,N}\mathbf{Q}_{0,N}\bar{\bar{\mathbf{u}}}_{sm,N}), \; m = 1,...,S,$$

$$\gamma_{4(s-1)+1,S+m,N} = -\frac{1}{N(T-1)} E(\bar{\bar{\mathbf{u}}}'_{sm,N}\mathbf{Q}_{0,N}\bar{\bar{\mathbf{u}}}_{sm,N}),\ m=1,...,S,$$

$$\gamma_{4(s-1)+1,S(m+1)-m(m-1)/2+l-m,N} = -\frac{2}{N(T-1)} E(\bar{\bar{\mathbf{u}}}'_{sm,N}\mathbf{Q}_{0,N}\bar{\bar{\mathbf{u}}}_{sl,N}),\ m=1,...,S-1,\ l=m+1,...,S,$$

$$\gamma_{4(s-1)+1,2S+S(S-1)/2+1,N} = \frac{1}{N} tr(\mathbf{M}'_{s,N}\mathbf{M}_{s,N}),$$

$$\gamma_{4(s-1)+1,2S+S(S-1)/2+2,N} = 0.$$

$M_{2,s}$ consists of $s=1,...,S$ rows $4(s-1)+2$ in (12):

$$\gamma_{4(s-1)+2,N} = \frac{1}{N(T-1)} E(\bar{\mathbf{u}}'_{s,N}\mathbf{Q}_{0,N}\mathbf{u}_{N}), \tag{16b}$$

$$\gamma_{4(s-1)+2,m,N} = \frac{1}{N(T-1)} E(\bar{\bar{\mathbf{u}}}'_{sm,N}\mathbf{Q}_{0,N}\mathbf{u}_{N} + \bar{\mathbf{u}}'_{s,N}\mathbf{Q}_{0,N}\bar{\mathbf{u}}_{m,N}),\ m=1,...,S,$$

$$\gamma_{4(s-1)+2,S+m,N} = -\frac{1}{N(T-1)} E(\bar{\bar{\mathbf{u}}}'_{sm,N}\mathbf{Q}_{0,N}\bar{\mathbf{u}}_{m,N}),\ m=1,...,S,$$

$$\gamma_{4(s-1)+2,S(m+1)-m(m-1)/2+l-m,N} = -\frac{1}{N(T-1)} E(\bar{\bar{\mathbf{u}}}'_{sl,N}\mathbf{Q}_{0,N}\bar{\mathbf{u}}_{m,N} + \bar{\bar{\mathbf{u}}}'_{sm,N}\mathbf{Q}_{0,N}\bar{\mathbf{u}}_{l,N}),\ m=1,...,S-1,$$

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

$$\gamma_{4(s-1)+2,2S+S(S-1)/2+1,N} = 0,$$

$$\gamma_{4(s-1)+2,2S+S(S-1)/2+2,N} = 0.$$

$M_{3,s}$ corresponds to $s=1,...,S$ rows $4(s-1)+3$ in (12):

$$\gamma_{4(s-1)+3,N} = \frac{1}{N} E(\bar{\mathbf{u}}'_{s,N}\mathbf{Q}_{1,N}\bar{\mathbf{u}}_{s,N}), \tag{16c}$$

$$\gamma_{4(s-1)+3,m,N} = \frac{2}{N} E(\bar{\mathbf{u}}'_{s,N}\mathbf{Q}_{1,N}\bar{\bar{\mathbf{u}}}_{sm,N}),\ m=1,...,S,$$

$$\gamma_{4(s-1)+3,S+m,N} = -\frac{1}{N} E(\bar{\bar{\mathbf{u}}}'_{sm,N}\mathbf{Q}_{1,N}\bar{\bar{\mathbf{u}}}_{sm,N}),\ m=1,...,S,$$

$$\gamma_{4(s-1)+3,S(m+1)-m(m-1)/2+l-m,N} = -\frac{2}{N} E(\bar{\bar{\mathbf{u}}}'_{sm,N}\mathbf{Q}_{1,N}\bar{\bar{\mathbf{u}}}_{sl,N}),\ m=1,...,S-1,\ l=m+1,...,S,$$

$$\gamma_{4(s-1)+3,2S+S(S-1)/2+1,N} = 0,$$

$$\gamma_{4(s-1)+3,2S+S(S-1)/2+2,N} = \frac{1}{N} tr(\mathbf{M}'_{s,N}\mathbf{M}_{s,N}).$$

$M_{4,s}$ represents $s=1,...,S$ rows $4(s-1)+4$ in (12):

$$\gamma_{4(s-1)+4,N} = \frac{1}{N} E(\bar{\mathbf{u}}'_{s,N}\mathbf{Q}_{1,N}\mathbf{u}_{N}), \tag{16d}$$

$$\gamma_{4(s-1)+4,m,N} = \frac{1}{N} E(\bar{\bar{\mathbf{u}}}'_{sm,N}\mathbf{Q}_{1,N}\mathbf{u}_{N} + \bar{\mathbf{u}}'_{s,N}\mathbf{Q}_{1,N}\bar{\mathbf{u}}_{m,N}),\ m=1,...,S,$$

$$\gamma_{4(s-1)+4,S+m,N} = -\frac{1}{N} E(\bar{\bar{\mathbf{u}}}'_{sm,N} \mathbf{Q}_{1,N} \bar{\mathbf{u}}_{m,N}),\ m = 1,\ldots,S,$$

$$\gamma_{4(s-1)+4,S(m+1)-m(m-1)/2+l-m,N} = -\frac{1}{N} E(\bar{\bar{\mathbf{u}}}'_{sl,N} \mathbf{Q}_{1,N} \bar{\mathbf{u}}_{m,N} + \bar{\bar{\mathbf{u}}}'_{sm,N} \mathbf{Q}_{1,N} \bar{\mathbf{u}}_{l,N}),\ m = 1,\ldots,S-1,\ l = m+1,\ldots,S,$$

$$\gamma_{4(s-1)+4,2S+S(S-1)/2+1,N} = 0,$$

$$\gamma_{4(s-1)+4,2S+S(S-1)/2+2,N} = 0.$$

M_a reflects the equation in row $(4S+1)$ of (12):

$$\gamma_{4S+1,N} = \frac{1}{N(T-1)} E(\mathbf{u}'_N \mathbf{Q}_{0,N} \mathbf{u}_N), \tag{16e}$$

$$\gamma_{4S+1,m,N} = \frac{2}{N(T-1)} E(\bar{\mathbf{u}}'_{m,N} \mathbf{Q}_{0,N} \mathbf{u}_N),\ m = 1,\ldots,S,$$

$$\gamma_{4S+1,S+m,N} = -\frac{1}{N(T-1)} E(\bar{\mathbf{u}}'_{m,N} \mathbf{Q}_{0,N} \bar{\mathbf{u}}_{m,N}),\ m = 1,\ldots,S,$$

$$\gamma_{4S+1,S(m+1)-m(m-1)/2+l-m,N} = -\frac{2}{N(T-1)} E(\bar{\mathbf{u}}'_{m,N} \mathbf{Q}_{0,N} \bar{\mathbf{u}}_{l,N}),\ m = 1,\ldots,S-1,\ l = m+1,\ldots,S,$$

$$\gamma_{4S+1,2S+S(S-1)/2+1,N} = 1,$$

$$\gamma_{4S+1,2S+S(S-1)/2+2,N} = 0.$$

M_b is associated with row $(4S+2)$ of (12):

$$\gamma_{4S+2,N} = \frac{1}{N} E(\mathbf{u}'_N \mathbf{Q}_{1,N} \mathbf{u}_N), \tag{16f}$$

$$\gamma_{4S+2,m,N} = \frac{2}{N} E(\bar{\mathbf{u}}'_{m,N} \mathbf{Q}_{1,N} \mathbf{u}_N),\ m = 1,\ldots,S,$$

$$\gamma_{4S+2,S+m,N} = -\frac{1}{N} E(\bar{\mathbf{u}}'_{m,N} \mathbf{Q}_{1,N} \bar{\mathbf{u}}_{m,N}),\ m = 1,\ldots,S,$$

$$\gamma_{4S+2,S(m+1)-m(m-1)/2+l-m,N} = -\frac{2}{N} E(\bar{\mathbf{u}}'_{m,N} \mathbf{Q}_{1,N} \bar{\mathbf{u}}_{l,N}),\ m = 1,\ldots,S-1,\ l = m+1,\ldots,S,$$

$$\gamma_{4S+2,2S+S(S-1)/2+1,N} = 0,$$

$$\gamma_{4S+2,2S+S(S-1)/2+2,N} = 1.$$

For future reference, we define the $(2S+1)\times 1$ vector $\boldsymbol{\gamma}_N^0$ as the sub-vector containing rows s and $(s+1)$, $s = 1,\ldots,S$ and row $(4S+1)$ of $\boldsymbol{\gamma}_N$, corresponding to $M_{1,s}$, $M_{2,s}$, and M_a. Moreover, we define the $(2S+1)\times[2S+S(S-1)/2+1]$ matrix $\boldsymbol{\Gamma}_N^0$ as the sub-matrix containing rows s and $(s+1)$, $s = 1,\ldots,S$, and row $(4S+1)$ of $\boldsymbol{\Gamma}_N$, corresponding to $M_{1,s}$, $M_{2,s}$, and M_a.

Analogously, we define the $(2S+1)\times 1$ vector $\boldsymbol{\gamma}_N^1$ as the sub-vector containing rows $2s$, $(2s+1)$, $s=1,...,S$, and row $(4S+2)$ of $\boldsymbol{\gamma}_N$, corresponding to $M_{3,s}$, $M_{4,s}$, and M_b. Finally, we define the $(2S+1)\times[2S+S(S-1)/2+1]$ matrix $\boldsymbol{\Gamma}_N^1$ as the sub-matrix containing rows $2s$, $(2s+1)$, $s=1,...,S$, and $(4S+2)$ of $\boldsymbol{\Gamma}_N$, corresponding to $M_{3,s}$, $M_{4,s}$, and M_b.

2. Definition of GM Estimators

We next define three alternative GM estimators for the spatial autoregressive parameters of the disturbance process given by (1c) and the variances of the error components.[5]

2.1. Initial GM Estimation

The initial GM estimator is a special case of (14), using the identity matrix as weighting matrix $\boldsymbol{\Theta}_N$ and a subset of moment conditions (M_a, $M_{1,s}$ and $M_{2,s}$) only. It is based on $\boldsymbol{\gamma}_N^0$ and $\boldsymbol{\Gamma}_N^0$. Define $\boldsymbol{\theta}_N^0$ as the corresponding parameter vector that excludes σ_1^2, i.e., $\boldsymbol{\theta}^0=(\boldsymbol{\rho}'_N,\sigma_v^2)=(\rho_{1,N},...,\rho_{S,N},\sigma_v^2)$, and accordingly

$$\mathfrak{b}_N^0=(\rho_{1,N},...,\rho_{S,N},\rho_{1,N}^2,...,\rho_{S,N}^2,\rho_{1,N}\rho_{2,N},...,\rho_{1,N}\rho_{S,N},...,\rho_{S-1,N}\rho_{S,N},\sigma_v^2)'.$$

The initial GM estimator is then obtained as the solution to

$$(\hat{\rho}_{1,N},...,\hat{\rho}_{S,N},\hat{\sigma}_{v,N}^2)=\arg\min\{\vartheta_N^0(\underline{\boldsymbol{\theta}}_N^0)'\vartheta_N^0(\underline{\boldsymbol{\theta}}_N^0),-\mathbf{a}^\rho\le\underline{\boldsymbol{\rho}}\le\mathbf{a}^\rho,\underline{\sigma}_{v,N}^2\in[0,b_v]\}, \tag{17a}$$

with $\vartheta_N^0(\underline{\boldsymbol{\theta}}^0)=\vartheta_N^0(\underline{\boldsymbol{\rho}},\underline{\sigma}_v^2)=(\tilde{\boldsymbol{\gamma}}_N^0-\tilde{\boldsymbol{\Gamma}}_N^0\underline{\mathfrak{b}}^0)$.

Using these initial estimates of $(\rho_{1,N},...,\rho_{S,N})$ and σ_v^2, σ_1^2 can be estimated from moment condition M_b:

$$\hat{\sigma}_{1,N}^2=\frac{1}{N}(\tilde{\mathbf{u}}_N-\sum_{m=1}^S\hat{\rho}_{m,N}\tilde{\bar{\mathbf{u}}}_{m,N})'\mathbf{Q}_{1,N}(\tilde{\mathbf{u}}_N-\sum_{m=1}^S\hat{\rho}_{m,N}\tilde{\bar{\mathbf{u}}}_{m,N}) \tag{17b}$$

$$=\tilde{\gamma}_{4S+2}-\tilde{\gamma}_{4S+2,1}\hat{\rho}_{1,N}-...-\tilde{\gamma}_{4S+2,S}\hat{\rho}_{S,N}-\tilde{\gamma}_{4S+2,S+1}\hat{\rho}_{1,N}^2\cdots$$
$$-\tilde{\gamma}_{4S+2,2S}\hat{\rho}_{S,N}^2-\tilde{\gamma}_{4S+2,2S+1}\hat{\rho}_{1,N}\hat{\rho}_{2,N}-...-\tilde{\gamma}_{4S+2,2S+S(S-1)/2}\hat{\rho}_{S-1,N}\hat{\rho}_{S,N}.$$

2.2. Weighted GM Estimation

While the initial GM estimator as defined in (17) is consistent, it is inefficient. First, it ignores the information contained in moment conditions (M_b, $M_{3,s}$ and $M_{4,s}$). Second, as is known from the literature on GMM-estimation, it is optimal to use as weighting matrix the inverse of the (properly normalized) variance-covariance matrix of the moments, evaluated at true

[5] See Kapoor, Kelejian, and Prucha (2007) for analogous conditions under SARAR(0,1) estimation, assuming only nonstochastic regressors in equation (1a).

parameter values. Denote the optimal weighting matrix, which will be derived in Subsection 3.2, by $\boldsymbol{\Psi}_N^{-1}$ and its estimate by $\tilde{\boldsymbol{\Psi}}_N^{-1}$. The optimally weighted GM estimator is based on all $(4S+2)$ moment conditions and uses $\tilde{\boldsymbol{\Theta}}_N = \tilde{\boldsymbol{\Psi}}_N^{-1}$ as the weighting matrix. It is defined as

$$(\tilde{\rho}_{1,N},...,\tilde{\rho}_{S,N},\tilde{\sigma}_{v,N}^2,\tilde{\sigma}_{1,N}^2) = \arg\min\{\vartheta_N(\underline{\boldsymbol{\theta}})'\tilde{\boldsymbol{\Theta}}_N\vartheta_N(\underline{\boldsymbol{\theta}}), -\mathbf{a}^\rho \le \underline{\boldsymbol{\rho}} \le \mathbf{a}^\rho, \underline{\sigma}_v^2 \in [0,b_v], \underline{\sigma}_1^2 \in [0,c]\},$$

$$\text{with } c \ge b_v + Tb_\mu \text{, and } \vartheta_N(\underline{\boldsymbol{\theta}}) = \vartheta_N(\underline{\boldsymbol{\rho}},\underline{\sigma}_v^2,\underline{\sigma}_1^2) = (\tilde{\boldsymbol{\gamma}}_N - \tilde{\boldsymbol{\Gamma}}_N\underline{\mathbf{b}}). \tag{18}$$

As already mentioned, the optimal weighting matrix is derived without distributional assumptions and involves third and fourth moments of the error components $v_{it,N}$ and $\mu_{i,N}$. Kapoor, Kelejian, and Prucha (2008) use the assumption that $\varepsilon_{it,N}$ is normally distributed to obtain a simplified weighting matrix as an approximation of the true optimal weighting matrix. For comparison, we also consider such a weighting matrix, which is a special case of $\boldsymbol{\Psi}_N^{-1}$ (see the Appendix) and referred to as $(\boldsymbol{\Psi}_N^\circ)^{-1}$. The simplified weighted GM estimator is defined as the weighted GM estimator given in (18), using $\tilde{\boldsymbol{\Theta}}_N = (\tilde{\boldsymbol{\Psi}}_N^\circ)^{-1}$.

3. Asymptotic Properties of the GM Estimator for $\boldsymbol{\theta}_N$

3.1 Consistency

For proving consistency, the following additional assumptions are introduced:

Assumption 4.

Assume that $\tilde{\mathbf{u}}_N - \mathbf{u}_N = \mathbf{D}_N\boldsymbol{\Delta}_N$, i.e., $\tilde{u}_{i,N} - u_{i,N} = \mathbf{d}_{i.,N}\boldsymbol{\Delta}_N$, for $i = 1,...,NT$,[6] where $\mathbf{D}_N$ is an $NT \times P$ matrix, the $1 \times P$ vector $\mathbf{d}_{i.,N}$ denotes the *i*-th row of $\mathbf{D}_N$ and $\boldsymbol{\Delta}_N$ is a $P \times 1$ vector. Let $d_{ij,N}$ be the *j*-th element of $\mathbf{d}_{i.,N}$. For some $\delta > 0$, we assume that $E|d_{ij,N}(t)|^{2+\delta} \le c_d < \infty$, where c_d does not depend on *N*, and that $N^{1/2}\|\boldsymbol{\Delta}_N\| = O_p(1)$.

Assumption 4 will hold in many settings, e.g., if model (1a) contains endogenous variables (such as spatial lags of $\mathbf{y}_N$) and is estimated using 2SLS. In that case, $\boldsymbol{\Delta}_N$ denotes the difference between the parameter estimates and the true parameter values and $\mathbf{d}_{i.,N}$ is the (negative of the) *i*-th row of the design matrix $\mathbf{Z}_N$ (compare Lemma 1 in Subsection 2 of Section IV).

Assumption 5.

---

[6] Note that we use single indexation $i = 1,...,NT$ to refer to the elements of the vectors that are stacked over time periods.

(a) The smallest eigenvalues of $\mathbf{\Gamma}_N^{0\prime}\mathbf{\Gamma}_N^0$ and $\mathbf{\Gamma}_N^{1\prime}\mathbf{\Gamma}_N^1$ are bounded away from zero, i.e., $\lambda_{\min}(\mathbf{\Gamma}_N^{i\prime}\mathbf{\Gamma}_N^i) \geq \lambda_* > 0$ for i = 1, 2. (b) $\tilde{\mathbf{\Theta}}_N - \mathbf{\Theta}_N = o_p(1)$, where $\mathbf{\Theta}_N$ are $(4S+2)\times(4S+2)$ nonstochastic, symmetric, positive definite matrices. (c) The largest eigenvalues of $\mathbf{\Theta}_N$ are bounded uniformly from above, and the smallest eigenvalues of $\mathbf{\Theta}_N$ are bounded uniformly away from zero.

Assumption 5 implies that the smallest eigenvalues of $\mathbf{\Gamma}_N'\mathbf{\Gamma}_N$ and $\mathbf{\Gamma}_N'\mathbf{\Theta}_N\mathbf{\Gamma}_N$ are bounded uniformly away from zero, ensuring that the true parameter vector $\mathbf{\theta}_N$ is identifiable unique. Moreover, by the equivalence of matrix norms, it follows from Assumption 5 that $\mathbf{\Theta}_N$ and $\mathbf{\Theta}_N^{-1}$ are $O(1)$.

Assumptions 1-5 ensure consistency of the GM estimators for $\mathbf{\theta}_N = (\mathbf{\rho}_N', \sigma_v^2, \sigma_1^2)$ as summarized in the following theorems (see Appendix B for a proof).

Theorem 1a. Consistency of Initial GM Estimator $\tilde{\mathbf{\theta}}_N^0$

Suppose Assumptions 1-5 hold. Then, provided the optimization space contains the parameter space, the initial GM estimators $\hat{\mathbf{\theta}}_N^0 = (\hat{\rho}_{1,N},...,\hat{\rho}_S, \hat{\sigma}_{v,N}^2)'$ defined by (17a), and $\hat{\sigma}_{1,N}^2$, defined by (17b) are consistent for $\rho_{1,N},...,\rho_{S,N}$, σ_v^2, and σ_1^2, i.e.,

$$\hat{\rho}_{s,N} - \rho_{s,N} \xrightarrow{p} 0,\ s = 1,...,S,\ \ \hat{\sigma}_{v,N}^2 - \sigma_v^2 \xrightarrow{p} 0\ \text{, and } \hat{\sigma}_{1,N}^2 - \sigma_1^2 \xrightarrow{p} 0 \text{ as } N \to \infty.$$

Theorem 1b. Consistency of Weighted GM Estimator $\tilde{\mathbf{\theta}}_N$

Suppose Assumptions 1-5 hold. Then, provided the optimization space contains the parameter space, the weighted GM estimators $\tilde{\mathbf{\theta}}_N(\tilde{\mathbf{\Theta}}_N) = [\tilde{\rho}_{1,N}(\tilde{\mathbf{\Theta}}_N),...,\tilde{\rho}_{S,N}(\tilde{\mathbf{\Theta}}_N), \tilde{\sigma}_{v,N}^2(\tilde{\mathbf{\Theta}}_N), \tilde{\sigma}_{1,N}^2(\tilde{\mathbf{\Theta}}_N)]'$ defined by (18) are consistent for $\rho_{1,N},...,\rho_{S,N}, \sigma_v^2$, and σ_1^2, i.e.,

$$\tilde{\rho}_{s,N}(\tilde{\mathbf{\Theta}}_N) - \rho_{s,N} \xrightarrow{p} 0,\ s = 1,...,S,\ \tilde{\sigma}_{v,N}^2(\tilde{\mathbf{\Theta}}_N) - \sigma_v^2 \xrightarrow{p} 0\ \text{, and } \tilde{\sigma}_{1,N}^2(\tilde{\mathbf{\Theta}}_N) - \sigma_1^2 \xrightarrow{p} 0 \text{ as } N \to \infty.$$

This result holds for an arbitrary weighting matrix (satisfying Assumption 5). Hence, it applies to both the optimally weighted GM estimator defined by (18) with $\tilde{\mathbf{\Theta}}_N = (\tilde{\mathbf{\Psi}}_N)^{-1}$ and its simplified variant $\tilde{\mathbf{\theta}}_N^\circ$ with $\tilde{\mathbf{\Theta}}_N = (\tilde{\mathbf{\Psi}}_N^\circ)^{-1}$.

3.2 Asymptotic Distribution of GM Estimator for $\mathbf{\theta}_N$

In the following we consider the asymptotic distribution of the optimally weighted GM estimator $\tilde{\mathbf{\theta}}_N$. To establish asymptotic normality of $\tilde{\mathbf{\theta}}_N = (\tilde{\mathbf{\rho}}_N, \tilde{\sigma}_{v,N}^2, \tilde{\sigma}_{1,N}^2)$, we introduce some additional assumptions.

Assumption 6.

Let $\mathbf{D}_N$ be defined as in Assumption 4, such that $\tilde{\mathbf{u}}_N - \mathbf{u}_N = \mathbf{D}_N \boldsymbol{\Delta}_N$. For any real $NT \times NT$ 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 bounded uniformly in absolute value and the row and column sums of $\boldsymbol{\Pi}_N$ are bounded uniformly in absolute value (see Kelejian and Prucha, 2010, Lemma C.2). This will be the case in many applications, e.g., for the model in Eq. (1a), if $\mathbf{D}_N$ equals (the negative of) matrix $\mathbf{Z}_N$ (compare Lemma 1 in Section IV).

Assumption 7.

Let $\boldsymbol{\Delta}_N$ be defined as in Assumption 4. Then,

$$(NT)^{1/2}\boldsymbol{\Delta}_N = (NT)^{-1/2}\mathbf{T}'_N \boldsymbol{\xi}_N + o_p(1) \text{, with } \mathbf{T}_N = (\mathbf{T}'_{v,N}, \mathbf{T}'_{\mu,N})' \text{, } \boldsymbol{\xi}_N = (\mathbf{v}'_N, \boldsymbol{\mu}'_N)' \text{, i.e.,}$$
$$(NT)^{1/2}\boldsymbol{\Delta}_N = (NT)^{-1/2}\mathbf{T}'_{v,N}\mathbf{v}_N + (NT)^{-1/2}\mathbf{T}'_{\mu,N}\boldsymbol{\mu}_N + o_p(1),$$

where $\mathbf{T}_N$ is an $(NT+N)\times P$-dimensional real nonstochastic matrix whose elements are bounded uniformly in absolute value; $\mathbf{T}_{v,N}$ is of dimension $(NT \times P)$ and $\mathbf{T}_{\mu,N}$ is of dimension $(N \times P)$. As remarked above, $\boldsymbol{\Delta}_N$ 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 it holds if the model in Eq. (1a) is estimated by TSLS or feasible generalized TSLS (FGTSLS).

The limiting distribution of the GM estimator of $\boldsymbol{\theta}_N$ will be shown to depend on (the inverse of) the matrix $\mathbf{J}'_N \boldsymbol{\Theta}_N \mathbf{J}_N$ and the variance-covariance matrix of a vector of quadratic forms in $\mathbf{v}_N$ and $\boldsymbol{\mu}_N$, denoted as $\mathfrak{q}_N$. We consider each of these expressions in the following. The $(4S+2)\times(S+2)$ matrix $\mathbf{J}_N$ of derivatives of the $(4S+2)\times 1$ vector of moment conditions in (11) is given by

$$\mathbf{J}_N(\boldsymbol{\theta}_N) = \frac{\partial(\boldsymbol{\gamma}_N - \boldsymbol{\Gamma}_N \mathfrak{b}_N)}{\partial \boldsymbol{\theta}'} = (j_{i,1,N}, \ldots, j_{i,S,N}, j_{i,\sigma_v,N}, j_{i,\sigma_1,N}) \text{, with} \tag{19a}$$
$$j_{i,s,N} = \frac{\partial(\boldsymbol{\gamma}_{i.,N} - \boldsymbol{\Gamma}_{i.,N}\mathfrak{b}_N)}{\partial \rho_s} \text{, } i = 1, \ldots, (4S+2) \text{, } s = 1, \ldots, S \text{,}$$

$$j_{i,\sigma_v,N} = \frac{\partial(\boldsymbol{\gamma}_{i.,N} - \boldsymbol{\Gamma}_{i.,N}\mathfrak{b}_N)}{\partial \sigma_v},\ i = 1,...,(4S+2),$$

$$j_{i,\sigma_1,N} = \frac{\partial(\boldsymbol{\gamma}_{i.,N} - \boldsymbol{\Gamma}_{i.,N}\mathfrak{b}_N)}{\partial \sigma_1},\ i = 1,...,(4S+2),$$

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

Using $\frac{\partial \boldsymbol{\gamma}_N}{\partial \boldsymbol{\theta}'} = \mathbf{0}$ and ignoring the negative sign, we have

$$\mathbf{J}_N(\boldsymbol{\rho}) = \frac{\partial}{\partial \boldsymbol{\theta}'}\boldsymbol{\Gamma}_N \mathfrak{b}_N = \boldsymbol{\Gamma}_N \mathfrak{B}_N, \tag{19b}$$

where $\boldsymbol{\Gamma}_N$ is defined above and of dimension $(4S+2)\times[2S+S(S-1)/2+2]$ and $\mathfrak{B}_N$ is a $[2S+S(S-1)/2+2]\times(S+2)$ matrix of the form

$$\mathfrak{B}_N = (\mathfrak{B}_1, \mathfrak{B}'_{2,N}, \mathfrak{B}'_{3,N}, \mathfrak{B}'_{4,N})', \tag{20a}$$

with $\mathfrak{B}_1 = (\mathbf{I}_S, \mathbf{0}_{S\times 2})$, $\mathfrak{B}_{2,N} = [diag_{s=1}^{S}(2\rho_{s,N}), \mathbf{0}_{S\times 2})]$, and $\mathfrak{B}_{3,N} = [(\mathfrak{B}'_{3,1,N},...,\mathfrak{B}'_{3,S-1,N})', \mathbf{0}_{S(S-1)/2\times 2}]$ is an $S(S-1)/2\times(S+2)$ matrix. The $(S-1)$ vertically arranged blocks, $\mathfrak{B}_{3,m,N}$, $m = 1,...,(S-1)$, have the following structure:

$$\mathfrak{B}_{3,m,N} = (\mathfrak{C}_{m,N}, \mathfrak{d}_{m,N}, \mathfrak{E}_{m,N}), \tag{20b}$$

where $\mathfrak{C}_{m,N}$ is a $(S-m)\times(m-1)$ matrix of zeros,[7] $\mathfrak{d}_{m,N}$ is a $(S-m)\times 1$ vector, defined as $\mathfrak{d}_{m,N} = (\rho_{m+1,N},...,\rho_{S,N})'$, and $\mathfrak{E}_{m,N} = \rho_{m,N}\mathbf{I}_{S-m}$. Finally, $\mathfrak{B}_{4,N}$ is a $2\times(S+2)$ matrix, defined as

$$\mathfrak{B}_{4,N} = \begin{bmatrix} \mathbf{0}_{1\times S}, 1, 0 \\ \mathbf{0}_{1\times S+1}, 1 \end{bmatrix}. \tag{20c}$$

We next consider the vector $\mathfrak{q}_N$ and its limiting distribution. First, define $\mathbf{q}_N(\boldsymbol{\theta}_N, \boldsymbol{\Delta}_N)$ as the $(4S+2)\times 1$ vector of sample moments as given by (11) with the expectation operator suppressed, evaluated at the true parameter values, and ignoring the deterministic constants:

[7] I.e., there is no block $\mathfrak{C}_{1,N}$ in $\mathfrak{B}_{3,1,N}$.

$$\mathbf{q}_N(\boldsymbol{\theta}_N, \boldsymbol{\Delta}_N) = N^{-1} \begin{bmatrix} \tilde{\mathbf{u}}_N' \mathbf{C}_{1,1,N} \tilde{\mathbf{u}}_N \\ \tilde{\mathbf{u}}_N' \mathbf{C}_{2,1,N} \tilde{\mathbf{u}}_N \\ \tilde{\mathbf{u}}_N' \mathbf{C}_{3,1,N} \tilde{\mathbf{u}}_N \\ \tilde{\mathbf{u}}_N' \mathbf{C}_{4,1,N} \tilde{\mathbf{u}}_N \\ . \\ \tilde{\mathbf{u}}_N' \mathbf{C}_{1,S,N} \tilde{\mathbf{u}}_N \\ \tilde{\mathbf{u}}_N' \mathbf{C}_{2,S,N} \tilde{\mathbf{u}}_N \\ \tilde{\mathbf{u}}_N' \mathbf{C}_{3,S,N} \tilde{\mathbf{u}}_N \\ \tilde{\mathbf{u}}_N' \mathbf{C}_{4,S,N} \tilde{\mathbf{u}}_N \\ \tilde{\mathbf{u}}_N' \mathbf{C}_{a,N} \tilde{\mathbf{u}}_N \\ \tilde{\mathbf{u}}_N' \mathbf{C}_{b,N} \tilde{\mathbf{u}}_N \end{bmatrix}, \tag{21}$$

where

$$\begin{aligned}
\mathbf{C}_{1,s,N} &= \frac{1}{(T-1)}[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N} \mathbf{M}_{m,N}')]\mathbf{Q}_{0,N}(\mathbf{I}_T \otimes \mathbf{M}_{s,N}' \mathbf{M}_{s,N})\mathbf{Q}_{0,N}[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N} \mathbf{M}_{m,N})], \\
\mathbf{C}_{2,s,N} &= \frac{1}{2(T-1)}[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N} \mathbf{M}_{m,N}')]\mathbf{Q}_{0,N}[\mathbf{I}_T \otimes (\mathbf{M}_{s,N} + \mathbf{M}_{s,N}')]\mathbf{Q}_{0,N}[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N} \mathbf{M}_{m,N})], \\
\mathbf{C}_{3,s,N} &= [\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N} \mathbf{M}_{m,N}')]\mathbf{Q}_{1,N}(\mathbf{I}_T \otimes \mathbf{M}_{s,N}' \mathbf{M}_{s,N})\mathbf{Q}_{1,N}[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N} \mathbf{M}_{m,N})], \\
\mathbf{C}_{4,s,N} &= \frac{1}{2}[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N} \mathbf{M}_{m,N}')]\mathbf{Q}_{1,N}[\mathbf{I}_T \otimes (\mathbf{M}_{s,N} + \mathbf{M}_{s,N}')]\mathbf{Q}_{1,N}[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N} \mathbf{M}_{m,N})], \\
\mathbf{C}_{a,N} &= \frac{1}{(T-1)}[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N} \mathbf{M}_{m,N}')]\mathbf{Q}_{0,N}[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N} \mathbf{M}_{m,N})], \\
\mathbf{C}_{b,N} &= [\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N} \mathbf{M}_{m,N}')]\mathbf{Q}_{1,N}[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N} \mathbf{M}_{m,N})].
\end{aligned} \tag{22}$$

By Assumption 3 and Remark A.1 in Appendix A, the row and column sums of the symmetric $NT \times NT$ matrices $\mathbf{C}_{p,s,N}$, $p = 1,...,4$, $s = 1,...,S$, $\mathbf{C}_{a,N}$, and $\mathbf{C}_{b,N}$ are bounded uniformly in absolute value. Also, note that $\mathbf{C}_{3,s,N}$, $\mathbf{C}_{4,s,N}$, $\mathbf{C}_{b,N}$ differ from $\mathbf{C}_{1,s,N}$, $\mathbf{C}_{2,s,N}$, $\mathbf{C}_{a,N}$ only by the normalization and the use of $\mathbf{Q}_{1,N}$ versus $\mathbf{Q}_{0,N}$.

In light of (21) and Lemma B.1 (see Appendix B), the elements of $N^{1/2}\mathbf{q}_N(\boldsymbol{\rho}_N, \boldsymbol{\Delta}_N)$ can be expressed as

$$N^{1/2}\mathbf{q}_N(\boldsymbol{\theta}_N,\boldsymbol{\Delta}_N)=\begin{bmatrix} N^{-1/2}\mathbf{u}'_N\mathbf{C}_{1,1,N}\mathbf{u}_N+\boldsymbol{\alpha}'_{1,1,N}N^{1/2}\boldsymbol{\Delta}_N \\ N^{-1/2}\mathbf{u}'_N\mathbf{C}_{2,1,N}\mathbf{u}_N+\boldsymbol{\alpha}'_{2,1,N}N^{1/2}\boldsymbol{\Delta}_N \\ N^{-1/2}\mathbf{u}'_N\mathbf{C}_{3,1,N}\mathbf{u}_N+\boldsymbol{\alpha}'_{3,1,N}N^{1/2}\boldsymbol{\Delta}_N \\ N^{-1/2}\mathbf{u}'_N\mathbf{C}_{4,1,N}\mathbf{u}_N+\boldsymbol{\alpha}'_{4,1,N}N^{1/2}\boldsymbol{\Delta}_N \\ . \\ N^{-1/2}\mathbf{u}'_N\mathbf{C}_{1,S,N}\mathbf{u}_N+\boldsymbol{\alpha}'_{1,S,N}N^{1/2}\boldsymbol{\Delta}_N \\ N^{-1/2}\mathbf{u}'_N\mathbf{C}_{2,S,N}\mathbf{u}_N+\boldsymbol{\alpha}'_{2,S,N}N^{1/2}\boldsymbol{\Delta}_N \\ N^{-1/2}\mathbf{u}'_N\mathbf{C}_{3,S,N}\mathbf{u}_N+\boldsymbol{\alpha}'_{3,S,N}N^{1/2}\boldsymbol{\Delta}_N \\ N^{-1/2}\mathbf{u}'_N\mathbf{C}_{4,S,N}\mathbf{u}_N+\boldsymbol{\alpha}'_{4,S,N}N^{1/2}\boldsymbol{\Delta}_N \\ N^{-1/2}\mathbf{u}'_N\mathbf{C}_{a,N}\mathbf{u}_N+\boldsymbol{\alpha}'_{a,N}N^{1/2}\boldsymbol{\Delta}_N \\ N^{-1/2}\mathbf{u}'_N\mathbf{C}_{a,N}\mathbf{u}_N+\boldsymbol{\alpha}'_{b,N}N^{1/2}\boldsymbol{\Delta}_N \end{bmatrix}+o_p(1), \qquad (23)$$

where $\boldsymbol{\alpha}_{p,s,N}=2N^{-1}E(\mathbf{D}'_N\mathbf{C}_{p,s,N}\mathbf{u}_N)$, $p=1,\dots,4$, $s=1,\dots,S$, $\boldsymbol{\alpha}_{a,N}=2N^{-1}E(\mathbf{D}'_N\mathbf{C}_{a,N}\mathbf{u}_N)$, and $\boldsymbol{\alpha}_{b,N}=2N^{-1}E(\mathbf{D}'_N\mathbf{C}_{b,N}\mathbf{u}_N)$. By Lemma B.1 the elements of the $P\times 1$ vectors $\boldsymbol{\alpha}_{p,s,N}$, $p=1,\dots,4$, $s=1,\dots,S$, $\boldsymbol{\alpha}_{a,N}$ and $\boldsymbol{\alpha}_{b,N}$ are bounded uniformly in absolute value.

Using (22), (3c), Assumption 7, and $\mathbf{Q}_{0,N}\boldsymbol{\varepsilon}_N=\mathbf{Q}_{0,N}\mathbf{v}_N$ we obtain:

$$N^{1/2}\mathbf{q}_N(\boldsymbol{\theta}_N,\boldsymbol{\Delta}_N)=N^{-1/2}\begin{bmatrix} \frac{1}{(T-1)}\mathbf{v}'_N\mathbf{Q}_{0,N}(\mathbf{I}_T\otimes\mathbf{M}'_{s,N}\mathbf{M}_{s,N})\mathbf{Q}_{0,N}\mathbf{v}_N+\mathbf{a}'_{1,s,N}\boldsymbol{\xi}_N \\ \frac{1}{2(T-1)}\mathbf{v}'_N\mathbf{Q}_{0,N}[\mathbf{I}_T\otimes(\mathbf{M}_{s,N}+\mathbf{M}'_{s,N})]\mathbf{Q}_{0,N}\mathbf{v}_N+\mathbf{a}'_{2,s,N}\boldsymbol{\xi}_N \\ \boldsymbol{\varepsilon}'_N\mathbf{Q}_{1,N}(\mathbf{I}_T\otimes\mathbf{M}'_{s,N}\mathbf{M}_{s,N})\mathbf{Q}_{1,N}\boldsymbol{\varepsilon}_N+\mathbf{a}'_{3,s,N}\boldsymbol{\xi}_N \\ \frac{1}{2}\boldsymbol{\varepsilon}'_N\mathbf{Q}_{1,N}[\mathbf{I}_T\otimes(\mathbf{M}_{s,N}+\mathbf{M}'_{s,N})]\mathbf{Q}_{1,N}\boldsymbol{\varepsilon}_N+\mathbf{a}'_{4,s,N}\boldsymbol{\xi}_N \\ . \\ \frac{1}{(T-1)}\mathbf{v}'_N\mathbf{Q}_{0,N}\mathbf{v}_N+\mathbf{a}'_{a,N}\boldsymbol{\xi}_N \\ \boldsymbol{\varepsilon}'_N\mathbf{Q}_{1,N}\boldsymbol{\varepsilon}_N+\mathbf{a}'_{b,N}\boldsymbol{\xi}_N \end{bmatrix}+o_p(1), \qquad (25)$$

for $s=1,\dots,S$. The $(NT+N)\times 1$ vector $\boldsymbol{\xi}_N=(\mathbf{v}'_N,\boldsymbol{\mu}'_N)'$, $\mathbf{a}_{p,s,N}=T^{-1}\mathbf{T}_N\boldsymbol{\alpha}_{p,s,N}$, $p=1,\dots,4$, $s=1,\dots,S$, $\mathbf{a}_{a,N}=T^{-1}\mathbf{T}_N\boldsymbol{\alpha}_{a,N}$, and $\mathbf{a}_{b,N}=T^{-1}\mathbf{T}_N\boldsymbol{\alpha}_{b,N}$, which can also be written as

$\mathbf{a}_{p,s,N}=(\mathbf{a}^{\nu\,\prime}_{p,s,N},\mathbf{a}^{\mu\,\prime}_{p,s,N})'=T^{-1}[(\mathbf{T}_{\nu,N}\boldsymbol{\alpha}_{p,s,N})',(\mathbf{T}_{\mu,N}\boldsymbol{\alpha}_{p,s,N})']'$, $s=1,\dots,S$, $p=1,\dots,4$, and

$\mathbf{a}_{a,N}=(\mathbf{a}^{\nu\prime}_{a,N},\mathbf{a}^{\mu\prime}_{a,N})'=T^{-1}[(\mathbf{T}_{\nu,N}\boldsymbol{\alpha}_{a,N})',(\mathbf{T}_{\mu,N}\boldsymbol{\alpha}_{a,N})']'$,

$\mathbf{a}_{b,N}=(\mathbf{a}^{\nu\prime}_{b,N},\mathbf{a}^{\mu\prime}_{b,N})'=T^{-1}[(\mathbf{T}_{\nu,N}\boldsymbol{\alpha}_{b,N})',(\mathbf{T}_{\mu,N}\boldsymbol{\alpha}_{b,N})']'$.

Observe that the elements of $\mathbf{a}_{p,s,N}$, $p=1,...,4$, $s=1,...,S$, $\mathbf{a}_{a,N}$, and $\mathbf{a}_{b,N}$ are bounded uniformly in absolute value by Assumption 7 and Lemma B.1. Utilizing

(26)

$$\boldsymbol{\varepsilon}'_N\mathbf{Q}_{1,N}(\mathbf{I}_T\otimes\mathbf{M}'_{s,N}\mathbf{M}_{s,N})\mathbf{Q}_{1,N}\boldsymbol{\varepsilon}_N=\mathbf{v}'_N\mathbf{Q}_{1,N}[\mathbf{I}_T\otimes(\mathbf{M}'_{s,N}\mathbf{M}_{s,N})]\mathbf{Q}_{1,N}\mathbf{v}_N+ \\ +T\boldsymbol{\mu}'_N(\mathbf{M}'_{s,N}\mathbf{M}_{s,N})\boldsymbol{\mu}_N+2\mathbf{v}'_N[\mathbf{e}_T\otimes(\mathbf{M}'_{s,N}\mathbf{M}_{s,N})]\boldsymbol{\mu}_N.$$

$$\frac{1}{2}\boldsymbol{\varepsilon}'_N\mathbf{Q}_{1,N}[\mathbf{I}_T\otimes(\mathbf{M}_{s,N}+\mathbf{M}'_{s,N})]\mathbf{Q}_{1,N}\boldsymbol{\varepsilon}_N=\frac{1}{2}\{\mathbf{v}'_N\mathbf{Q}_{1,N}[\mathbf{I}_T\otimes(\mathbf{M}_{s,N}+\mathbf{M}'_{s,N})]\mathbf{Q}_{1,N}\mathbf{v}_N \\ +T\boldsymbol{\mu}'_N(\mathbf{M}_{s,N}+\mathbf{M}'_{s,N})\boldsymbol{\mu}_N+2\mathbf{v}'_N[\mathbf{e}_T\otimes(\mathbf{M}_{s,N}+\mathbf{M}'_{s,N})]\boldsymbol{\mu}_N\},$$

$$\boldsymbol{\varepsilon}'_N\mathbf{Q}_{1,N}\boldsymbol{\varepsilon}_N=\mathbf{v}'_N\mathbf{Q}_{1,N}\mathbf{v}_N+T\boldsymbol{\mu}'_N\boldsymbol{\mu}_N+2\mathbf{v}'_N(\mathbf{e}_T\otimes\mathbf{I}_N)\boldsymbol{\mu}_N,$$

we have

(27)

$$N^{1/2}\mathbf{q}_N(\boldsymbol{\theta}_N,\boldsymbol{\Delta}_N)=$$

$$N^{-1/2}\begin{bmatrix}\frac{1}{(T-1)}\mathbf{v}'_N\mathbf{Q}_{0,N}(\mathbf{I}_T\otimes\mathbf{M}'_{s,N}\mathbf{M}_{s,N})\mathbf{Q}_{0,N}\mathbf{v}_N+\mathbf{a}'_{1,s,N}\boldsymbol{\xi}_N\\ \frac{1}{2(T-1)}\mathbf{v}'_N\mathbf{Q}_{0,N}[\mathbf{I}_T\otimes(\mathbf{M}_{s,N}+\mathbf{M}'_{s,N})]\mathbf{Q}_{0,N}\mathbf{v}_N+\mathbf{a}'_{2,s,N}\boldsymbol{\xi}_N\\ \mathbf{v}'_N\mathbf{Q}_{1,N}(\mathbf{I}_T\otimes\mathbf{M}'_{s,N}\mathbf{M}_{s,N})\mathbf{Q}_{1,N}\mathbf{v}_N+T\boldsymbol{\mu}'_N\mathbf{M}'_{s,N}\mathbf{M}_{s,N}\boldsymbol{\mu}_N+2\mathbf{v}'_N[\mathbf{e}_T\otimes(\mathbf{M}'_{s,N}\mathbf{M}_{s,N})]\boldsymbol{\mu}_N+\mathbf{a}'_{3,s,N}\boldsymbol{\xi}_N\\ \frac{1}{2}\mathbf{v}'_N\mathbf{Q}_{1,N}[\mathbf{I}_T\otimes(\mathbf{M}_{s,N}+\mathbf{M}'_{s,N})]\mathbf{Q}_{1,N}\mathbf{v}_N+\frac{T}{2}\boldsymbol{\mu}'_N(\mathbf{M}_{s,N}+\mathbf{M}'_{s,N})\boldsymbol{\mu}_N+\mathbf{v}'_N[\mathbf{e}_T\otimes(\mathbf{M}_{s,N}+\mathbf{M}'_{s,N})]\boldsymbol{\mu}_N+\mathbf{a}'_{4,s,N}\boldsymbol{\xi}_N\\ .\\ \frac{1}{(T-1)}\mathbf{v}'_N\mathbf{Q}_{0,N}\mathbf{v}_N+\mathbf{a}'_{a,N}\boldsymbol{\xi}_N\\ \mathbf{v}'_N\mathbf{Q}_{1,N}\mathbf{v}_N+T\boldsymbol{\mu}'_N\boldsymbol{\mu}_N+2\mathbf{v}'_N(\mathbf{e}_T\otimes\mathbf{I}_N)\boldsymbol{\mu}_N+\mathbf{a}'_{b,N}\boldsymbol{\xi}_N\end{bmatrix}+o_p(1)$$

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

Next, consider the $(4S+2)\times 1$ vector

$$\mathbf{q}_N=N^{-1/2}\mathbf{q}^*_N=N^{-1/2}\begin{bmatrix}\mathbf{q}^*_{1,N}\\ .\\ \mathbf{q}^*_{S,N}\\ q^*_{a,N}\\ q^*_{b,N}\end{bmatrix}. \tag{28}$$

Each element $\mathbf{q}^*_{s,N}$, $s=1,...,S$, is a 4×1 vector, given by

$$\mathbf{q}^*_{s,N} = \begin{bmatrix} q^*_{1,s,N} \\ q^*_{2,s,N} \\ q^*_{3,s,N} \\ q^*_{4,s,N} \end{bmatrix}, \quad (29)$$

where $q^*_{p,s,N}$, $p = 1,...,4$, $s = 1,...,S$, $q^*_{a,N}$, and $q^*_{b,N}$ can be written as linear quadratic forms in the $(NT + N) \times 1$ vector $\xi_N = (\mathbf{v}'_N, \boldsymbol{\mu}'_N)'$, i.e., we have

$$q^*_{p,s,N} = \xi'_N \mathbf{A}_{p,s,N} \xi_N + \mathbf{a}'_{p,s,N} \xi_N, \; p = 1,...,4, \; s = 1,...,S, \quad (30)$$
$$q^*_{a,N} = \xi'_N \mathbf{A}_{a,N} \xi_N + \mathbf{a}'_{a,N} \xi_N, \text{ and}$$
$$q^*_{b,N} = \xi'_N \mathbf{A}_{b,N} \xi_N + \mathbf{a}'_{b,N} \xi_N.$$

We consider each of these terms in the following.

$$q^*_{1,s,N} = \xi'_N \mathbf{A}_{1,s,N} \xi_N + \mathbf{a}'_{1,s,N} \xi_N, \text{ where} \quad (31)$$
$$\mathbf{A}_{1,s,N} = \begin{bmatrix} \mathbf{A}^{v}_{1,s,N} & \mathbf{A}^{v,\mu}_{1,s,N} \\ (\mathbf{A}^{v,\mu}_{1,s,N})' & \mathbf{A}^{\mu}_{1,s,N} \end{bmatrix} = \begin{bmatrix} \frac{1}{(T-1)} \mathbf{Q}_{0,N} (\mathbf{I}_T \otimes \mathbf{M}'_{s,N} \mathbf{M}_{s,N}) \mathbf{Q}_{0,N} & \mathbf{0}_{NT \times N} \\ \mathbf{0}_{N \times NT} & \mathbf{0}_{N \times N} \end{bmatrix}, \text{ and}$$
$$\mathbf{a}'_{1,s,N} = (\mathbf{a}^{v\,\prime}_{1,s,N}, \mathbf{a}^{\mu\,\prime}_{1,s,N}),$$

and the $\mathbf{0}$ terms denote zero-matrices, whose dimensions are indicated by the subscript.

$$q^*_{2,s,N} = \xi'_N \mathbf{A}_{2,s,N} \xi_N + \mathbf{a}'_{2,s,N} \xi_N, \text{ where}$$
$$\mathbf{A}_{2,s,N} = \begin{bmatrix} \mathbf{A}^{v}_{2,s,N} & \mathbf{A}^{v,\mu}_{2,s,N} \\ (\mathbf{A}^{v,\mu}_{2,s,N})' & \mathbf{A}^{\mu}_{2,s,N} \end{bmatrix} = \begin{bmatrix} \frac{1}{2(T-1)} \mathbf{Q}_{0,N} [\mathbf{I}_T \otimes (\mathbf{M}_{s,N} + \mathbf{M}'_{s,N})] \mathbf{Q}_{0,N} & \mathbf{0}_{NT \times N} \\ \mathbf{0}_{N \times NT} & \mathbf{0}_{N \times N} \end{bmatrix}, \text{ and} \quad (32a)$$
$$\mathbf{a}'_{2,s,N} = (\mathbf{a}^{v\,\prime}_{2,s,N}, \mathbf{a}^{\mu\,\prime}_{2,s,N}).$$

$$q^*_{3,s,N} = \xi'_N \mathbf{A}_{3,s,N} \xi_N + \mathbf{a}'_{3,s,N} \xi_N, \text{ where}$$
$$\mathbf{A}_{3,s,N} = \begin{bmatrix} \mathbf{A}^{v}_{3,s,N} & \mathbf{A}^{v,\mu}_{3,s,N} \\ (\mathbf{A}^{v,\mu}_{3,s,N})' & \mathbf{A}^{\mu}_{3,s,N} \end{bmatrix} = \begin{bmatrix} \mathbf{Q}_{1,N} (\mathbf{I}_T \otimes \mathbf{M}'_{s,N} \mathbf{M}_{s,N}) \mathbf{Q}_{1,N} & [\mathbf{e}_T \otimes (\mathbf{M}'_{s,N} \mathbf{M}_{s,N})] \\ [\mathbf{e}'_T \otimes (\mathbf{M}'_{s,N} \mathbf{M}_{s,N})] & T \mathbf{M}'_{s,N} \mathbf{M}_{s,N} \end{bmatrix}, \text{ and} \quad (32b)$$
$$\mathbf{a}'_{3,s,N} = (\mathbf{a}^{v\,\prime}_{3,s,N}, \mathbf{a}^{\mu\,\prime}_{3,s,N}).$$
$$q^*_{4,s,N} = \xi'_N \mathbf{A}_{4,s,N} \xi_N + \mathbf{a}'_{4,s,N} \xi_N, \text{ where}$$

$$\mathbf{A}_{4,s,N} = \begin{bmatrix} \mathbf{A}^{v}_{4,s,N} & \mathbf{A}^{v,\mu}_{4,s,N} \\ (\mathbf{A}^{v,\mu}_{4,s,N})' & \mathbf{A}^{v}_{4,s,N} \end{bmatrix} = \begin{bmatrix} \frac{1}{2}\mathbf{Q}_{1,N}[\mathbf{I}_T \otimes (\mathbf{M}_{s,N} + \mathbf{M}'_{s,N})]\mathbf{Q}_{1,N} & \frac{1}{2}[\mathbf{e}_T \otimes (\mathbf{M}_{s,N} + \mathbf{M}'_{s,N})] \\ \frac{1}{2}[\mathbf{e}'_T \otimes (\mathbf{M}_{s,N} + \mathbf{M}'_{s,N})] & \frac{T}{2}(\mathbf{M}_{s,N} + \mathbf{M}'_{s,N}) \end{bmatrix}, \text{ and (32c)}$$

$$\mathbf{a}'_{4,s,N} = (\mathbf{a}^{v\,\prime}_{4,s,N}, \mathbf{a}^{\mu\,\prime}_{4,s,N}).$$

$q^*_{a,N} = \boldsymbol{\xi}'_N \mathbf{A}_{a,N} \boldsymbol{\xi}_N + \mathbf{a}'_{a,N}\boldsymbol{\xi}_N$, where

$$\mathbf{A}_{a,N} = \begin{bmatrix} \mathbf{A}^{v}_{a,N} & \mathbf{A}^{v,\mu}_{a,N} \\ (\mathbf{A}^{v,\mu}_{a,N})' & \mathbf{A}^{\mu}_{a,N} \end{bmatrix} = \begin{bmatrix} \frac{1}{(T-1)}\mathbf{Q}_{0,N} & \mathbf{0}_{NT\times N} \\ \mathbf{0}_{N\times NT} & \mathbf{0}_{N\times N} \end{bmatrix}, \text{ and} \tag{32e}$$

$$\mathbf{a}'_{a,N} = (\mathbf{a}^{v\,\prime}_{a,N}, \mathbf{a}^{\mu\,\prime}_{a,N}).$$

$q^*_{b,N} = \boldsymbol{\xi}'_N \mathbf{A}_{b,N} \boldsymbol{\xi}_N + \mathbf{a}'_{b,N}\boldsymbol{\xi}_N$, where

$$\mathbf{A}_{b,N} = \begin{bmatrix} \mathbf{A}^{v}_{b,N} & \mathbf{A}^{v,\mu}_{b,N} \\ (\mathbf{A}^{v,\mu}_{b,N})' & \mathbf{A}^{\mu}_{b,N} \end{bmatrix} = \begin{bmatrix} \mathbf{Q}_{1,N} & (\mathbf{e}_T \otimes \mathbf{I}_N) \\ (\mathbf{e}'_T \otimes \mathbf{I}_N) & T\mathbf{I}_N \end{bmatrix}, \text{ and} \tag{32f}$$

$$\mathbf{a}'_{b,N} = (\mathbf{a}^{v\,\prime}_{b,N}, \mathbf{a}^{\mu\,\prime}_{b,N}).$$

Note that the row and column sums of the symmetric $(NT+N)\times(NT+N)$ matrices $\mathbf{A}_{1,s,N},...,\mathbf{A}_{4,s,N}$, $s=1,...,S$, $\mathbf{A}_{a,N}$, and $\mathbf{A}_{b,N}$, are bounded uniformly in absolute value by Assumption 3 and Remark A.1. Moreover, the elements of the $\boldsymbol{\xi}_N = (\mathbf{v}'_N, \boldsymbol{\mu}'_N)'$ are independently distributed by Assumption 1. Hence, the variance-covariance matrix of $\boldsymbol{\xi}_N$ is

$$\boldsymbol{\Omega}_{\xi,N} = \begin{bmatrix} \sigma^2_v \mathbf{I}_{NT} & \mathbf{0}_{NT\times N} \\ \mathbf{0}_{N\times NT} & \sigma^2_\mu \mathbf{I}_N \end{bmatrix}. \tag{33}$$

In order to calculate the variance-covariance matrix of $\mathfrak{q}_N$, denoted as $\boldsymbol{\Psi}_N$, we invoke Lemma A.1 in Kelejian and Prucha (2010). It is given by $\boldsymbol{\Psi}_N = N^{-1}E(\mathfrak{q}^*_N \mathfrak{q}^{*\prime}_N)$, which is a symmetric $(4S+2)\times(4S+2)$ matrix, and takes the following form:

$$\boldsymbol{\Psi}_N = (\mathfrak{E}_{r,s,N}),\ r,s = 1,...,S+1, \text{ i.e.,} \tag{34a}$$

$$\boldsymbol{\Psi}_N = \begin{bmatrix} \mathfrak{E}_{1,1,N} & . & \mathfrak{E}_{1,S,N} & \mathfrak{E}_{1,S+1,N} \\ . & & & . \\ \mathfrak{E}_{S,1,N} & & \mathfrak{E}_{S,S,N} & \mathfrak{E}_{S,S+1,N} \\ \mathfrak{E}_{S+1,1,N} & & \mathfrak{E}_{S+1,S,N} & \mathfrak{E}_{S+1,S+1,N} \end{bmatrix}. \tag{34b}$$

Observe that the matrix $\boldsymbol{\Psi}_N$ contains three parts.

i) The upper left block is of dimension $4S \times 4S$, consisting of S^2 blocks of dimension 4 × 4, which are defined as

$$\mathfrak{E}_{r,s,N} = N^{-1} E(\mathfrak{q}^*_{r,N} \mathfrak{q}^{*\prime}_{s,N}) = (\mathfrak{E}^{p,q}_{r,s}), \ r,s = 1,...,S, \ p,q = 1,...,4. \quad (34c)$$

The elements $\mathfrak{E}^{p,q}_{r,s,N}$, $p,q = 1,...,4$, $r,s = 1,...,S$ are defined as

$$\begin{aligned}
\mathfrak{E}^{p,q}_{r,s,N} &= N^{-1} Cov(q^*_{p,r,N}, q^*_{q,s,N}) \quad (34d) \\
&= 2\sigma_v^4 N^{-1} Tr(\mathbf{A}^v_{p,r,N} \mathbf{A}^v_{q,s,N}) + 2\sigma_\mu^4 N^{-1} Tr(\mathbf{A}^\mu_{p,r,N} \mathbf{A}^\mu_{q,s,N}) + 4\sigma_\mu^2 \sigma_v^2 N^{-1} Tr[(\mathbf{A}^{v,\mu}_{p,r,N})' \mathbf{A}^{v,\mu}_{q,s,N}] \\
&+ \sigma_v^2 N^{-1} \mathbf{a}^{v\prime}_{p,r,N} \mathbf{a}^v_{q,s,N} + \sigma^2_{\mu,N} N^{-1} \mathbf{a}^{\mu\prime}_{p,r,N} \mathbf{a}^\mu_{q,s,N} \\
&+ (\sigma_v^{(4)} - 3\sigma_v^4) N^{-1} \sum_{i=1}^{NT} a^v_{p,r,ii,N} a^v_{q,s,ii,N} + (\sigma_\mu^{(4)} - 3\sigma_\mu^4) N^{-1} \sum_{i=1}^{N} a^\mu_{p,r,ii,N} a^\mu_{q,s,ii,N} \\
&+ \sigma_v^{(3)} N^{-1} \sum_{i=1}^{NT} (a^v_{p,r,i,N} a^v_{q,s,ii,N} + a^v_{p,r,ii,N} a^v_{q,s,i,N}) + \sigma_\mu^{(3)} N^{-1} \sum_{i=1}^{N} (a^\mu_{p,r,i,N} a^\mu_{q,s,ii,N} + a^\mu_{p,r,ii,N} a^\mu_{q,s,i,N}),
\end{aligned}$$

where $a^v_{p,r,ii,N}$ and $a^\mu_{p,r,ii,N}$ denote the *i*-th main diagonal element of the matrices $\mathbf{A}^v_{p,r,N}$ and $\mathbf{A}^\mu_{p,r,N}$, respectively, and $a^v_{p,r,i,N}$ and $a^v_{p,r,i,N}$ denote the *i*-th element of the vectors $\mathbf{a}^v_{p,r,N}$ and $\mathbf{a}^\mu_{p,r,N}$ respectively. The terms $\sigma_v^{(3)}$, $\sigma_\mu^{(3)}$ and $\sigma_v^{(4)}$, $\sigma_\mu^{(4)}$ denote the third and fourth moment of $v_{it,N}$ and $\mu_{it,N}$, respectively.

ii) The last two rows and columns are matrices of dimension $(2 \times 4S)$ and $(4S \times 2)$, respectively, each of which is made up by S blocks of dimension (2×4) (4×2), defined as

$$\mathfrak{E}_{S+1,s,N} = (\mathfrak{E}^{p,q}_{S+1,s,N}) = N^{-1} Cov(q^*_{p,N}, q^*_{q,s,N}), \ p = a,b, \ q = 1,...,4, \text{ and } s = 1,...,S, \quad (34e)$$

and $\mathfrak{E}_{s,S+1} = (\mathfrak{E}_{S+1,s})'$, $s = 1,...,S$. The elements $\mathfrak{E}^{p,q}_{S+1,s,N}$ are defined as in (34d), using the corresponding indexation.

iii) Finally, the lower right block of dimension 2 × 2, is defined as

$$\mathfrak{E}_{S+1,S+1,N} = (\mathfrak{E}^{p,q}_{S+1,S+1,N}) = N^{-1} Cov(q^*_{p,N}, q^*_{q,N}), \ p,q = a,b, \quad (34f)$$

where the elements $\mathfrak{E}^{p,q}_{S+1,s,N}$ are defined as in (34d), using the corresponding indexation.

The expression given by (34d) holds generally. Part of the elements of $\boldsymbol{\Psi}_N$ can be stated in simpler terms. E.g., due to the orthogonality of $\mathbf{Q}_{0,N}$ and $\mathbf{Q}_{1,N}$, the terms in the first line drop out when $\mathbf{Q}_{0,N}$ and $\mathbf{Q}_{1,N}$ meet in the trace expression. Moreover, if $\mathbf{v}_N$ and $\boldsymbol{\mu}_N$ are normally distributed, the terms involving the third and fourth moments of $\mathbf{v}_N$ and $\boldsymbol{\mu}_N$ drop out for all elements of $\boldsymbol{\Psi}_N$.

To derive the asymptotic distribution of $\mathfrak{q}_N$ and $\tilde{\boldsymbol{\theta}}_N$ we invoke the central limit theorem for vectors of linear quadratic forms given by Kelejian and Prucha (2010, Theorem A.1) and Corollary F4 in Pötscher and Prucha (1997). We summarize the results regarding the asymptotic distribution of $\tilde{\boldsymbol{\theta}}_N$ in the following Theorem, which is proved in Appendix B.

Theorem 2. (Asymptotic Normality of $\tilde{\boldsymbol{\theta}}_N$)

Let $\tilde{\boldsymbol{\theta}}_N$ be the GM estimator defined by (18). Suppose Assumptions 1-7 hold and, furthermore, that $\lambda_{\min}(\boldsymbol{\Psi}_N) \geq c^*_{\boldsymbol{\Psi}} > 0$. Then, provided the optimization space contains the parameter space, we have

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

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

$$\boldsymbol{\xi}_N = \boldsymbol{\Psi}_N^{-1/2} \mathfrak{q}_N \xrightarrow{d} N(0, \mathbf{I}_{4S+2}),$$

where $\boldsymbol{\Psi}_N = E(\mathfrak{q}_N \mathfrak{q}'_N)$ and $\boldsymbol{\Psi}_N = (\boldsymbol{\Psi}_N^{1/2})(\boldsymbol{\Psi}_N^{1/2})'$.

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

$$\boldsymbol{\Omega}_{\tilde{\boldsymbol{\theta}}_N}(\boldsymbol{\Theta}_N) = (\mathbf{J}'_N \boldsymbol{\Theta}_N \mathbf{J}_N)^{-1} \mathbf{J}'_N \boldsymbol{\Theta}_N \boldsymbol{\Psi}_N \boldsymbol{\Theta}_N \mathbf{J}_N (\mathbf{J}'_N \boldsymbol{\Theta}_N \mathbf{J}_N)^{-1},$$

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

Theorem 2 implies that the difference between the cumulative distribution function of $N^{1/2}(\tilde{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_N)$ and that of $N(0, \boldsymbol{\Omega}_{\tilde{\boldsymbol{\theta}}_N})$ converges pointwise to zero, which justifies the use of the latter as an approximation of the former.[8]

[8] Compare Corollary F4 in Pötscher and Prucha (1997).

Note that $\mathbf{\Omega}_{\tilde{\mathbf{\theta}}_N}(\mathbf{\Psi}_N^{-1}) = (\mathbf{J}_N'\mathbf{\Psi}_N^{-1}\mathbf{J}_N)^{-1}$ and that $\mathbf{\Omega}_{\tilde{\mathbf{\theta}}_N}(\mathbf{\Theta}_N) - \mathbf{\Omega}_{\tilde{\mathbf{\theta}}_N}(\mathbf{\Psi}_N^{-1})$ is positive semidefinite. Thus, using a consistent estimator of $\mathbf{\Psi}_N^{-1}$ (which will be derived below) as weighting matrix $\mathbf{\Theta}_N$ leads to the efficient GM estimator.

3.3 Estimation of the Variance-Covariance Matrix of $\tilde{\mathbf{\theta}}_N$

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

$$\tilde{\mathbf{J}}_N = \tilde{\mathbf{\Gamma}}_N \tilde{\mathfrak{B}}_N . \tag{35}$$

We next specify estimators for $\mathbf{a}_{p,s,N} = \mathbf{T}_N \boldsymbol{\alpha}_{p,s,N}$, $p = 1,...,4$, $s = 1,...,S$, $\mathbf{a}_{a,N} = \mathbf{T}_N \boldsymbol{\alpha}_{a,N}$, and $\mathbf{a}_{b,N} = \mathbf{T}_N \boldsymbol{\alpha}_{b,N}$. The matrix $\mathbf{T}_N$ will often be of the form

$$\mathbf{T}_N = \mathbf{F}_N \mathbf{P}_N \quad \text{with } \mathbf{F}_N = (\mathbf{F}_{v,N}', \mathbf{F}_{\mu,N}')', \tag{36a}$$

which can also be written as

$$\mathbf{T}_N = (\mathbf{T}_{v,N}', \mathbf{T}_{\mu,N}')' \text{ with } \mathbf{T}_{v,N} = \mathbf{F}_{v,N}\mathbf{P}_N,\ \mathbf{T}_{\mu,N} = \mathbf{F}_{\mu,N}\mathbf{P}_N,$$

and

$$\mathbf{F}_{v,N} = [\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N}\mathbf{M}_{m,N}')^{-1}]\mathbf{H}_N, \tag{36b}$$

$$\mathbf{F}_{\mu,N} = (\mathbf{e}_T' \otimes \mathbf{I}_N)[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N}\mathbf{M}_{m,N}')^{-1}]\mathbf{H}_N,$$

or, alternatively,

$$\mathbf{F}_{v,N} = \mathbf{\Omega}_{\boldsymbol{\varepsilon},N}^{-1}[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N}\mathbf{M}_{m,N})]\mathbf{H}_N, \tag{36c}$$

$$\mathbf{F}_{\mu,N} = [\sigma_1^{-2}(\mathbf{e}_T' \otimes \mathbf{I}_N)][\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N}\mathbf{M}_{m,N})]\mathbf{H}_N,$$

where $\mathbf{F}_{v,N}$ is a real nonstochastic $NT \times P_*$ matrix, $\mathbf{F}_{\mu,N}$ is a real nonstochastic $N \times P_*$ matrix, $\mathbf{H}_N$ is a real nonstochastic $NT \times P_*$ 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, when Eq. (1a) is estimated using two-stage least squares (TSLS), $\mathbf{\Delta}_N = (\tilde{\boldsymbol{\delta}}_N - \boldsymbol{\delta}_N)$ and the matrix $\mathbf{P}_N$ will be of the structure as defined above and can be estimated consistently by some estimator $\tilde{\mathbf{P}}_N$ (see Section IV).

The estimators for $\mathbf{T}_N$ are defined as

$$\tilde{\mathbf{T}}_{v,N} = \tilde{\mathbf{F}}_{v,N}\tilde{\mathbf{P}}_N, \quad \tilde{\mathbf{T}}_{\mu,N} = \tilde{\mathbf{F}}_{\mu,N}\tilde{\mathbf{P}}_N, \tag{37a}$$

$$\tilde{\mathbf{F}}_{v,N} = [\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S}\tilde{\rho}_{m,N}\mathbf{M}'_{m,N})^{+}]\mathbf{H}_N \text{, or} \tag{37b}$$

$$\tilde{\mathbf{F}}_{\mu,N} = (\mathbf{e}'_T \otimes \mathbf{I}_N)[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S}\tilde{\rho}_{m,N}\mathbf{M}'_{m,N})^{+}]\mathbf{H}_N,$$

or

$$\tilde{\mathbf{F}}_{v,N} = \tilde{\mathbf{\Omega}}^{-1}_{\boldsymbol{\varepsilon},N}[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S}\tilde{\rho}_{m,N}\mathbf{M}_{m,N})]\mathbf{H}_N, \tag{37c}$$

$$\tilde{\mathbf{F}}_{\mu,N} = [\tilde{\sigma}^{-2}_{1,N}(\mathbf{e}'_T \otimes \mathbf{I}_N)][\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S}\tilde{\rho}_{m,N}\mathbf{M}_{m,N})]\mathbf{H}_N.$$

The estimators of $\mathbf{a}_{p,s,N} = \mathbf{T}_N\boldsymbol{\alpha}_{p,s,N}$, $p = 1,...,4,a,b$, $s = 1...,S$, $\mathbf{a}_{a,N} = \mathbf{T}_N\boldsymbol{\alpha}_{a,N}$, and $\mathbf{a}_{b,N} = \mathbf{T}_N\boldsymbol{\alpha}_{b,N}$ are then given by

$$\tilde{\mathbf{a}}_{p,s,N} = \tilde{\mathbf{T}}_N\tilde{\boldsymbol{\alpha}}_{p,s,N} \tag{38}$$

with $\tilde{\boldsymbol{\alpha}}_{p,s,N} = 2N^{-1}(\mathbf{D}'_N\tilde{\mathbf{C}}_{p,s,N}\tilde{\mathbf{u}}_N)$, and the matrices $\tilde{\mathbf{C}}_{p,s,N}$, $p = 1,...,4$, $s = 1,...,S$, $\tilde{\mathbf{C}}_{a,N}$, and $\tilde{\mathbf{C}}_{b,N}$ are given by (22) with $\boldsymbol{\rho}_N$ replaced by $\tilde{\boldsymbol{\rho}}_N$.

The elements of the estimated $(4S+2)\times(4S+2)$ matrix $\tilde{\mathbf{\Psi}}_N$ are defined in (34d), with $\sigma_{v,N}$ and $\sigma_{\mu,N}$ replaced by $\tilde{\sigma}_{v,N}$ and $\tilde{\sigma}_{v,N}$. The third and fourth moments of $\mu_{i,N}$ and $v_{it,N}$, denoted as $\sigma^{(3)}_{\mu}, \sigma^{(3)}_{v}$ and $\sigma^{(4)}_{\mu}, \sigma^{(4)}_{v}$, can be estimated consistently as follows (see Appendix B for a proof):

$$\tilde{\sigma}^{(3)}_{\varepsilon,N} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\varepsilon}^{3}_{it,N}, \tag{39a}$$

$$\tilde{\sigma}^{(3)}_{\mu,N} = \frac{1}{NT(T-1)} \sum_{i=1}^{N} \sum_{s=1}^{T} \sum_{\substack{t=1 \\ t \neq s}}^{T} \tilde{\varepsilon}_{is,N} \tilde{\varepsilon}^{2}_{it} , \tag{39b}$$

$$\tilde{\sigma}^{(3)}_{v,N} = \tilde{\sigma}^{(3)}_{\varepsilon,N} - \tilde{\sigma}^{(3)}_{\mu,N} , \tag{39c}$$

as well as

$$\tilde{\sigma}^{(4)}_{\mu,N} = \frac{1}{NT(T-1)} \sum_{i=1}^{N} \sum_{s=1}^{T} \sum_{\substack{t=1 \\ t \neq s}}^{T} \tilde{\varepsilon}_{is,N} \tilde{\varepsilon}^{3}_{it,N} \tag{40a}$$

$$- \frac{3}{NT(T-1)} \sum_{i=1}^{N} \sum_{s=1}^{T} \sum_{\substack{t=1 \\ t \neq s}}^{T} \tilde{\varepsilon}_{is,N} \tilde{\varepsilon}_{it,N} \left(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{\varepsilon}^{2}_{it,N} - \frac{1}{NT(T-1)} \sum_{i=1}^{N} \sum_{s=1}^{T} \sum_{\substack{t=1 \\ t \neq s}}^{T} \tilde{\varepsilon}_{is,N} \tilde{\varepsilon}_{it,N} \right) , \tag{40b}$$

$$\tilde{\sigma}^{(4)}_{v,N} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{\varepsilon}^{4}_{it,N} - \frac{1}{NT(T-1)} \sum_{i=1}^{N} \sum_{s=1}^{T} \sum_{\substack{t=1 \\ t \neq s}}^{T} \tilde{\varepsilon}_{is,N} \tilde{\varepsilon}^{3}_{it,N} \tag{40c}$$

$$- \frac{3}{NT(T-1)} \sum_{i=1}^{N} \sum_{s=1}^{T} \sum_{\substack{t=1 \\ t \neq s}}^{T} \tilde{\varepsilon}_{is,N} \tilde{\varepsilon}_{it,N} \left(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{\varepsilon}^{2}_{it,N} - \frac{1}{NT(T-1)} \sum_{i=1}^{N} \sum_{s=1}^{T} \sum_{\substack{t=1 \\ t \neq s}}^{T} \tilde{\varepsilon}_{is,N} \tilde{\varepsilon}_{it,N} \right) ,$$

where $\tilde{\boldsymbol{\varepsilon}}_N = \mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S} \tilde{\rho}_{m,N} \mathbf{M}_{m,N}) \tilde{\mathbf{u}}_N$.[9] Based on $\tilde{\boldsymbol{\Psi}}_N$, we can now define the estimator for $\boldsymbol{\Omega}_{\tilde{\boldsymbol{\theta}}_N}$ as

$$\tilde{\boldsymbol{\Omega}}_{\tilde{\boldsymbol{\theta}}_N}(\tilde{\boldsymbol{\Theta}}_N) = (\tilde{\mathbf{J}}'_N \tilde{\boldsymbol{\Theta}}_N \tilde{\mathbf{J}}_N)^{+} \tilde{\mathbf{J}}'_N \tilde{\boldsymbol{\Theta}}_N \tilde{\boldsymbol{\Psi}}_N \tilde{\boldsymbol{\Theta}}_N \tilde{\mathbf{J}}_N (\tilde{\mathbf{J}}'_N \tilde{\boldsymbol{\Theta}}_N \tilde{\mathbf{J}}_N)^{+} . \tag{41}$$

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

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 bounded uniformly. Suppose furthermore (a) that the elements of the nonstochastic matrices $\mathbf{H}_N$ are bounded uniformly in absolute value, (b) $\sup_N \sum_{s=1}^{S} \left| \rho_{s,N} \right| < 1$ 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) $\tilde{\mathbf{P}}_N - \mathbf{P}_N = o_p(1)$ with $\mathbf{P}_N = O(1)$. Then, $\tilde{\boldsymbol{\Psi}}_N - \boldsymbol{\Psi}_N = o_p(1)$ and $\tilde{\boldsymbol{\Psi}}^{-1}_N - \boldsymbol{\Psi}^{-1}_N = o_p(1)$. Furthermore, if Assumption 5 holds, then also $\tilde{\boldsymbol{\Omega}}_{\tilde{\boldsymbol{\theta}}_N} - \boldsymbol{\Omega}_{\tilde{\boldsymbol{\theta}}_N} = o_p(1)$.

---

[9] Compare Gilbert (2002) for the estimation of third and fourth moments in error component models without spatial lags and without spatial autoregressive disturbances.

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

Note that both $N^{1/2}(\tilde{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_N)$ and $(NT)^{1/2}\boldsymbol{\Delta}_N$, and thus also $N^{1/2}\boldsymbol{\Delta}_N$ are asymptotically linear in $\boldsymbol{\xi}_N$. Hence, the joint distribution of the vector $[N^{1/2}\boldsymbol{\Delta}'_N, N^{1/2}(\tilde{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_N)']'$ can be derived invoking the central limit theorem for vectors of quadratic forms by Kelejian and Prucha (2010).

Consider the $(P_* + 4S + 2)\times 1$ vector of linear and linear quadratic forms in $\boldsymbol{\xi}_N$:

$$\mathbf{w}_N = \begin{bmatrix} (NT)^{-1/2}\mathbf{F}'_N\boldsymbol{\xi}_N \\ \mathbf{q}_N \end{bmatrix}. \tag{42}$$

Its variance-covariance matrix is of dimension $(P_* + 4S + 2)\times(P_* + 4S + 2)$ and given by:

$$Var(\mathbf{w}_N) = \boldsymbol{\Psi}_{o,N} = E\begin{bmatrix} (NT)^{-1}\mathbf{F}'_N\boldsymbol{\xi}_N\boldsymbol{\xi}'_N\mathbf{F}_N & (NT)^{-1/2}\mathbf{F}'_N\boldsymbol{\xi}_N\mathbf{q}'_N \\ (NT)^{-1/2}\mathbf{q}_N\boldsymbol{\xi}'_N\mathbf{F}_N & \mathbf{q}_N\mathbf{q}'_N \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Psi}_{\Delta\Delta,N} & \boldsymbol{\Psi}_{\Delta\theta,N} \\ \boldsymbol{\Psi}'_{\Delta\theta,N} & \boldsymbol{\Psi}_N \end{bmatrix}, \tag{43a}$$

where the $(4S+2)\times(4S+2)$ matrix $\boldsymbol{\Psi}_N$ is defined above, $\boldsymbol{\Psi}_{\Delta\Delta,N}$ is of dimension $P_* \times P_*$ and defined as

$$\boldsymbol{\Psi}_{\Delta\Delta,N} = E[(NT)^{-1}\mathbf{F}'_N\boldsymbol{\xi}_N\boldsymbol{\xi}'_N\mathbf{F}_N] = (NT)^{-1}(\sigma_\nu^2\mathbf{F}'_{\nu,N}\mathbf{F}_{\nu,N} + \sigma_\mu^2\mathbf{F}'_{\mu,N}\mathbf{F}_{\mu,N}), \tag{43b}$$

and the $P_* \times (4S+2)$ matrix $\boldsymbol{\Psi}_{\Delta\theta,N}$ is given by

$$\begin{aligned}
\boldsymbol{\Psi}_{\Delta\theta,N} &= E[(NT)^{-1/2}\mathbf{F}'_N\boldsymbol{\xi}_N\mathbf{q}'_N] \\
&= (NT)^{-1/2}N^{-1/2}[\mathbf{F}'_{\nu,N}(\sigma_\nu^{(3)}\boldsymbol{\kappa}_{\mathbf{A}^\nu_{1,1,N}} + \sigma_\nu^2\mathbf{a}^\nu_{1,1,N}) + \mathbf{F}'_{\mu,N}(\sigma_\mu^{(3)}\boldsymbol{\kappa}_{\mathbf{A}^\mu_{1,1,N}} + \sigma_\mu^2\mathbf{a}^\mu_{1,1,N}),\ldots, \\
&\ldots,\mathbf{F}'_{\nu,N}(\sigma_\nu^{(3)}\boldsymbol{\kappa}_{\mathbf{A}^\nu_{a,N}} + \sigma_\nu^2\mathbf{a}^\nu_{a,N}) + \mathbf{F}'_{\mu,N}(\sigma_\mu^{(3)}\boldsymbol{\kappa}_{\mathbf{A}^\mu_{a,N}} + \sigma_\mu^2\mathbf{a}^\mu_{a,N}), \\
&\mathbf{F}'_{\nu,N}(\sigma_\nu^{(3)}\boldsymbol{\kappa}_{\mathbf{A}^\nu_{b,N}} + \sigma_\nu^2\mathbf{a}^\nu_{b,N}) + \mathbf{F}'_{\mu,N}(\sigma_\mu^{(3)}\boldsymbol{\kappa}_{\mathbf{A}^\mu_{b,N}} + \sigma_\mu^2\mathbf{a}^\mu_{b,N}),
\end{aligned} \tag{43c}$$

where $\boldsymbol{\kappa}_{\mathbf{A}^\nu_{p,q,N}}$ and $\boldsymbol{\kappa}_{\mathbf{A}^\mu_{p,q,N}}$ are $NT\times 1$ and $N\times 1$ vectors, whose *i*-th element corresponds to the *i*-th main diagonal element of $\mathbf{A}^\nu_{p,q,N}$ and $\mathbf{A}^\mu_{p,q,N}$, respectively.

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

$$\widetilde{\mathbf{\Psi}}_{o,N}=\begin{bmatrix}\widetilde{\mathbf{\Psi}}_{\Delta\Delta,N} & \widetilde{\mathbf{\Psi}}_{\Delta\theta,N}\\ \widetilde{\mathbf{\Psi}}'_{\Delta\theta,N} & \widetilde{\mathbf{\Psi}}_{N}\end{bmatrix}, \text{ where} \tag{44}$$

$$\widetilde{\mathbf{\Psi}}_{\Delta\Delta,N}=(NT)^{-1}[\widetilde{\sigma}_{v}^{2}\widetilde{\mathbf{F}}'_{v,N}\widetilde{\mathbf{F}}_{v,N}+\widetilde{\sigma}_{\mu}^{2}\widetilde{\mathbf{F}}'_{\mu,N}\widetilde{\mathbf{F}}_{\mu,N}],$$

$$\widetilde{\mathbf{\Psi}}_{\Delta\theta,N}=(NT)^{-1/2}N^{-1/2}[\widetilde{\mathbf{F}}'_{v,N}(\widetilde{\sigma}_{v}^{(3)}\mathbf{\kappa}_{\mathbf{A}_{1,1,N}^{v}}+\widetilde{\sigma}_{v}^{2}\widetilde{\mathbf{a}}_{1,1,N}^{v})+\widetilde{\mathbf{F}}'_{\mu,N}(\widetilde{\sigma}_{\mu}^{(3)}\mathbf{\kappa}_{\mathbf{A}_{1,1,N}^{\mu}}+\widetilde{\sigma}_{\mu}^{2}\widetilde{\mathbf{a}}_{1,1,N}^{\mu}),...,$$

$$...,\widetilde{\mathbf{F}}'_{v,N}(\widetilde{\sigma}_{v}^{(3)}\mathbf{\kappa}_{\mathbf{A}_{a,N}^{v}}+\widetilde{\sigma}_{v}^{2}\widetilde{\mathbf{a}}_{a,N}^{v})+\widetilde{\mathbf{F}}'_{\mu,N}(\widetilde{\sigma}_{\mu}^{(3)}\mathbf{\kappa}_{\mathbf{A}_{a,N}^{\mu}}+\widetilde{\sigma}_{\mu}^{2}\widetilde{\mathbf{a}}_{a,N}^{\mu}),$$

$$\widetilde{\mathbf{F}}'_{v,N}(\widetilde{\sigma}_{v}^{(3)}\mathbf{\kappa}_{\mathbf{A}_{b,N}^{v}}+\widetilde{\sigma}_{v}^{2}\widetilde{\mathbf{a}}_{b,N}^{v})+\widetilde{\mathbf{F}}'_{\mu,N}(\widetilde{\sigma}_{\mu}^{(3)}\mathbf{\kappa}_{\mathbf{A}_{b,N}^{\mu}}+\widetilde{\sigma}_{\mu}^{2}\widetilde{\mathbf{a}}_{b,N}^{\mu})].$$

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

Theorem 4. Joint Distribution of $\widetilde{\mathbf{\theta}}_N$ and Other Model Parameters

Suppose all assumptions used in Theorem 3 hold and $\lambda_{\min}(\mathbf{\Psi}_{o,N})\geq c^{*}_{\mathbf{\Psi}_o}>0$. Then,

$$\begin{bmatrix}N^{1/2}\mathbf{\Delta}_N\\ N^{1/2}(\widetilde{\mathbf{\theta}}_N-\mathbf{\theta}_N)\end{bmatrix}=\begin{bmatrix}T^{-1/2}\mathbf{P}'_N & \mathbf{0}\\ \mathbf{0} & (\mathbf{J}'_N\mathbf{\Theta}_N\mathbf{J}_N)^{-1}\mathbf{J}'_N\mathbf{\Theta}_N\end{bmatrix}\mathbf{\Psi}_{o,N}^{1/2}\mathbf{\xi}_{o,N}+o_p(1), \text{ with}$$

$$\mathbf{\xi}_{o,N}=\mathbf{\Psi}_{o,N}^{-1/2}[N^{-1/2}\mathbf{\xi}'_N\mathbf{F}_N,\mathbf{q}'_N]'\xrightarrow{d}N(\mathbf{0},\mathbf{I}_{P^*+4S+2}),$$

$$\mathbf{\Omega}_{o,N}=\begin{bmatrix}T^{-1/2}\mathbf{P}'_N & \mathbf{0}\\ \mathbf{0} & (\mathbf{J}'_N\mathbf{\Theta}_N\mathbf{J}_N)^{-1}\mathbf{J}'_N\mathbf{\Theta}_N\end{bmatrix}\mathbf{\Psi}_{o,N}\begin{bmatrix}T^{-1/2}\mathbf{P}_N & \mathbf{0}\\ \mathbf{0} & \mathbf{\Theta}_N\mathbf{J}_N(\mathbf{J}'_N\mathbf{\Theta}_N\mathbf{J}_N)^{-1}\end{bmatrix}, \text{ and}$$

$$\widetilde{\mathbf{\Omega}}_{o,N}=\begin{bmatrix}T^{-1/2}\widetilde{\mathbf{P}}'_N & \mathbf{0}\\ \mathbf{0} & (\widetilde{\mathbf{J}}'_N\widetilde{\mathbf{\Theta}}_N\widetilde{\mathbf{J}}_N)^{+}\widetilde{\mathbf{J}}'_N\widetilde{\mathbf{\Theta}}_N\end{bmatrix}\widetilde{\mathbf{\Psi}}_{o,N}\begin{bmatrix}T^{-1/2}\widetilde{\mathbf{P}}_N & \mathbf{0}\\ \mathbf{0} & \widetilde{\mathbf{\Theta}}_N\widetilde{\mathbf{J}}_N(\widetilde{\mathbf{J}}'_N\widetilde{\mathbf{\Theta}}_N\widetilde{\mathbf{J}}_N)^{+}\end{bmatrix}.$$

Moreover,

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

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

Remark 1.

As in Kelejian and Prucha (2010, p. 17), Theorem 4 can also be used to obtain the joint distribution of $(\tilde{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_N)$ and some other estimator $\boldsymbol{\Delta}_N^{**}$, where $(NT)^{1/2}\boldsymbol{\Delta}_N^{**} = (NT)^{1/2}\mathbf{T}_N^{**\prime}\boldsymbol{\xi}_N + o_p(1)$, $\mathbf{T}_N^{**} = \mathbf{F}_N^{**}\mathbf{P}_N^{**}$, $\tilde{\mathbf{T}}_N^{**} = \tilde{\mathbf{F}}_N^{**}\tilde{\mathbf{P}}_N^{**}$, assuming that analogous assumptions are maintained for this estimator. In particular, the results remain valid, but with $\mathbf{F}_N$, $\mathbf{P}_N$ replaced by $\mathbf{F}_N^{**}$, $\mathbf{P}_N^{**}$, and $\tilde{\mathbf{F}}_N$, $\tilde{\mathbf{P}}_N$ replaced by $\tilde{\mathbf{F}}_N^{**}$, $\tilde{\mathbf{P}}_N^{**}$, in the definitions of $\boldsymbol{\Psi}_{\Delta\Delta,N}$, $\boldsymbol{\Psi}_{\Delta\theta,N}$, $\tilde{\boldsymbol{\Psi}}_{\Delta\Delta,N}$, and $\tilde{\boldsymbol{\Psi}}_{\Delta\theta,N}$.

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

1. Instruments

It is evident from model (1), that $E(\mathbf{u}_N'\overline{\mathbf{Y}}_N) \neq \mathbf{0}$. In line with Kelejian and Prucha (2010), 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) and uniformly bounded elements in absolute value.

Assumption 9.

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

Assumption 10.

$\mathbf{Q}_{\mathbf{HH}} = \lim_{N\to\infty}[(NT)^{-1}\mathbf{H}_N'\mathbf{H}_N]$ and $\mathbf{Q}_{\mathbf{HZ}} = \text{p}\lim_{N\to\infty}[(NT)^{-1}\mathbf{H}_N'\mathbf{Z}_N]$ are finite and nonsingular.

Regarding the choice of instruments, note that

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

$$= \sum_{r=1}^{R}\mathbf{W}_{r,N}\{\mathbf{I}_T \otimes [\mathbf{I}_N + \sum_{i=1}^{\infty}(\sum_{r'=1}^{R}\lambda_{r',N}\mathbf{W}_{r',N})^i]\}\mathbf{X}_N\boldsymbol{\beta}_N, \tag{45}$$

provided that $\left\|\sum_{r'=1}^{R}\lambda_{r',N}\mathbf{W}_{r',N}\right\| < 1$. The matrices $\mathbf{H}_N$ are used to instrument $\mathbf{Z}_N = (\mathbf{X}_N, \overline{\mathbf{Y}}_N)$ in terms of their predicted values $\hat{\mathbf{Z}}_N = \mathbf{P}_{\mathbf{H}_N}\mathbf{Z}_N$, where $\mathbf{P}_{\mathbf{H}_N} = \mathbf{H}_N(\mathbf{H}_N'\mathbf{H}_N)^{-1}\mathbf{H}_N'$. In light of (45) $\mathbf{H}_N$ may include $\mathbf{X}_N$ and a subset of the linearly independent columns of terms of the sum

$$[\mathbf{I}_T \otimes \sum_{i=1}^{Q}(\sum_{r'=1}^{R}\mathbf{W}_{r',N})^i]\mathbf{X}_N, \tag{46}$$

where Q is some predefined constant. Such specification of $\mathbf{H}_N$ complies with the second part of Assumption 9 (by Assumptions 3 and 8).

2. Definition of TSLS Estimator and Asymptotic Results

Estimation of (1) proceeds in three steps. In the first step, (1a) is estimated by TSLS using instruments $\mathbf{H}_N$. In the second step, $\rho_{1,N},...,\rho_{S,N}$, σ_v^2, and σ_1^2 are estimated using the GM estimators defined in Section III in (17) and (18), based on consistent estimates of $\mathbf{u}_N$ from the first step. In the third step, the model is re-estimated by feasible generalized TSLS (FGTSLS), which is equivalent to TSLS on transformed Eq. (1). This approach allows for testing joint hypotheses about $\boldsymbol{\delta}_N$ and $\boldsymbol{\theta}_N$.

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

$$\tilde{\boldsymbol{\delta}}_N = (\hat{\mathbf{Z}}_N'\mathbf{Z}_N)^{-1}\hat{\mathbf{Z}}_N'\mathbf{y}_N, \text{ where} \tag{47}$$

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

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

In the second step, the parameters $\rho_{s,N}$, $s = 1,...,S$, σ_v^2, and σ_1^2, are estimated using the GM estimator defined by (18), 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 $\tilde{\rho}_{s,N}$, $s = 1,...,S$, $\tilde{\sigma}_{v,N}^2$, and $\tilde{\sigma}_{1,N}^2$.

The following lemma shows that the various assumptions maintained in Section III are automatically satisfied by the TSLS estimator $\tilde{\boldsymbol{\delta}}_N$ and the corresponding residuals $\tilde{\mathbf{u}}_N$. A proof is given in Appendix C.

Lemma 1.[10]

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

(a) $(NT)^{1/2}(\tilde{\boldsymbol{\delta}}_N - \boldsymbol{\delta}_N) = (NT)^{-1/2}\mathbf{T}'_N\boldsymbol{\xi}_N + o_p(1) = (NT)^{-1/2}\mathbf{T}'_{v,N}\mathbf{v}_N + (NT)^{-1/2}\mathbf{T}'_{\mu,N}\boldsymbol{\mu}_N + o_p(1)$, where

$\boldsymbol{\xi}_N = (\mathbf{v}'_N, \boldsymbol{\mu}'_N)'$, $\mathbf{T}_N = (\mathbf{T}'_{v,N}, \mathbf{T}'_{\mu,N})'$,

$\mathbf{T}_{v,N} = \mathbf{F}_{v,N}\mathbf{P}_N$, $\mathbf{T}_{\mu,N} = \mathbf{F}_{\mu,N}\mathbf{P}_N$,

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

$\mathbf{F}_{v,N} = [\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S}\rho_{m,N}\mathbf{M}'_{m,N})^{-1}]\mathbf{H}_N$, and

$\mathbf{F}_{\mu,N} = (\mathbf{e}'_T \otimes \mathbf{I}_N)[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S}\rho_{m,N}\mathbf{M}'_{m,N})^{-1}]\mathbf{H}_N$.

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

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

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

Condition $\sup_N \|\boldsymbol{\beta}_N\| \le b < \infty$ 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 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 $\tilde{\boldsymbol{\theta}}_N$ based on TSLS residuals. The lemma also establishes that the elements of $\mathbf{D}_N$ are bounded uniformly 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 cover the GM estimator $\tilde{\boldsymbol{\theta}}_N$ and the TSLS estimator $\tilde{\boldsymbol{\delta}}_N$. In particular, Theorem 4 gives the joint limiting distribution of $N^{1/2}(\tilde{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_N)$ and $N^{1/2}(\tilde{\boldsymbol{\delta}}_N - \boldsymbol{\delta}_N)$, where $\mathbf{D}_N = -\mathbf{Z}_N$, the matrices $\mathbf{P}_N, \tilde{\mathbf{P}}_N$, $\mathbf{F}_{v,N}$, $\mathbf{F}_{\mu,N}$ are as

[10] Compare Kelejian and Prucha (2008) for analogous results in case of a cross-section SARAR(1,1) model and Badinger and Egger (2011) in case of a cross-section SARAR(*R*,*S*) model.

in Lemma 1, $\tilde{\mathbf{F}}_{v,N} = [\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S} \tilde{\rho}_{m,N} \mathbf{M}'_{m,N})^{+}]\mathbf{H}_N$ and $\tilde{\mathbf{F}}_{\mu,N} = (\mathbf{e}'_T \otimes \mathbf{I}_N)[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S} \tilde{\rho}_{m,N} \mathbf{M}'_{m,N})^{+}]\mathbf{H}_N$.

We now turn to the third step of estimation. Consider the transformed model (1b), with

$$\mathbf{y}_N^{**} = \mathbf{Z}_N^{**}\boldsymbol{\delta}_N + \mathbf{u}_N^{**}, \tag{48}$$

where

$$\mathbf{y}_N^{**} = \boldsymbol{\Omega}_{\boldsymbol{\varepsilon},N}^{-1/2}[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N} \mathbf{M}_{m,N})]\mathbf{y}_N$$

$$\mathbf{Z}_N^{**} = \boldsymbol{\Omega}_{\boldsymbol{\varepsilon},N}^{-1/2}[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N} \mathbf{M}_{m,N})]\mathbf{Z}_N,$$

$$\mathbf{u}_N^{**} = \boldsymbol{\Omega}_{\boldsymbol{\varepsilon},N}^{-1/2}[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N} \mathbf{M}_{m,N})]\mathbf{u}_N = \boldsymbol{\Omega}_{\varepsilon,N}^{-1/2}\boldsymbol{\varepsilon}_N,$$

$$\mathbf{H}_N^{**} = \boldsymbol{\Omega}_{\boldsymbol{\varepsilon},N}^{-1/2}[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N} \mathbf{M}_{m,N})]\mathbf{H}_N = \boldsymbol{\Omega}_{\varepsilon,N}^{-1/2}\mathbf{H}_N^{*},$$

$$\mathbf{H}_N^{*} = [\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S} \rho_{m,N} \mathbf{M}_{m,N})]\mathbf{H}_N,$$

$$\hat{\mathbf{Z}}_N^{**} = \mathbf{P}_{\mathbf{H}_N^{**}}\mathbf{Z}_N^{**} = \mathbf{H}_N^{**}(\mathbf{H}_N^{**\prime}\mathbf{H}_N^{**})^{-1}\mathbf{H}_N^{**\prime}\mathbf{Z}_N^{**}.$$

The generalized TSLS (GTSLS) estimator, denoted as $\hat{\boldsymbol{\delta}}_N$, is then obtained as a TSLS estimator applied to the transformed model (56), using the transformed instruments $\mathbf{H}_N^{**} = (\mathbf{I} - \sum_{m=1}^{S} \rho_{m,N} \mathbf{M}_{m,N})\mathbf{H}_N$, i.e.,

$$\hat{\boldsymbol{\delta}}_N = (\hat{\mathbf{Z}}_N^{**\prime}\mathbf{Z}_N^{**})^{-1}\hat{\mathbf{Z}}_N^{**\prime}\mathbf{y}_N^{**}. \tag{49}$$

The FGTSLS estimator, denoted as $\hat{\tilde{\boldsymbol{\delta}}}_N$, is defined analogously, after replacing $\boldsymbol{\rho}_N$ by $\tilde{\boldsymbol{\rho}}_N$ ($\boldsymbol{\Omega}_{\boldsymbol{\varepsilon},N}$ by $\tilde{\boldsymbol{\Omega}}_{\boldsymbol{\varepsilon},N}$), i.e.,

$$\hat{\tilde{\boldsymbol{\delta}}}_N = (\hat{\tilde{\mathbf{Z}}}_N^{**\prime}\tilde{\mathbf{Z}}_N^{**})^{-1}\hat{\tilde{\mathbf{Z}}}_N^{**\prime}\tilde{\mathbf{y}}_N^{**}, \tag{50}$$

where

$$\hat{\tilde{\mathbf{Z}}}_N^{**} = \mathbf{P}_{\tilde{\mathbf{H}}_N^{**}} \tilde{\mathbf{Z}}_N^{**}, \text{ with } \mathbf{P}_{\mathbf{H}_N} = \tilde{\mathbf{H}}_N^{**}(\tilde{\mathbf{H}}_N^{**\prime}\tilde{\mathbf{H}}_N^{**})^{-1}\tilde{\mathbf{H}}_N^{**\prime},$$

$$\tilde{\mathbf{H}}_N^{**} = \tilde{\mathbf{\Omega}}_{\boldsymbol{\varepsilon},N}^{-1/2}[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S}\tilde{\rho}_{m,N}\mathbf{M}_{m,N})]\mathbf{H}_N,$$

$$\tilde{\mathbf{Z}}_N^{**} = \tilde{\mathbf{\Omega}}_{\boldsymbol{\varepsilon},N}^{-1/2}[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S}\tilde{\rho}_{m,N}\mathbf{M}_{m,N})]\mathbf{Z}_N,$$

$$\tilde{\mathbf{y}}_N^{**} = \tilde{\mathbf{\Omega}}_{\boldsymbol{\varepsilon},N}^{-1/2}[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S}\tilde{\rho}_{m,N}\mathbf{M}_{m,N})]\mathbf{y}_N.$$

Kelejian and Prucha (2010) and Arraiz, Drukker, Kelejian and Prucha (2010) use the untransformed instrument matrix $\mathbf{H}_N$ in FGTSLS estimation of cross-section SARAR(1,1) models. In light of (45), the ideal instruments matrix for $\overline{\mathbf{Y}}_N^{**}$ in the transformed model is given by $\mathbf{H}_N^{**}$.

The following lemma shows that the various assumptions maintained in Section III are automatically satisfied by (F)GTSLS estimator $\hat{\tilde{\boldsymbol{\delta}}}_N$ and the corresponding residuals. The proof is given in Appendix C.

Lemma 2.

Suppose the Assumptions of Lemma 1 hold,[11] and define $\hat{\breve{\boldsymbol{\delta}}}_N$ as in Eq. (50), where $\breve{\boldsymbol{\theta}}_N$ is any $N^{1/2}$-consistent estimator of $\boldsymbol{\theta}_N$ (such as the GM estimator $\tilde{\boldsymbol{\theta}}_N$ based on TSLS residuals). Then

(a) $(NT)^{1/2}\boldsymbol{\Delta}_N^{**} = (NT)^{-1/2}\mathbf{T}_N^{**\prime}\boldsymbol{\xi}_N + o_p(1) = (NT)^{-1/2}\mathbf{T}_{v,N}^{**\prime}\mathbf{v}_N + (NT)^{-1/2}\mathbf{T}_{\mu,N}^{**\prime}\boldsymbol{\mu}_N + o_p(1)$, where

$\boldsymbol{\xi}_N = (\mathbf{v}_N', \boldsymbol{\mu}_N')'$, $\mathbf{T}_N^{**} = (\mathbf{T}_{v,N}^{**\prime}, \mathbf{T}_{\mu,N}^{**\prime})'$,

$\mathbf{T}_{v,N}^{**} = \mathbf{F}_{v,N}^{**}\mathbf{P}_N^{**}$, $\mathbf{T}_{\mu,N}^{**} = \mathbf{F}_{\mu,N}^{**}\mathbf{P}_N^{**}$,

$$\mathbf{P}_N^{**} = \mathbf{Q}_{\mathbf{H}^{**}\mathbf{H}^{**}}^{-1}\mathbf{Q}_{\mathbf{H}^{**}\mathbf{Z}^{**}}(\mathbf{Q}_{\mathbf{H}^{**}\mathbf{Z}^{**}}'\mathbf{Q}_{\mathbf{H}^{**}\mathbf{H}^{**}}^{-1}\mathbf{Q}_{\mathbf{H}^{**}\mathbf{Z}^{**}})^{-1},$$

$$\mathbf{F}_{v,N}^{**} = (\sigma_v^{-2}\mathbf{Q}_{0,N} + \sigma_1^{-2}\mathbf{Q}_{1,N})\mathbf{H}_N^{*} = \mathbf{\Omega}_{\boldsymbol{\varepsilon},N}^{-1}[\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N})]\mathbf{H}_N,$$

$$\mathbf{F}_{\mu,N}^{**} = [\sigma_1^{-2}(\mathbf{e}_T' \otimes \mathbf{I}_N)]\mathbf{H}_N^{*} = [\sigma_1^{-2}(\mathbf{e}_T' \otimes \mathbf{I}_N)][\mathbf{I}_T \otimes (\mathbf{I}_N - \sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N})]\mathbf{H}_N.$$

---

[11] In light of the properties of $(\mathbf{I}_N - \sum_{m=1}^{S}\rho_{m,N}\mathbf{M}_{m,N})$ and $\mathbf{\Omega}_{\boldsymbol{\varepsilon},N}$, this implies that Assumptions 9 and 10 will also be satisfied for the transformed instruments $\mathbf{H}_N^{**}$.

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

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

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

In light of Lemmata 1 and 2 the joint limiting distribution of the (F)GTSLS estimator $\hat{\breve{\boldsymbol{\delta}}}_N$ and the GM estimator $\breve{\boldsymbol{\theta}}_N$ follows from Theorem 4 and the discussion thereafter, with $\boldsymbol{\Delta}_N^{**} = \hat{\breve{\boldsymbol{\delta}}}_N - \boldsymbol{\delta}_N$. The asymptotic variance-covariance matrix and its corresponding estimator are provided in Theorem 4 with the modifications as described in Remark 1 thereafter.

V. Monte Carlo Evidence

To illustrate the based performance of the proposed estimation procedure, we consider a limited 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 $\mathbf{X}_N$ includes two explanatory variables. To economize on notation, let us suppress subscript N to indicate triangular arrays in the remainder of this section. Hence we have[12]

$$\mathbf{y} = \mathbf{x}_1\beta_1 + \mathbf{x}_2\beta_2 + \sum_{r=1}^{3}\lambda_r(\mathbf{I}_T \otimes \mathbf{W}_r)\mathbf{y} + \mathbf{u}, \tag{51a}$$

$$\mathbf{u} = \sum_{s=1}^{3}\rho_s(\mathbf{I}_T \otimes \mathbf{W}_s)\mathbf{u} + \boldsymbol{\varepsilon}. \tag{51b}$$

We consider two sample sizes: $N = 100$ and $N = 500$ and assume $T = 3$ throughout. 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 $\mathbf{W}^0$ consists of three $N \times N$ matrices $\mathbf{W}_1^0, \mathbf{W}_2^0$, and $\mathbf{W}_3^0$, where $\mathbf{W}_1^0 + \mathbf{W}_2^0 + \mathbf{W}_3^0 = \mathbf{W}^0$. The matrices $\mathbf{W}_1^0$, $\mathbf{W}_2^0$, and $\mathbf{W}_3^0$ are specified such that they

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

contain the elements of $\mathbf{W}^0$ for a different band of neighbours each. Otherwise, they have zero elements. We choose a design, where $\mathbf{W}_1^0$ corresponds to an 'up to three ahead and up to three behind' specification, $\mathbf{W}_2^0$ corresponds to a 'four to six ahead and four to six behind' specification, and $\mathbf{W}_3^0$ corresponds to a 'seven to nine ahead and seven to nine 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 row-normalized matrices $\mathbf{W}_1$, $\mathbf{W}_2$, and $\mathbf{W}_3$, the parameter space for $\boldsymbol{\lambda}$ and $\boldsymbol{\rho}$ must satisfy $0 \le |\lambda_1| + |\lambda_2| + |\lambda_3| < 1$ and $0 \le |\rho_1| + |\rho_2| + |\rho_3| < 1$. We consider three parameter constellations. In parameter constellation (1) there is third order spatial dependence in both the dependent variable and the disturbances, which is non-increasing in the order of neighbourhood, i.e., $\lambda_1 \ge \lambda_2 \ge \lambda_3$ and $\rho_1 \ge \rho_2 \ge \rho_3$. In (2), there is first order spatial dependence in both $\mathbf{y}$ and $\mathbf{u}$. Finally, constellation (3) considers zero dependence parameters for all spatial lags in $\mathbf{y}$ and $\mathbf{u}$, i.e., a non-spatial model.

< Table 1 here >

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

$$\mathbf{H} = (\mathbf{X}, (\mathbf{I}_T \otimes \mathbf{W}_1)\mathbf{X}, (\mathbf{I}_T \otimes \mathbf{W}_2)\mathbf{X}, (\mathbf{I}_T \otimes \mathbf{W}_3)\mathbf{X}, (\mathbf{I}_T \otimes \mathbf{W}_1^2)\mathbf{X}, (\mathbf{I}_T \otimes \mathbf{W}_2^2)\mathbf{X}, (\mathbf{I}_T \otimes \mathbf{W}_3^2)\mathbf{X}, \\ (\mathbf{I}_T \otimes \mathbf{W}_1\mathbf{W}_2)\mathbf{X}, (\mathbf{I}_T \otimes \mathbf{W}_2\mathbf{W}_3)\mathbf{X}). \tag{52}$$

We assume further that the error components v_{it} and μ_{it} are drawn from a standard normal distribution with zero mean and unit variance, i.e., $v_{it} = \zeta_{v,it}$ and $\mu_{it} = \zeta_{\mu,i}$ where each $\zeta_{v,it}$ and $\zeta_{\mu,i}$ are i.i.d. $N(0,1)$. One of the merits of spatial GM estimators relative to spatial maximum likelihood estimators is their suitability for non-normally distributed disturbances. This is not specific, however, to higher- versus lower-order or panel versus cross-section models. Hence, we refer readers interested in the performance of spatial GM versus spatial maximum likelihood panel data estimators to the Monte Carlo results in Lee and Yu (2010) or Baltagi, Egger, and Pfaffermayr (2012).

For each Monte Carlo experiment, we consider 1000 draws. Results for the estimates of $\rho_{1,N}, \rho_{2,N}$, and $\rho_{3,N}$ are obtained by the GM estimator defined in Eq. (18), using the optimal weighting matrix under normality $(\tilde{\mathbf{\Psi}}_N^{\circ})^{-1}$. The estimates reported for the regression parameters are FGTSLS estimates as defined in (50) using the transformed set of instruments $\tilde{\mathbf{H}}^{**}$.

For each coefficient, we report the average bias and root mean squared error (RMSE) for each parameter constellation and the rejection rates for the test that the coefficient is equal to the true parameter value. Under parameter constellation (2) we also test the SARAR(3,3) against the SARAR(1,1) model, using $H_0^{\lambda,\rho,*}: \lambda_2 = \lambda_3 = \rho_2 = \rho_3 = 0$. For the non-spatial model under parameter constellation (3), we report results for the tests of the joint hypothesis $H_0^{\lambda,\rho}: \lambda_1 = \lambda_2 = \lambda_3 = \rho_1 = \rho_2 = \rho_3 = 0$.

Table 2 reports the results of the Monte Carlo analysis for the two sample sizes considered.[13] In terms of bias and RMSE, the estimator performs well, even at $N = 100$. Across all parameter constellations, the bias and RMSE amount to 0.0007 and 0.0229 for the estimates of $\boldsymbol{\lambda} = (\lambda_1,...,\lambda_3)'$ and to 0.0054 and 0.1096 for the estimates of $\boldsymbol{\rho} = (\rho_1,...,\rho_3)'$. With an average rejection rate of 0.0082, the performance of the single hypothesis tests referring to $\boldsymbol{\lambda}$ and $\boldsymbol{\rho}$ is satisfactory. The actual size of the joint hypothesis tests, however, differs significantly from the nominal size with an average rejection rate of 0.1395.

< Table 2 >

However, performance improves quickly with growing sample size. For $N = 500$, the bias virtually disappears and the average RMSE of the estimates of $\boldsymbol{\lambda} = (\lambda_1,...,\lambda_3)'$ shrinks to 0.0010, that of the estimates of $\boldsymbol{\rho} = (\rho_1,...,\rho_3)'$ shrinks to 0.0440. 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 0.0089, that for the FGTSLS estimates of $\boldsymbol{\lambda}$ to 0.053. The average size of the joint hypothesis amounts to 0.084 for joint tests.

Overall, the Monte Carlo experiments illustrate that the proposed estimators work reasonably well in terms of bias and RMSE, even in very small samples. Regarding the estimates of the variance-covariance matrix of the parameter estimates and implied tests of single and joint hypothesis, some care is warranted in the interpretation of the results in small samples, though

[13] Results for the variances of the error components are very similar and thus omitted for the sake of brevity.

the difference to the true size of the tests is moderate at least for the single hypothesis tests. Hence, in small samples it might be worth exploiting additional moment conditions as outlined in footnote 3. As the sample size increases, the rejection rates of single and joint hypothesis tests converge reasonably quickly to the true size such that they 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 and Suggestions for Future Research

This paper derives GM and FGTSLS estimators for the parameters of SARAR(R,S) models allowing the applied econometrician to study the strength and pattern of spatial interdependence quite flexibly. We study the asymptotic properties of the proposed two-step estimators of the model parameters and derive their joint asymptotic distribution. This enables tests of the fairly general SARAR(R,S) model against restricted alternatives such as SARAR(0,S) and SARAR(R,0) or SARAR(1,1) with panel data.

One suggestion for future research is to extend the analysis of tests towards a study of conditional and unconditional tests on the relevance of error components and spatial interaction. In particular, a comprehensive Monte Carlo study of GM estimators using alternative weighting schemes of the moments and alternative distributional assumptions may be instructive.

References

Anselin, L. (2003). Spatial externalities, spatial multipliers, and spatial econometrics. *International Regional Science Review*, 26(2), 153-166.

Anselin, L. (2007). Spatial econometrics in RSUE: Retrospect and prospect. *Regional Science and Urban* Economics, 37(4), 450-456.

Arraiz, I., Drukker, D.M., Kelejian, H., and Prucha, I. (2010). A spatial Cliff-Ord-type model with heteroskedastic innovations: Small and large sample results. *Journal of Regional Science*, 50(2), 592-614.

Badinger, H. and Egger, P. (2011). GM estimation of higher-order spatial autoregressive processes in cross-section models with heteroskedastic disturbances. *Papers in Regional Science,* 90(1), 213-236.

Baltagi, B.H., Egger, P.H., and Pfaffermayr, M. (2012). A generalized spatial panel data model with random effects. *Econometric Reviews*, forthcoming.

Baltagi, B.H., Song, S.H., and Koh, W. (2003). Testing panel data regression models with spatial error correlation. *Journal of Econometrics*, 117(1), 123-150.

Beck, N., K.S. Gleditsch, Beardsley, K. (2006). Space is more than geography: Using spatial econometrics in the study of political economy. *International Studies Quarterly*, 50(1), 27-44.

Elhorst, J.P., D.J. Lacombe, and G. Piras (2012) On model specification and parameter space definitions in higher order spatial econometric models, *Regional Science and Urban Economics*, 42(1-2), 211-220.

Gilbert, S. (2002). Testing the distribution of error components in panel data models. *Economics Letters*, 77(1), 47-53.

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(1), 97-130.

Kelejian, H. and Robinson, D. (1992). Spatial autocorrelation: A new computationally simple test with an application to per capita county police expenditures. *Regional Science and Urban Economics*, 22(3), 317–331.

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(1), 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(2), 509-533.

Kelejian, H.H. and Prucha, I.R. (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics*, 157(1), 53-67.

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.

Koch, M. (2011a). Stationary parameter spaces in spatial econometric modelling, paper presented at the Vth World Conference of the Spatial Econometrics Association at Toulouse.

Koch, M. (2011b). A research note on asymptotic parameter spaces: A proof for the spatial one forward one behind pattern, paper presented at the Vth World Conference of the Spatial Econometrics Association at Toulouse.

LeSage, J.P. and Pace, R.K. (2009). *Introduction to Spatial Econometrics*. Boca Raton et al. CRC Press.

LeSage, J.P. and Pace, R.K. (2012). Pitfalls in higher order model extensions of basic spatial regression methodology, *Review of Regional Studies*, forthcoming.

Lee, L.F. and Liu, X. (2010). Efficient GMM estimation of high order spatial autoregressive models. *Econometric Theory*, 26(1), 187-230.

Lee, L.F. and Yu, J. (2010). Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics*, 154(2), 165-185.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica*, 46(1), 69-85.

Mutl, J. and Pfaffermayr, M. (2011). The Hausman test in a Cliff and Ord panel model. *Econometrics Journal*, 14(1), 48-76.

Pinkse, J. and Slade, M. (2010). The future of spatial econometrics. Journal of Regional Science, 50(1), p. 103 – 117.

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

Table 1. Parameter Constellations in Monte Carlo Experiments

Parameter constellation	λ_1	λ_2	λ_3	ρ_1	ρ_2	ρ_3
(1)	0.5	0.3	0.1	0.4	0.25	0.1
(2)	0.5	0	0	0.4	0	0
(3)	0	0	0	0	0	0

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

Table 2. Monte Carlo Results

Parameter constellation[1)]	N = 100			N = 500		
	(1)	(2)	(3)	(1)	(2)	(3)
λ_1	0.5	0.4	0	0.5	0.4	0
Bias	0.0004	0.0014	0.0013	0.0005	0.0000	0.0003
RMSE	0.0203	0.0230	0.0244	0.0088	0.0100	0.0099
Rej. Rate	0.0540	0.0590	0.0490	0.0590	0.0710	0.0380
λ_2	0.3	0	0	0.3	0	0
Bias	0.0008	0.0001	0.0001	-0.0002	0.0000	-0.0001
RMSE	0.0213	0.0226	0.0251	0.0094	0.0097	0.0104
Rej. Rate	0.0490	0.0520	0.0620	0.0620	0.0410	0.0480
λ_3	0.1	0	0	0.1	0	0
Bias	-0.0003	-0.0005	0.0010	0.0001	-0.0002	0.0000
RMSE	0.0213	0.0232	0.0250	0.0093	0.0102	0.0101
Rej. Rate	0.0520	0.0490	0.0690	0.0630	0.0530	0.0490
β_1	1	1	1	1	1	1
Bias	0.0001	-0.0004	-0.0003	0.0001	0.0000	0.0000
RMSE	0.0134	0.0132	0.0138	0.0061	0.0060	0.0061
Rej. Rate	0.0560	0.0500	0.0560	0.0550	0.0600	0.0480
β_2	1	1	1	1	1	1
Bias	-0.0007	-0.0002	-0.0001	0.0001	-0.0001	0.0002
RMSE	0.0130	0.0142	0.0133	0.0060	0.0058	0.0059
Rej. Rate	0.0460	0.0740	0.0550	0.0500	0.0520	0.0510
ρ_1	0.4	0.3	0	0.4	0.3	0
Bias	-0.0050	-0.0064	-0.0073	0.0013	0.0025	0.0027
RMSE	0.0946	0.1037	0.1261	0.0385	0.0426	0.0496
Rej. Rate	0.1070	0.1200	0.1330	0.0890	0.0910	0.0940
ρ_2	0.25	0	0	0.25	0	0
Bias	-0.0091	-0.0036	-0.0047	-0.0007	0.0002	0.0008
RMSE	0.1077	0.1107	0.1214	0.0444	0.0433	0.0477
Rej. Rate	0.1180	0.1090	0.1140	0.0870	0.0810	0.0790
ρ_3	0.1	0	0	0.1	0	0
Bias	-0.0079	-0.0020	-0.0028	-0.0027	0.0002	-0.0003
RMSE	0.1005	0.1044	0.1169	0.0404	0.0423	0.0475
Rej. Rate	0.0900	0.0980	0.0920	0.0790	0.0780	0.0860
Joint Tests[2)]						
Rej. Rate	-	0.1280	0.1510	-	0.0790	0.0880

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