ELSEVIER

Panel data models with spatially correlated error components

Mudit Kapoor[a], Harry H. Kelejian[b], Ingmar R. Prucha[b],*

[a]*Indian School of Business (ISB), Gachibowli, Hyderabad 500 019, India*
[b]*Department of Economics, University of Maryland, College Park, MD 20742, USA*

Abstract

In this paper we consider a panel data model with error components that are both spatially and time-wise correlated. The model blends specifications typically considered in the spatial literature with those considered in the error components literature. We introduce generalizations of the generalized moments estimators suggested in Kelejian and Prucha (1999. A generalized moments estimator for the autoregressive parameter in a spatial model. International Economic Review 40, 509–533) for estimating the spatial autoregressive parameter and the variance components of the disturbance process. We then use those estimators to define a feasible generalized least squares procedure for the regression parameters. We give formal large sample results for the proposed estimators. We emphasize that our estimators remain computationally feasible even in large samples.
© 2006 Elsevier B.V. All rights reserved.

1. Introduction

This paper considers the estimation of spatial models from panel data. Spatial models are important tools in economics, regional science and geography in analyzing a wide range of empirical issues. Spatial interactions could be due to competition between cross

---

*Corresponding author. Tel.: +1 301 405 3499; fax: +1 301 405 3542.
E-mail address: prucha@econ.umd.edu (I.R. Prucha).

sectional units, copy-cat policies, net work issues, spill-overs, externalities, regional issues, etc. Applications in the recent literature include, for example, the determinants of various forms of productivity, various categories of local public expenditures, vote seeking and tax setting behavior, population and employment growth, contagion problems, and the determinants of welfare expenditures.[1] To facilitate the empirical analysis of spatial issues the formal development of estimation methods for spatial models has received increasing attention in recent years.[2]

In spatial models, interactions between cross sectional units are typically modelled in terms of some measure of distance between them. By far the most widely used spatial models are variants of the ones considered in Cliff and Ord (1973, 1981). One method of estimation of these models is maximum likelihood (ML). However, even in its simplest form, the ML estimation of Cliff–Ord-type models entails substantial computational problems if the number of cross sectional units is large. Against this background, Kelejian and Prucha (1998) suggested an alternative instrumental variable estimation procedure for these models, which is based on a generalized moments (GM) estimator of a parameter in the spatial autoregressive process. This GM estimator was suggested by Kelejian and Prucha (1999) in an earlier paper.[3] The procedures suggested in Kelejian and Prucha (1998, 1999) are computationally feasible even for large sample sizes. Also, Kelejian and Prucha gave formal large sample results for their procedures.[4] As in most of the spatial literature, they consider the case where a single cross section of data is available. Monte Carlo results in Das et al. (2003) suggest that both the GM and the instrumental variable estimators are "virtually" as efficient as the corresponding ML estimators in small samples.

The purpose of this paper is two-fold. First, we introduce generalizations of the GM procedure in Kelejian and Prucha (1999) to panel data models involving a first order spatially autoregressive disturbance term, whose innovations have an error component structure. In particular, we introduce three GM estimators which correspond to alternative weighting schemes for the moments. Our specifications are such that the model's disturbances are potentially both spatially and time-wise autocorrelated, as well as heteroskedastic. These specifications merge those typically considered in the spatial literature with those considered in the error component literature.

Second, we define a feasible generalized least squares (FGLS) estimator for our model's regression parameters. This FGLS estimator is based on a spatial counterpart to the

---

[1]Some applications along these lines are, e.g., Audretsch and Feldmann (1996), Bernat (1996), Besley and Case (1995), Bollinger and Ihlanfeldt (1997), Buettner (1999), Case (1991), Case et al. (1993), Dowd and LeSage (1997), Holtz-Eakin (1994), LeSage (1999), Kelejian and Robinson (1997, 2000), Pinkse and Slade (1998), Pinkse et al. (2002), Shroder (1995), and Vigil (1998).

[2]Recent theoretical contributions include Baltagi and Li (2001a, b, 2004), Baltagi et al. (2003), Conley (1999), Das et al. (2003), Kelejian and Prucha (1997, 1998, 1999, 2001, 2004), Lee (2001a, b, 2002, 2003, 2004), LeSage (1997, 2000), Pace and Barry (1997), Pinkse and Slade (1998), Pinkse et al. (2002), and Rey and Boarnet (2004). Classic references concerning spatial models are Anselin (1988), Cliff and Ord (1973, 1981), and Cressie (1993).

[3]Due to publication lags, Kelejian and Prucha (1999) was published at a later date than Kelejian and Prucha (1998), even though it was written at an earlier point in time.

[4]The instrumental variable estimator suggested in Kelejian and Prucha (1998) is based on an approximation of the ideal instruments. In recent paper, Lee (2003) and Kelejian et al. (2004) extend their approach towards the ideal instruments.

Cochrane–Orcutt transformation, as well as transformations utilized in the estimation of classical error component models.[5]

We give formal results relating to our suggested procedures. Let T and N denote the sample sizes in the time and cross sectional dimensions, respectively. Then our analysis corresponds to the case where T is fixed and $N \to \infty$ and thus is geared towards samples where N is large relative to T, as is frequently the case. In more detail, we first demonstrate that our generalizations of the GM approach lead to consistent estimators of the parameters of the disturbance process. Second, we formally derive the large sample distribution of our FGLS estimators. In doing this we demonstrate that the parameters of the disturbance process are nuisance parameters. That is, we show that the true GLS and FGLS estimators have the same large sample distribution.

The model is specified and discussed in Section 2. Results relating to the GM procedures are given in Section 3, and those relating to the FGLS estimators are given in Section 4. Monte Carlo results relating to our suggested GM estimators are given in Section 5. Section 6 contains suggestions for future research. Technical details are relegated to Appendix A.

2. The model

2.1. Notation

In this section we specify the panel data model and then interpret its assumptions. We assume that the data are available for spatial units $i = 1, \ldots, N$ for time periods $t = 1, \ldots, T$. It proves helpful to introduce the following notation: Let $A_N(t)$ be some matrix; then we denote the (i, j)th element of $A_N(t)$ as $a_{ijt,N}$. Similarly, if $b_N(t)$ is a vector, then $b_{it,N}$ denotes the ith element of $b_N(t)$. If matrices or vectors do not depend on the index t or N, then those indices are suppressed on the elements. If A_N is a square matrix, then A_N^{-1} denotes the inverse of A_N. If A_N is singular, then A_N^{-1} should be interpreted as the generalized inverse of A_N. Now let B_N, $N \geq 1$, be some sequence of $kN \times kN$ matrices with k some fixed positive integer. We will then say that the row and column sums of the (sequence of) matrices B_N are bounded uniformly in absolute value if there exists a constant $c < \infty$, that does not depend on N, such that

$$\max_{1 \leq i \leq kN} \sum_{j=1}^{kN} |b_{ij,N}| \leq c \quad \text{and} \quad \max_{1 \leq j \leq kN} \sum_{i=1}^{kN} |b_{ij,N}| \leq c \text{ for all } N \geq 1.$$

As a point of interest, we note that the above condition is identical to the condition that the sequences of the maximum column sum matrix norms and maximum row sum matrix norms of B_N are bounded; cp. Horn and Johnson (1985, pp. 294–295). Finally, we denote the unit vector of dimension $S \times 1$ as e_S.

2.2. Model specification

As remarked, in this paper we consider a linear regression panel data model that allows for the disturbances to be correlated over time and across spatial units. In particular, we

[5]See, e.g., Balestra and Nerlove (1966), Wallace and Hussain (1969), Amemiya (1971) and Nerlove (1971) for seminal contributions to the literature on error components models. Early extensions to the systems case include Baltagi (1980, 1981) and Prucha (1984, 1985). Baltagi (2001) and Hsiao (2003) provide excellent reviews.

assume that in each time period $t = 1, \ldots, T$ the data are generated according to the following model:

$$y_N(t) = X_N(t)\beta + u_N(t), \tag{1}$$

where $y_N(t)$ denotes the $N \times 1$ vector of observation on the dependent variable in period t, $X_N(t)$ denotes the $N \times K$ matrix of observations on exogenous regressors in period t (which may contain the constant term), β is the corresponding $K \times 1$ vector of regression parameters, and $u_N(t)$ denotes the $N \times 1$ vector of disturbance terms. We conditionalize our model on the realized value of the regressors and so we will view $X_N(t)$, $t = 1, \ldots, T$ as matrices of constants.

A widely used approach to model spatial dependence is that of Cliff and Ord (1973, 1981). We follow this approach in modelling the disturbance process in each period as the following first order spatial autoregressive process:

$$u_N(t) = \rho W_N u_N(t) + \varepsilon_N(t), \tag{2}$$

where W_N is an $N \times N$ weighting matrix of known constants which does not involve t, ρ is a scalar autoregressive parameter, and $\varepsilon_N(t)$ is an $N \times 1$ vector of innovations in period t. For reasons of generality, we permit the elements of $y_N(t)$, $X_N(t)$, $u_N(t)$, $\varepsilon_N(t)$ and W_N to depend on N, that is, to form triangular arrays.

Stacking the observations in (1) and (2) we have

$$y_N = X_N \beta + u_N, \tag{3}$$

and

$$u_N = \rho(I_T \otimes W_N)u_N + \varepsilon_N, \tag{4}$$

where $y_N = [y_N'(1), \ldots, y_N'(T)]'$, $X_N = [X_N'(1), \ldots, X_N'(T)]'$, $u_N = [u_N'(1), \ldots, u_N'(T)]'$, and $\varepsilon_N = [\varepsilon_N'(1), \ldots, \varepsilon_N'(T)]'$.

To allow for the innovations to be correlated over time we assume the following error component structure for the innovation vector ε_N:

$$\varepsilon_N = (e_T \otimes I_N)\mu_N + v_N, \tag{5}$$

where μ_N represents the vector of unit specific error components, and $v_N = [v_N'(1), \ldots, v_N'(T)]'$ contains the error components that vary over both the cross-sectional units and time periods. In scalar notation we have

$$\varepsilon_{it,N} = \mu_{i,N} + v_{it,N}. \tag{6}$$

We note that the specification of ε_N corresponds to that of a classical one-way error component model; see, e.g., Baltagi (2001, p. 10). However, in contrast to much of the classical error component literature we group the data by time periods rather than units because this grouping is more convenient for modelling spatial correlation via (2).

We maintain the following assumptions.

Assumption 1. Let T be a fixed positive integer. (a) For all $1 \leqslant t \leqslant T$ and $1 \leqslant i \leqslant N$, $N \geqslant 1$ the error components $v_{it,N}$ are identically distributed with zero mean and variance σ_v^2, $0 < \sigma_v^2 < b_v < \infty$, and finite fourth moments. In addition for each $N \geqslant 1$ and $1 \leqslant t \leqslant T$, $1 \leqslant i \leqslant N$ the error components $v_{it,N}$ are independently distributed. (b) For all $1 \leqslant i \leqslant N$, $N \geqslant 1$ the unit specific error components $\mu_{i,N}$ are identically distributed with zero mean and variance σ_μ^2, $0 < \sigma_\mu^2 < b_\mu < \infty$, and finite fourth moments. In addition for each $N \geqslant 1$ and

$1 \leqslant i \leqslant N$ the unit specific error components $\mu_{i,N}$ are independently distributed. (c) The processes $\{v_{it,N}\}$ and $\{\mu_{i,N}\}$ are independent.

Assumption 2. (a) All diagonal elements of W_N are zero. (b) $|\rho| < 1$. (c) The matrix $I_N - \rho W_N$ is nonsingular.

2.3. Assumption implications

In light of (5) or (6) it is readily seen that Assumption 1 implies $\mathrm{E}\varepsilon_{it,N} = 0$ and

$$
\mathrm{E}(\varepsilon_{it,N}\varepsilon_{js,N}) = \begin{bmatrix} \sigma_\mu^2 + \sigma_v^2 & \text{if } i = j; \ t = s \\ \sigma_\mu^2 & \text{if } i = j; \ t \neq s \\ 0 & \text{otherwise} \end{bmatrix}.
$$

The innovations $\varepsilon_{it,N}$ are autocorrelated over time, but are not spatially correlated across units. In matrix notation we have $\mathrm{E}(\varepsilon_N) = 0$ and the variance–covariance matrix of ε_N is

$$
\begin{aligned}
\Omega_{\varepsilon,N} = \mathrm{E}(\varepsilon_N \varepsilon_N') &= \sigma_\mu^2 (J_T \otimes I_N) + \sigma_v^2 I_{NT} \\
&= \sigma_v^2 Q_{0,N} + \sigma_1^2 Q_{1,N},
\end{aligned} \tag{7}
$$

where $\sigma_1^2 = \sigma_v^2 + T\sigma_\mu^2$ and

$$
Q_{0,N} = \left(I_T - \frac{J_T}{T} \right) \otimes I_N,
$$

$$
Q_{1,N} = \frac{J_T}{T} \otimes I_N, \tag{8}
$$

and where $J_T = e_T e_T'$ is a $T \times T$ matrix of unit elements. The matrices $Q_{0,N}$ and $Q_{1,N}$ are standard transformation matrices utilized in the error component literature, with the appropriate adjustments implied by our adopted ordering of the data; compare, e.g., Baltagi (2001, p. 10). The matrices $Q_{0,N}$ and $Q_{1,N}$ are symmetric and idempotent, and orthogonal to each other. Furthermore $Q_{0,N} + Q_{1,N} = I_{NT}$, $tr(Q_{0,N}) = N(T-1)$, and $tr(Q_{1,N}) = N$. Also $Q_{0,N}(e_T \otimes I_N) = 0$ and $Q_{1,N}(e_T \otimes I_N) = (e_T \otimes I_N)$. Given these properties it is readily seen that

$$
\Omega_{\varepsilon,N}^{-1} = \sigma_v^{-2} Q_{0,N} + \sigma_1^{-2} Q_{1,N}, \tag{9}
$$

which is important in that it facilitates the computation of $\Omega_{\varepsilon,N}^{-1}$ even for large sample sizes N. As an illustration of the transformations implied by $Q_{0,N}$ and $Q_{1,N}$ consider the transformed vectors $Q_{0,N}\varepsilon_N$ and $Q_{1,N}\varepsilon_N$. It is readily seen that the (i,t)th element of those vectors is equal to $\varepsilon_{it,N} - \bar{\varepsilon}_{i.,N}$ and $\bar{\varepsilon}_{i.,N}$, respectively, where $\bar{\varepsilon}_{i.,N} = T^{-1}\sum_{t=1}^{T} \varepsilon_{it,N}$ denotes the sample mean taken over time corresponding to the ith unit. This shows that the $Q_{0,N}$ transformation subtracts unit specific sample means from the original variable. Clearly, $Q_{0,N}\varepsilon_N = Q_{0,N}v_N$, i.e., the $Q_{0,N}$ transformation eliminates the unit specific error components.

Assumption 2, part (a) is a normalization of the model. The nonzero elements of the weighting matrix are typically specified to be those which correspond to units which are related in a meaningful way. Such units are said to be neighbors. Therefore, part (a) of Assumption 2 also implies that no unit is a neighbor to itself. Note that although the elements of W_N are assumed not to depend on t, we do allow them to form triangular

arrays. Among other things, this is consistent with models in which the weighting matrix is row normalized and the number of neighbors a given unit may have depends on the sample size. Parts (b) and (c) of Assumption 2 ensure that the model is closed in that it can be uniquely solved for the disturbance vector $u_N(t)$ in terms of the innovation vector, $\varepsilon_N(t)$, and therefore for the dependent vector, $y_N(t)$, in terms of the exogenous regressor matrix, $X_N(t)$, and the innovation vector, $\varepsilon_N(t)$. Specifically, we have in light of (1) and (2)

$$u_N(t) = (I_N - \rho W_N)^{-1} \varepsilon_N(t),$$

$$y_N(t) = X_N(t)\beta + (I_N - \rho W_N)^{-1} \varepsilon_N(t),$$

or in stacked notation

$$u_N = [I_T \otimes (I_N - \rho W_N)^{-1}] \varepsilon_N,$$

$$y_N = X_N \beta + [I_T \otimes (I_N - \rho W_N)^{-1}] \varepsilon_N. \tag{10}$$

It follows from (10) that the model disturbances are such that $E(u_N) = 0$ and $E(u_N u_N') = \Omega_{u,N}$ where

$$\Omega_{u,N} = [I_T \otimes (I_N - \rho W_N)^{-1}] \Omega_{\varepsilon,N} [I_T \otimes (I_N - \rho W_N')^{-1}]. \tag{11}$$

Clearly our assumptions imply that the model disturbances are, unless $\rho = 0$, both spatially and time-wise autocorrelated. In particular

$$E[u_N(t) u_N'(t)] = (\sigma_\mu^2 + \sigma_v^2)(I_N - \rho W_N)^{-1}(I_N - \rho W_N')^{-1}.$$

We also note that, in general, the elements of $(I_N - \rho W_N)^{-1}$ will form triangular arrays in that they will depend on the sample size, N. Consequently, the elements of u_N and y_N as well as the elements of their variance covariance matrices will also form triangular arrays. Finally, we note that the model disturbances will generally be heteroskedastic.

Anselin (1988, p. 153), and more recently Baltagi and Li (2004) and Baltagi et al. (2003), have also considered error component specifications within the context of spatial models.[6] We note that our specification differs somewhat from theirs in that we allow for spatial interactions involving not only the error components $v_{it,N}$, but also the unit specific error components, μ_i. We note that for $T = 1$ our specification reduces to the standard Cliff–Ord first order spatial autoregressive model.

3. A Generalised Moments Procedure

In the following we define three GM estimators of ρ, σ_v^2 and σ_μ^2, or equivalently of ρ, σ_v^2 and σ_1^2, since in our analysis T is fixed and $N \to \infty$. These GM estimators are defined in terms of six moment conditions. The first of these estimators provides initial estimates of ρ, σ_v^2 and σ_1^2, and is only based on a subset of the moment conditions. The initial estimates for σ_v^2 and σ_1^2 are then used in the formulation of the other two GM estimators. More specifically, our second GM estimator is based on all six moment conditions with the sample moments weighted by an approximation to the inverse of their variance covariance

---

[6]In contrast to this paper, which introduces generalizations of the GM estimation procedure for the spatial autoregressive parameter and aims at establishing the asymptotic properties of the considered estimators, their analysis had a different focus.

matrix.[7] The third GM estimator is based on a simplified weighting scheme. The reason for considering the simplified weighting scheme is that the corresponding estimator is very easy to compute even in large samples. The considered GM estimators generalize, in essence, the GM estimators given in Kelejian and Prucha (1999) for the case of a single cross section, which are based on three moment conditions, giving equal weight to each sample moment. For notational convenience, let

$$\bar{u}_N = (I_T \otimes W_N)u_N,$$

$$\bar{\bar{u}}_N = (I_T \otimes W_N)\bar{u}_N,$$

$$\bar{\varepsilon}_N = (I_T \otimes W_N)\varepsilon_N. \tag{12}$$

3.1. Moment conditions

Given Assumptions 1 and 2 we demonstrate in Appendix A that, for (finite) $T \geqslant 2$

$$\mathrm{E} \begin{bmatrix} \dfrac{1}{N(T-1)}\varepsilon'_N Q_{0,N}\varepsilon_N \\ \dfrac{1}{N(T-1)}\bar{\varepsilon}'_N Q_{0,N}\bar{\varepsilon}_N \\ \dfrac{1}{N(T-1)}\bar{\varepsilon}'_N Q_{0,N}\varepsilon_N \\ \dfrac{1}{N}\varepsilon'_N Q_{1,N}\varepsilon_N \\ \dfrac{1}{N}\bar{\varepsilon}'_N Q_{1,N}\bar{\varepsilon}_N \\ \dfrac{1}{N}\bar{\varepsilon}'_N Q_{1,N}\varepsilon_N \end{bmatrix} = \begin{bmatrix} \sigma_v^2 \\ \sigma_v^2 \dfrac{1}{N} tr(W'_N W_N) \\ 0 \\ \sigma_1^2 \\ \sigma_1^2 \dfrac{1}{N} tr(W'_N W_N) \\ 0 \end{bmatrix}. \tag{13}$$

Our three GM estimators of ρ, σ_v^2, and σ_1^2 are based on these moment relationships. They naturally generalize the moment relationships introduced in Kelejian and Prucha (1998, 1999) for a single cross section to the case of a panel of cross sections. If ε_N were observed, then $\varepsilon'_N Q_{0,N}\varepsilon_N/(N(T-1))$ and $\varepsilon'_N Q_{1,N}\varepsilon_N/N$ represent the (unbiased) analysis of variance estimators of σ_v^2 and σ_1^2, respectively; compare, e.g., Amemiya (1971). In case of a single cross section, i.e., $T = 1$, the innovations ε_N are i.i.d. with variance σ_1^2 (where $\sigma_1^2 = \sigma_v^2 + \sigma_\mu^2$), and in this case it is intuitively clear that we would only be able to identify σ_1^2 and not both σ_v^2 and σ_1^2. This is also reflected by the moment conditions (13) in that if $T = 1$ we have $Q_{0,N} = 0$ and the first three moment conditions become uninformative, and the last three equations reduce to those considered in Kelejian and Prucha (1998, 1999). In the following we maintain $T \geqslant 2$.

Towards defining our GM estimators we note that in light of (4) and (12)

$$\varepsilon_N = u_N - \rho\bar{u}_N,$$

$$\bar{\varepsilon}_N = \bar{u}_N - \rho\bar{\bar{u}}_N. \tag{14}$$

---

[7]Under normality the approximation is perfect.

Substituting these expressions for ε_N and $\bar{\varepsilon}_N$ into (13) we obtain a system of six equations involving the second moments of u_N, $\bar{u}_N$ and $\bar{\bar{u}}_N$. This system involves ρ, σ_v^2, and σ_1^2 and can be expressed as

$$\Gamma_N[\rho, \rho^2, \sigma_v^2, \sigma_1^2]' - \gamma_N = 0, \tag{15}$$

where

$$\Gamma_N = \begin{bmatrix} \gamma_{11,N}^0 & \gamma_{12,N}^0 & \gamma_{13,N}^0 & 0 \\ \gamma_{21,N}^0 & \gamma_{22,N}^0 & \gamma_{23,N}^0 & 0 \\ \gamma_{31,N}^0 & \gamma_{32,N}^0 & \gamma_{33,N}^0 & 0 \\ \gamma_{11,N}^1 & \gamma_{12,N}^1 & 0 & \gamma_{13,N}^1 \\ \gamma_{21,N}^1 & \gamma_{22,N}^1 & 0 & \gamma_{23,N}^1 \\ \gamma_{31,N}^1 & \gamma_{32,N}^1 & 0 & \gamma_{33,N}^1 \end{bmatrix}, \quad \gamma_N = \begin{bmatrix} \gamma_{1,N}^0 \\ \gamma_{2,N}^0 \\ \gamma_{3,N}^0 \\ \gamma_{1,N}^1 \\ \gamma_{2,N}^1 \\ \gamma_{3,N}^1 \end{bmatrix},$$

and $(i = 0, 1)$

$$\gamma_{11,N}^i = \frac{2}{N(T-1)^{1-i}} \mathrm{E} u_N' Q_{i,N} \bar{u}_N, \quad \gamma_{12,N}^i = \frac{-1}{N(T-1)^{1-i}} \mathrm{E} \bar{u}_N' Q_{i,N} \bar{u}_N,$$

$$\gamma_{21,N}^i = \frac{2}{N(T-1)^{1-i}} \mathrm{E} \bar{u}_N' Q_{i,N} \bar{u}_N, \quad \gamma_{22,N}^i = \frac{-1}{N(T-1)^{1-i}} \mathrm{E} \bar{\bar{u}}_N' Q_{i,N} \bar{\bar{u}}_N,$$

$$\gamma_{31,N}^i = \frac{1}{N(T-1)^{1-i}} \mathrm{E} (u_N' Q_{i,N} \bar{\bar{u}}_N + \bar{u}_N' Q_{i,N} \bar{u}_N), \quad \gamma_{32,N}^i = \frac{-1}{N(T-1)^{1-i}} \mathrm{E} \bar{u}_N' Q_{i,N} \bar{\bar{u}}_N,$$

$$\gamma_{13,N}^i = 1, \quad \gamma_{1,N}^i = \frac{1}{N(T-1)^{1-i}} \mathrm{E} u_N' Q_{i,N} u_N,$$

$$\gamma_{23,N}^i = \frac{1}{N} tr(W_N' W_N), \quad \gamma_{2,N}^i = \frac{1}{N(T-1)^{1-i}} \mathrm{E} \bar{u}_N' Q_{i,N} \bar{u}_N,$$

$$\gamma_{33,N}^i = 0, \quad \gamma_{3,N}^i = \frac{1}{N(T-1)^{1-i}} \mathrm{E} u_N' Q_{i,N} \bar{u}_N.$$

The equations underlying our GM procedures are the sample counterparts to the six equations in (15) based on estimated disturbances. In particular, let $\tilde{\beta}_N$ be an estimator of β, and let

$$\tilde{u}_N = y_N - X_N \tilde{\beta}_N,$$

$$\tilde{\bar{u}}_N = (I_T \otimes W_N) \tilde{u}_N,$$

$$\tilde{\bar{\bar{u}}}_N = (I_T \otimes W_N) \tilde{\bar{u}}_N = (I_T \otimes W_N^2) \tilde{u}_N. \tag{16}$$

Then a sample analogue to (15) in terms of $\tilde{u}_N$, $\tilde{\bar{u}}_N$ and $\tilde{\bar{\bar{u}}}_N$ is

$$G_N[\rho, \rho^2, \sigma_v^2, \sigma_1^2]' - g_N = \xi_N(\rho, \sigma_v^2, \sigma_1^2), \tag{17}$$

where

$$G_N = \begin{bmatrix} g^0_{11,N} & g^0_{12,N} & g^0_{13,N} & 0 \\ g^0_{21,N} & g^0_{22,N} & g^0_{23,N} & 0 \\ g^0_{31,N} & g^0_{32,N} & g^0_{33,N} & 0 \\ g^1_{11,N} & g^1_{12,N} & 0 & g^1_{13,N} \\ g^1_{21,N} & g^1_{22,N} & 0 & g^1_{23,N} \\ g^1_{31,N} & g^1_{32,N} & 0 & g^1_{33,N} \end{bmatrix}, \quad g_N = \begin{bmatrix} g^0_{1,N} \\ g^0_{2,N} \\ g^0_{3,N} \\ g^1_{1,N} \\ g^1_{2,N} \\ g^1_{3,N} \end{bmatrix},$$

$$g^i_{11,N} = \frac{2}{N(T-1)^{1-i}}\widetilde{u}'_N Q_{i,N}\widetilde{\overline{u}}_N, \quad g^i_{12,N} = \frac{-1}{N(T-1)^{1-i}}\widetilde{u}'_N Q_{i,N}\widetilde{u}_N,$$

$$g^i_{21,N} = \frac{2}{N(T-1)^{1-i}}\widetilde{\overline{u}}'_N Q_{i,N}\widetilde{\overline{u}}_N, \quad g^i_{22,N} = \frac{-1}{N(T-1)^{1-i}}\widetilde{\overline{u}}'_N Q_{i,N}\widetilde{\overline{u}}_N,$$

$$g^i_{31,N} = \frac{1}{N(T-1)^{1-i}}(\widetilde{u}'_N Q_{i,N}\widetilde{\overline{\overline{u}}}_N + \widetilde{\overline{u}}'_N Q_{i,N}\widetilde{u}_N), \quad g^i_{32,N} = \frac{-1}{N(T-1)^{1-i}}\widetilde{u}'_N Q_{i,N}\widetilde{\overline{\overline{u}}}_N,$$

$$g^i_{13,N} = 1, \quad g^i_{1,N} = \frac{1}{N(T-1)^{1-i}}\widetilde{u}'_N Q_{i,N}\widetilde{u}_N,$$

$$g^i_{23,N} = \frac{1}{N}tr(W'_N W_N), \quad g^i_{2,N} = \frac{1}{N(T-1)^{1-i}}\widetilde{\overline{u}}'_N Q_{i,N}\widetilde{\overline{u}}_N,$$

$$g^i_{33,N} = 0, \quad g^i_{3,N} = \frac{1}{N(T-1)^{1-i}}\widetilde{u}'_N Q_{i,N}\widetilde{\overline{u}}_N,$$

where $\xi_N(\rho,\sigma^2_v,\sigma^2_1)$ is a vector of residuals.

Towards defining our GM procedures it proves helpful to observe that the first three equations in (15) do not involve σ^2_1, while the last three do not involve σ^2_v. In particular, let $\Gamma^0_N = (\gamma^0_{ij,N})_{i,j=1,2,3}$, and $\Gamma^1_N = (\gamma^1_{ij,N})_{i,j=1,2,3}$. Then (15) can be rewritten as

$$\Gamma^0_N[\rho,\rho^2,\sigma^2_v]' - \gamma^0_N = 0,$$

$$\Gamma^1_N[\rho,\rho^2,\sigma^2_1]' - \gamma^1_N = 0. \tag{18}$$

Analogously let $G^0_N = (g^0_{ij,N})_{i,j=1,2,3}$, and $G^1_N = (g^1_{ij,N})_{i,j=1,2,3}$. Then (17) can be rewritten as

$$G^0_N[\rho,\rho^2,\sigma^2_v]' - g^0_N = \xi^0_N(\rho,\sigma^2_v),$$

$$G^1_N[\rho,\rho^2,\sigma^2_1]' - g^1_N = \xi^1_N(\rho,\sigma^2_1), \tag{19}$$

where $\xi_N(\rho,\sigma^2_v,\sigma^2_1) = [\xi^0_N(\rho,\sigma^2_v)', \xi^1_N(\rho,\sigma^2_1)']'$.

3.2. Additional assumptions

We need the following additional assumptions.

Assumption 3. The elements of X_N are bounded uniformly in absolute value by $k_x < \infty$. Furthermore, for $i = 0, 1$, the matrices

$$M_{xx}^i = \lim_{N \to \infty} \frac{1}{NT} X_N^*(\rho)' Q_{i,N} X_N^*(\rho)', \tag{20}$$

with $X_N^*(\rho) = [I_T \otimes (I_N - \rho W_N)] X_N$, are finite, and the matrices

$$\lim_{N \to \infty} \frac{1}{NT} X_N' X_N, \quad \lim_{N \to \infty} \frac{1}{NT} X_N^*(\rho)' X_N^*(\rho), \quad \lim_{N \to \infty} \frac{1}{NT} X_N^*(\rho)' \Omega_{\varepsilon,N}^{-1} X_N^*(\rho) \tag{21}$$

are finite and nonsingular.

Assumption 4. The row and column sums of W_N and $P_N(\rho) = (I_N - \rho W_N)^{-1}$ are bounded uniformly in absolute value by, respectively, $k_w < \infty$ and $k_p < \infty$ where k_p may depend on ρ.

Assumption 5. The smallest eigenvalues of $(\Gamma_N^0)'(\Gamma_N^0)$ and $(\Gamma_N^1)'(\Gamma_N^1)$ are bounded away from zero, i.e., $\lambda_{\min}[(\Gamma_N^i)'(\Gamma_N^i)] \geqslant \lambda_* > 0$ for $i = 1, 2$, where λ_* may depend on ρ, σ_v^2 and σ_1^2.

Conditions like those maintained in Assumption 3 are typical in large sample analysis, see e.g., Kelejian and Prucha (1998, 1999). The large sample properties of the generalized least squares (GLS) estimator of β considered below will involve the limit of $(NT)^{-1} X_N' \Omega_{u,N}^{-1} X_N$. It follows from (11) that

$$\Omega_{u,N}^{-1} = [I_T \otimes (I_N - \rho W_N')] \Omega_{\varepsilon,N}^{-1} [I_T \otimes (I_N - \rho W_N)]. \tag{22}$$

Recalling the expression for $\Omega_{\varepsilon,N}^{-1}$ in (9) and utilizing Assumption 3 it then follows that:

$$\lim_{N \to \infty} (NT)^{-1} X_N' \Omega_{u,N}^{-1} X_N = \lim_{N \to \infty} \frac{1}{NT} X_N^*(\rho)' \Omega_{\varepsilon,N}^{-1} X_N^*(\rho)$$
$$= \sigma_v^{-2} M_{xx}^0 + \sigma_1^{-2} M_{xx}^1. \tag{23}$$

Assumption 4 has two parts. The first relates directly to the weighting matrix and restricts the extent of neighborliness of the cross sectional units. In practice this part should be satisfied if each unit has only a limited number of neighbors, and is itself a neighbor to a limited number of other units—i.e., W_N is sparse. It will also be satisfied if W_N is not sparse, but its elements decline with a distance measure that increases sufficiently rapidly as the sample increases. The second component of Assumption 4 relates to $(I_N - \rho W_N)^{-1}$. From (11) and the subsequent discussion, the variance covariance matrix of the disturbance vector $u_N(t)$ is seen to be proportional to $(I_N - \rho W_N)^{-1}(I_N - \rho W_N')^{-1}$. The second component of Assumption 4 implies that row and column sums of this matrix are bounded uniformly in absolute value, since this property is preserved under matrix multiplication; see Remark A2 in Appendix A. Assumption 4 thus restricts the degree of cross sectional correlation between the model disturbances. For perspective, we note that in virtually all large sample theory it is necessary to restrict the degree of permissible correlations, see e.g., Amemiya (1985, Chapters 3 and 4) or Pötscher and Prucha (1997, Chapters 5 and 6).

Assumption 5 ensures identifiable uniqueness of the parameters ρ, σ_v^2, and σ_1^2. As demonstrated in Appendix A, it implies that also the smallest eigenvalue of $\Gamma_N' \Gamma_N$

is bounded away from zero. A similar assumption was made in Kelejian and Prucha (1998).[8]

3.3. The GM estimators

As remarked, we consider three GM estimators of ρ, σ_v^2, and σ_1^2. The first set of GM estimators will be denoted as $\tilde{\rho}_N$, $\tilde{\sigma}_{v,N}^2$, and $\tilde{\sigma}_{1,N}^2$. These estimators are intended as initial estimators and will be referred to as such. They are based only on a subset of the moment conditions in (15) or (18). In particular, the estimators $\tilde{\rho}_N$ and $\tilde{\sigma}_{v,N}^2$ use only the first three sample moments to estimate ρ and σ_v^2, giving equal weights to each of those moments. They are defined as the unweighted nonlinear least squares estimators based on the first three equations in (17) or (19):

$$(\tilde{\rho}_N, \tilde{\sigma}_{v,N}^2) = \arg\min\{\xi_N^0(\underline{\rho}, \underline{\sigma}_v^2)'\xi_N^0(\underline{\rho}, \underline{\sigma}_v^2), \ \underline{\rho} \in [-a, a], \ \underline{\sigma}_v^2 \in [0, b]\}, \tag{24}$$

where $a \geqslant 1$, $b \geqslant b_v$. By assumption $|\rho| < 1$. If the bound a defining the optimization space is sufficiently large, then $\tilde{\rho}_N$ is essentially the unconstrained nonlinear least squares estimator of ρ.[9] The assumption of a compact optimization space is standard in the literature on nonlinear estimation; see, e.g., Gallant and White (1988) and Pötscher and Prucha (1997). Lee (2004) also maintains a compactness assumption in deriving the asymptotic properties of the ML estimator for the spatial autoregressive parameter in a cross section framework. We note that in contrast to the ML estimator, the objective function of the GM estimator remains well defined for values ρ for which $I_N - \rho\, W_N$ is singular.

For clarity we note that $\xi_N^0(\underline{\rho}, \underline{\sigma}_v^2) = G_N^0[\underline{\rho}, \underline{\rho}^2, \underline{\sigma}_v^2]' - g_N^0$. The elements of the G_N^0 and g_N^0 are observed. The estimator in (24) can thus be computed from a nonlinear regression of the "dependent" variable g_N^0 on the three "independent" variables composing G_N^0, based on three observations on those variables. The nonlinearity arises because of the restriction between the first two elements of the parameter vector $(\underline{\rho}, \underline{\rho}^2, \underline{\sigma}_v^2)$. It follows that the procedure described in (24) can be implemented in any software package which contains a nonlinear least squares option. It should also be clear that an over-parameterized ordinary least squares (OLS) estimator can be obtained as $[(G_N^0)'(G_N^0)]^{-1}(G_N^0)'g_N^0$. In their study dealing with only a single cross section, Kelejian and Prucha (1999) found that their over-parameterized OLS estimator was less efficient than their nonlinear least squares estimator, and so we will not further consider this estimator here.

Given $\tilde{\rho}_N$ and $\tilde{\sigma}_{v,N}^2$ we can estimate σ_1^2 from the fourth moment condition as

$$\begin{aligned}
\tilde{\sigma}_{1,N}^2 &= \frac{1}{N}(\tilde{u}_N - \tilde{\rho}_N\overline{\tilde{u}}_N)' Q_{1,N}(\tilde{u}_N - \tilde{\rho}_N\overline{\tilde{u}}_N) \\
&= g_{1,N}^1 - g_{11,N}^1\tilde{\rho}_N - g_{12,N}^1\tilde{\rho}_N^2.
\end{aligned} \tag{25}$$

The fourth moment condition corresponds to the fourth equation in (17) or (19). We note that the estimator $\tilde{\sigma}_{1,N}^2$ has the interpretation of an analysis of variance estimator for σ_1^2 based on the estimated innovations $\tilde{\varepsilon} = \tilde{u}_N - \tilde{\rho}_N\overline{\tilde{u}}_N$.

---

[8]For a further discussion of the identifiable uniqueness assumption in nonlinear estimation see, e.g., Gallant and White (1988) and Pötscher and Prucha (1997).

[9]We note that in small samples we may obtain estimates $\tilde{\rho}_N$ with $|\tilde{\rho}_N| \geqslant 1$. To force the estimator into the interval $(-1, 1)$ one may consider the reparameterization $\rho = \tan h\,(\varkappa)$ and optimize w.r.t. $\varkappa$.

Theorem 1 below establishes the consistency of the GM estimators $\tilde{\rho}_N$, $\tilde{\sigma}_{v,N}^2$, and $\tilde{\sigma}_{1,N}^2$ defined by (24) and (25). The proof of the theorem is given in Appendix A.

Theorem 1. *Suppose Assumptions 1–5 hold. Then, if $\tilde{\beta}_N$ is a consistent estimator of β, the GM estimators $\tilde{\rho}_N, \tilde{\sigma}_{v,N}^2, \tilde{\sigma}_{1,N}^2$ are consistent for $\rho, \sigma_v^2, \sigma_1^2$ i.e.,*

$$(\tilde{\rho}_N, \tilde{\sigma}_{v,N}^2, \tilde{\sigma}_{1,N}^2) \xrightarrow{P} (\rho, \sigma_v^2, \sigma_1^2) \quad as \ N \to \infty.$$

We show in Appendix A that Assumptions 1–4 imply that the OLS estimator of β corresponding to (3) is consistent, and thus it can be used to calculate the estimated disturbances employed in the GM procedure described by (24) and (25).

It is well known from the literature on generalized method of moments estimators that for purposes of asymptotic efficiency it is optimal to use the inverse of the (properly normalized) variance covariance matrix of the sample moments at the true parameter values as a weighting matrix. In our case the sample moments at the true parameter value are given by the first vector in (13) with the expectations operator suppressed. In Appendix A we derive the following expression for the variance covariance matrix of those sample moments, after multiplication by N:

$$\Xi_N = \begin{bmatrix} \dfrac{1}{T-1}\sigma_v^4 & 0 \\ 0 & \sigma_1^4 \end{bmatrix} \otimes T_W, \tag{26}$$

where

$$T_W = \begin{bmatrix} 2 & 2\,tr\left(\dfrac{W_N'W_N}{N}\right) & 0 \\ 2\,tr\left(\dfrac{W_N'W_N}{N}\right) & 2\,tr\left(\dfrac{W_N'W_NW_N'W_N}{N}\right) & tr\left(\dfrac{W_N'W_N(W_N'+W_N)}{N}\right) \\ 0 & tr\left(\dfrac{W_N'W_N(W_N'+W_N)}{N}\right) & tr\left(\dfrac{W_NW_N+W_N'W_N}{N}\right) \end{bmatrix}.$$

The above variance covariance matrix was derived under the assumption that the innovations are normally distributed. Hence the use of this matrix will not be strictly optimal in the absence of normality. However this matrix has the advantage of relative simplicity. In the absence of normality the variance covariance matrix in (26) can be viewed as an approximation to the more complex true variance covariance matrix. We note that our consistency results below do not depend upon the normality assumption.

The variance covariance matrix Ξ_N depends on σ_v^2 and σ_1^2, and is hence unobserved. It seems natural to define our estimator $\tilde{\Xi}_N$ to be identical to Ξ_N except that σ_v^4 and σ_1^4 are replaced by estimators. Clearly $\tilde{\Xi}_N$ is a consistent estimator for Ξ_N, if the estimators for σ_v^2 and σ_1^2 are consistent. Our second GM estimator is then defined as the nonlinear least squares estimators based on (17) with the sample moments weighted by $\tilde{\Xi}_N^{-1}$:

$$(\hat{\rho}_N, \hat{\sigma}_{v,N}^2, \hat{\sigma}_{1,N}^2) = \arg\min\{\xi_N(\underline{\rho}, \underline{\sigma}_v^2, \underline{\sigma}_1^2)' \tilde{\Xi}_N^{-1} \xi_N(\underline{\rho}, \underline{\sigma}_v^2, \underline{\sigma}_1^2),$$
$$\underline{\rho} \in [-a, a], \ \underline{\sigma}_v^2 \in [0, b], \ \underline{\sigma}_1^2 \in [0, c]\}, \tag{27}$$

where $a \geqslant 1$, $b \geqslant b_v$, $c \geqslant T b_\mu + b_v$. We refer to this estimator in the following as the weighted GM estimator.

A discussion analogous to that after the definition of the initial GM estimator also applies here. In particular, the weighted GM estimator defined in (27) can be computed by running a weighted nonlinear regression of g_N on G_N with $\tilde{\Xi}_N^{-1}$ as the weights matrix. In light of Theorem 1 above we can use, in particular, the initial GM estimators $\tilde{\sigma}_{v,N}^2$ and $\tilde{\sigma}_{1,N}^2$ to construct a consistent estimator $\tilde{\Xi}_N$ for Ξ_N.

The next theorem establishes the consistency of the weighted GM estimator defined by (27). The proof of the theorem is given in Appendix A.

Theorem 2. *Suppose Assumptions 1–5 hold and the smallest and largest eigenvalues of the matrices Ξ_N^{-1} satisfy $0 < \bar{\lambda}_* \leqslant \lambda_{\min}(\Xi_N^{-1}) \leqslant \lambda_{\max}(\Xi_N^{-1}) \leqslant \bar{\lambda}_{**} < \infty$. Suppose furthermore that $\tilde{\beta}_N$ and $\tilde{\Xi}_N$ are consistent estimators of β and Ξ_N, respectively. Then, the GM estimators $\hat{\rho}_N, \hat{\sigma}_{v,N}^2, \hat{\sigma}_{1,N}^2$ defined by (27) are consistent for $\rho, \sigma_v^2, \sigma_1^2$, i.e.,*

$$(\hat{\rho}_N, \hat{\sigma}_{v,N}^2, \hat{\sigma}_{1,N}^2) \xrightarrow{P} (\rho, \sigma_v^2, \sigma_1^2) \quad \text{as } N \to \infty.$$

The assumptions relating to the eigenvalues of Ξ_N^{-1} together with Assumption 5 ensure the identifiably uniqueness of the parameters $\rho, \sigma_v^2, \sigma_1^2$. They also ensure that the elements of Ξ_N^{-1} are O(1).

The third GM estimator considered here is motivated mostly by computational considerations. To calculate the estimator $\tilde{\Xi}_N$ considered above we have to compute the elements of T_W. While T_W can be computed accurately even for large n, nevertheless it is readily seen that the computation of, in particular, the $(2,2)$ element of that matrix involves a computational count of up to O(n^3). Thus it seems of interest to also consider an alternative estimator that is computationally simpler. Let

$$\Upsilon_N = \begin{bmatrix} \dfrac{1}{T-1}\sigma_v^4 & 0 \\ 0 & \sigma_1^4 \end{bmatrix} \otimes I_3 \tag{28}$$

and let $\tilde{\Upsilon}_N$ be the corresponding estimator where σ_v^2 and σ_1^2 are replaced by estimators. That is, Υ_N and $\tilde{\Upsilon}_N$ have the same form as Ξ_N and $\tilde{\Xi}_N$, except that T_W is replaced by the identity matrix I_3. Our third GM estimator is then defined as the nonlinear least squares estimators based on (17) where the sample moments are weighted by $\tilde{\Upsilon}_N^{-1}$:

$$(\check{\rho}_N, \check{\sigma}_{v,N}^2, \check{\sigma}_{1,N}^2) = \arg\min\{\xi_N(\underline{\rho}, \underline{\sigma}_v^2, \underline{\sigma}_1^2)' \tilde{\Upsilon}_N^{-1} \xi_N(\underline{\rho}, \underline{\sigma}_v^2, \underline{\sigma}_1^2)$$
$$\underline{\rho} \in [-a, a], \quad \underline{\sigma}_v^2 \in [0, b], \quad \underline{\sigma}_1^2 \in [0, c]\}, \tag{29}$$

where, again, $a \geqslant 1$, $b \geqslant b_v$, $c \geqslant T b_\mu + b_v$. We refer to this estimator as the partially weighted GM estimator. It places the same weight on each of the first three moment equations, and the same weight on each of the last three moment equations, but the weight given to the first three moment equations is different than that given to the last three moment equations. This weighting scheme is motivated by the estimator considered in Kelejian and Prucha (1998, 1999), which is based on three equally weighted moment equations. Those studies report that the estimator exhibits a small sample behavior that is quite similar to the ML estimator both for a set of idealized and real world weighting matrices W_N. This suggest that, at least for some matrices W_N, the weighting scheme implied by T_W^{-1} as compared to the identity matrix is of lesser importance. The Monte Carlo results given

below provide some support for this conjecture also in the present context of panel data. In light of Theorem 1 above we can use again the initial GM estimators $\tilde{\sigma}_{v,N}^2$ and $\tilde{\sigma}_{1,N}^2$ to construct a consistent estimator $\tilde{\Upsilon}_N$ for Υ_N.

The next theorem establishes the consistency of the partially weighted GM estimator defined by (29). The proof of the theorem is given in Appendix A.

Theorem 3. *Suppose Assumptions 1–5 hold. Suppose furthermore that $\tilde{\beta}_N$ and $\tilde{\Upsilon}_N^{-1}$ are consistent estimators of β and Υ_N, respectively. Then, the GM estimators $\check{\rho}_N, \check{\sigma}_{v,N}^2, \check{\sigma}_{1,N}^2$ defined by (29) are consistent for $\rho, \sigma_v^2, \sigma_1^2$, i.e.,*

$$(\check{\rho}_N, \check{\sigma}_{v,N}^2, \check{\sigma}_{1,N}^2) \xrightarrow{P} (\rho, \sigma_v^2, \sigma_1^2) \quad \text{as } N \to \infty.$$

A more extreme simplified weighting scheme would be to give equal weight to all six moment equations. The corresponding GM estimator is also readily seen to be consistent. However, preliminary Monte Carlo results suggest that the small sample behavior of this estimator can be substantially worse than that of the other three estimators, and hence we do not further consider this estimator in this study.

4. Spatial GLS estimation

Consider again the regression model in (3)–(5). The true GLS estimator of β is given by

$$\begin{aligned}
\hat{\beta}_{\text{GLS},N} &= \{X_N'[\Omega_{u,N}^{-1}(\rho, \sigma_v^2, \sigma_1^2)]X_N\}^{-1} X_N'[\Omega_{u,N}^{-1}(\rho, \sigma_v^2, \sigma_1^2)]y_N \\
&= \{X_N^*(\rho)'[\Omega_{\varepsilon,N}^{-1}(\sigma_v^2, \sigma_1^2)]X_N^*(\rho)\}^{-1} X_N^*(\rho)'[\Omega_{\varepsilon,N}^{-1}(\sigma_v^2, \sigma_1^2)]y_N^*(\rho),
\end{aligned} \tag{30}$$

where

$$y_N^*(\rho) = [I_T \otimes (I_N - \rho W_N)]y_N,$$

$$X_N^*(\rho) = [I_T \otimes (I_N - \rho W_N)]X_N, \tag{31}$$

and where we have denoted the explicit dependence of $\Omega_{u,N}^{-1}$ and $\Omega_{\varepsilon,N}^{-1}$ on ρ and/or σ_v^2 and σ_1^2. The second expression in (30) is obtained by utilizing the expression for $\Omega_{u,N}$ given in (11). The variables $y_N^*(\rho)$ and $X_N^*(\rho)$ can be viewed as the result of a spatial Cochrane–Orcutt type transformation of the original model. More specifically, premultiplication of (3) and (4) with $I_T \otimes (I_N - \rho W_N)$ yields[10]

$$y_N^*(\rho) = X_N^*(\rho)\beta + \varepsilon_N. \tag{32}$$

In light of (9) and the properties of $Q_{0,N}$ and $Q_{1,N}$ we have $\Omega_{\varepsilon,N}^{-1} = \Omega_{\varepsilon,N}^{-1/2}\Omega_{\varepsilon,N}^{-1/2}$ with

$$\Omega_{\varepsilon,N}^{-1/2} = \sigma_v^{-1}Q_{0,N} + \sigma_1^{-1}Q_{1,N}. \tag{33}$$

Guided by the classical error component literature, we note that a convenient way of computing the GLS estimator $\hat{\beta}_{\text{GLS},N}$ is to further transform the model in (32) by premultiplying it by $\Omega_{\varepsilon,N}^{-1/2}$, or equivalently by

$$\sigma_v \Omega_{\varepsilon,N}^{-1/2} = I_{NT} - \theta Q_{1,N}, \tag{34}$$

---

[10]This is, of course, equivalent to premultiplying (1) and (2) in each period t by $(I_N - \rho W_N)$ and then stacking the data.

where $\theta = 1 - \sigma_v/\sigma_1$. The GLS estimator of β is then identical to the OLS estimator of β computed from the resulting transformed model. We note that the transformation based on (34) simply amounts to subtracting from each variable its sample mean over time multiplied by θ.

Let $\check{\rho}_N, \check{\sigma}^2_{v,N}, \check{\sigma}^2_{1,N}$ be estimators of $\rho, \sigma^2_{v,N}, \sigma^2_{1,N}$. The corresponding feasible GLS estimator of β, say $\hat{\beta}_{\text{FGLS},N}$, is then obtained by replacing $\rho, \sigma^2_{v,N}, \sigma^2_{1,N}$ by those estimators in the expression for the GLS estimator, i.e.,

$$
\begin{aligned}
\hat{\beta}_{\text{FGLS},N} = \{X^*_N(\check{\rho}_N)'[\Omega^{-1}_{\varepsilon,N}(\check{\sigma}^2_{v,N}, \check{\sigma}^2_{1,N})]X^*_N(\check{\rho}_N)\}^{-1} \\
\times X^*_N(\check{\rho}_N)'[\Omega^{-1}_{\varepsilon,N}(\check{\sigma}^2_{v,N}, \check{\sigma}^2_{1,N})]y^*_N(\check{\rho}_N).
\end{aligned}
\tag{35}
$$

This estimator can again be computed conveniently as the OLS estimator after transforming the model in a way analogously to what was described for the GLS estimator.

The next theorem establishes consistency and asymptotic normality of the true and feasible GLS estimators. The proof of the theorem is given in Appendix A.

Theorem 4. *Given Assumptions* 1–4 *hold*:

(a) *Then*

$$
(NT)^{1/2}[\hat{\beta}_{\text{GLS},N} - \beta] \xrightarrow{D} N\{0, \Psi\} \quad as\ N \to \infty,
$$

with

$$
\Psi = [\sigma^{-2}_v M^0_{xx} + \sigma^{-2}_1 M^1_{xx}]^{-1}.
$$

(b) *If* $\check{\rho}_N, \check{\sigma}^2_{v,N}, \check{\sigma}^2_{1,N}$ *are (any) consistent estimators of* $\rho, \sigma^2_v, \sigma^2_1$, *then*

$$
(NT)^{1/2}[\hat{\beta}_{\text{GLS},N} - \hat{\beta}_{\text{FGLS},N}] \xrightarrow{P} 0 \quad as\ N \to \infty.
$$

(c) *Furthermore*,

$$
\check{\Psi}_N - \Psi \xrightarrow{P} 0 \quad as\ N \to \infty,
$$

where

$$
\check{\Psi}_N = \left\{ \frac{1}{NT} X^*_N(\check{\rho}_N)'[\Omega^{-1}_{\varepsilon,N}(\check{\sigma}^2_{v,N}, \check{\sigma}^2_{1,N})]X^*_N(\check{\rho}_N) \right\}^{-1}.
$$

Part (a) implies that the true GLS estimator, $\hat{\beta}_{\text{GLS},N}$, is consistent and asymptotically normal. Part (b) implies that if the feasible GLS estimator, $\hat{\beta}_{\text{FGLS},N}$, is based on any consistent estimators of $\rho, \sigma^2_v, \sigma^2_1$, then $\hat{\beta}_{\text{GLS},N}$ and $\hat{\beta}_{\text{FGLS},N}$ are asymptotically equivalent, and hence $\hat{\beta}_{\text{FGLS},N}$ is also consistent and asymptotically normal with the same asymptotic distribution as $\hat{\beta}_{\text{GLS},N}$. Therefore, $\rho, \sigma^2_v, \sigma^2_1$ are "nuisance" parameters. Together with part (c), these results suggest that small sample inferences can be based on the approximation

$$
\hat{\beta}_{\text{FGLS},N} \dot\sim N(\beta, (NT)^{-1}\check{\Psi}_N).
$$

In Theorems 2 and 3 we establish the consistency of the weighted and partially weighted GM estimators for ρ, σ_v^2, and σ_1^2, respectively. Therefore, these estimators can be used to construct the feasible GLS estimator considered in Theorem 4.

5. A Monte Carlo investigation

In the following we report on a small Monte Carlo study of the small sample properties of the estimators suggested in this paper. We consider the three suggested GM estimators of the spatial autoregressive parameter ρ and variances σ_μ^2 and σ_1^2, and corresponding feasible GLS estimators for β. We also consider iterated versions of these estimators. For purposes of comparison, we also consider the ML estimator of these parameters. A larger Monte Carlo study relating to a wider set of experiments than those described below is left for future research.

In all of our Monte Carlo experiments $N = 100$ and $T = 5$. The data are generated according to (3)–(5) with the regressor matrix containing two regressors x_1 and x_2 with corresponding parameters β_1 and β_2. In our illustrative experiments we take x_1 to be the constant term. The observations on x_2 corresponds to per capita income (measured in $1000) in 100 contiguous counties in Virginia over the periods 1996–2000. The parameters β_1 and β_2 are taken to be equal to one. We assume that both of the error components, μ_N and v_N, are normally distributed. The generation of the disturbance vector u_N requires the specification of σ_μ^2, σ_v^2, ρ, and the weighting matrix W_N. In all of our Monte Carlo experiments $\sigma_\mu^2 = \sigma_v^2 = 1$ and thus $\sigma_1^2 = 6$.[11] As in Kelejian and Prucha (1999), we consider seven values of ρ, namely -0.9, -0.5, -0.25, 0, 0.25, 0.5, and 0.9. Also, as in the earlier study we consider three weighting matrices which essentially differ in their degrees of sparseness. The first matrix is such that its ith row, $1 < i < N$, has nonzero elements in positions $i+1$ and $i-1$ so that the ith element of $u_N(t)$, $1 < i < N$, is directly related to the one immediately after it, and the one immediately before it. This matrix is defined in a circular world so that the nonzero elements in rows 1 and N are, respectively, in positions $(1,2)$, $(1,N)$ and $(N,1)$, $(N,N-1)$. This matrix is row normalized so that all of it's nonzero elements are equal to $\frac{1}{2}$. As in Kelejian and Prucha (1999, p. 520) we refer to this specification of the weighting matrix as ''1 ahead and 1 behind.'' The other two weighting matrices we consider are defined in a corresponding way as 3 ahead and 3 behind, and 5 ahead and 5 behind. The nonzero elements in these matrices are, respectively, $\frac{1}{6}$ and $\frac{1}{10}$. In Tables 1–5, below we reference these three weighting matrices by $J = 2, 6, 10$, where J is the number of nonzero elements in a given row. To summarize, our experimental design contains 21 cases, which result from 7 selections of ρ and 3 specifications of the weighting matrix. We have also calculated R^2 values (obtained from a simple least squares regression of y on the regressors). The median R^2 values for respective experiments ranges between 0.61–0.95, 0.77–0.95 and 0.83–0.95 for $J = 2$, 6 and 10, respectively.

Recall that $(\tilde{\rho}, \tilde{\sigma}_v^2, \tilde{\sigma}_v^2)$, $(\hat{\rho}, \hat{\sigma}_v^2, \hat{\sigma}_1^2)$, and $(\breve{\rho}, \breve{\sigma}_v^2, \breve{\sigma}_1^2)$, defined in (24), (27), and (29) denote the initial (unweighted), the weighted and the partially weighted GM estimators for $(\rho, \sigma_v^2, \sigma_1^2)$,

---

[11]We considered other values of the ratio σ_μ^2/σ_v^2 but the results were qualitatively similar to those reported below for the case $\sigma_\mu^2 = \sigma_v^2 = 1$. Again, to conserve space these results are not reported, but are available upon request.

Table 1
RMSEs of the estimators of ρ

Parameter values		RMSE						
J	ρ	$\hat{\rho}^{\mathrm{ML}}$	$\tilde{\rho}$	$\hat{\rho}$	$\check{\rho}$	$\tilde{\rho}^{(1)}$	$\hat{\rho}^{(1)}$	$\check{\rho}^{(1)}$
2	−0.9	0.0111	0.0191	0.0149	0.0182	0.0150	0.0122	0.0133
2	−0.5	0.0344	0.0416	0.0354	0.0355	0.0409	0.0351	0.0353
2	−0.25	0.0420	0.0496	0.0422	0.0419	0.0493	0.0415	0.0421
2	0	0.0445	0.0522	0.0446	0.0437	0.0520	0.0450	0.0444
2	0.25	0.0420	0.0499	0.0431	0.0416	0.0502	0.0431	0.0420
2	0.5	0.0350	0.0420	0.0359	0.0354	0.0421	0.0363	0.0354
2	0.9	0.0111	0.0157	0.0131	0.0142	0.0158	0.0125	0.0132
6	−0.9	0.0857	0.1088	0.0872	0.0975	0.1080	0.0872	0.0979
6	−0.5	0.0902	0.1037	0.0861	0.0925	0.1051	0.0890	0.0938
6	−0.25	0.0846	0.0962	0.0825	0.0865	0.0989	0.0826	0.0873
6	0	0.0757	0.0869	0.0750	0.0769	0.0883	0.0748	0.0776
6	0.25	0.0653	0.0739	0.0648	0.0652	0.0745	0.0649	0.0657
6	0.5	0.0503	0.0572	0.0504	0.0498	0.0573	0.0511	0.0499
6	0.9	0.0160	0.0188	0.0175	0.0171	0.0188	0.0171	0.0166
10	−0.9	0.1336	0.1563	0.1431	0.1458	0.1718	0.1402	0.1530
10	−0.5	0.1265	0.1489	0.1302	0.1317	0.1564	0.1301	0.1355
10	−0.25	0.1184	0.1360	0.1187	0.1186	0.1410	0.1212	0.1213
10	0	0.1049	0.1197	0.1049	0.1032	0.1223	0.1068	0.1034
10	0.25	0.0879	0.1007	0.0844	0.0848	0.1005	0.0849	0.0857
10	0.5	0.0657	0.0753	0.0646	0.0637	0.0753	0.0641	0.0646
10	0.9	0.0207	0.0243	0.0206	0.0216	0.0241	0.0205	0.0211
Column averages		0.0641	0.0756	0.0647	0.0660	0.0766	0.0648	0.0666

$\hat{\rho}^{\mathrm{ML}}$ is the maximum likelihood estimator; $\tilde{\rho}$ the initial (unweighted) GM estimator; $\hat{\rho}$ the weighted GM estimator; $\check{\rho}$ the partially weighted GM estimator; $\tilde{\rho}^{(1)}$ the initial (unweighted) GM estimator, iterated once; $\hat{\rho}^{(1)}$ the weighted GM estimator, iterated once; and $\check{\rho}^{(1)}$ the partially weighted GM estimator, iterated once.
$N = 100$, $T = 5$, and $(\sigma_\mu^2, \sigma_v^2, \sigma_1^2, \beta_1, \beta_2) = (1, 1, 6, 1, 1)$.

respectively.[12] These estimators are computed from OLS residuals. Based on the respective sets of GM estimators for $(\rho, \sigma_v^2, \sigma_1^2)$ we compute feasible GLS estimators for β_1 and β_2 by substituting into the formula for the feasible GLS estimator in (35) the respective sets of GM estimators for $(\rho, \sigma_v^2, \sigma_1^2)$. We denote the corresponding feasible GLS estimators as $\tilde{\beta}_{i,\mathrm{FGLS}}$, $\hat{\beta}_{i,\mathrm{FGLS}}$ and $\check{\beta}_{i,\mathrm{FGLS}}$, $i = 1, 2$, respectively. We also compute iterated version of these estimators. Towards computing these iterated estimators we first compute feasible GLS residuals corresponding to $\tilde{\beta}_{i,\mathrm{FGLS}}$, $\hat{\beta}_{i,\mathrm{FGLS}}$ and $\check{\beta}_{i,\mathrm{FGLS}}$, $i = 1, 2$, respectively. We then use those residuals to compute new sets of the initial (unweighted), the weighted, and the partially weighted GM estimators for $(\rho, \sigma_v^2, \sigma_1^2)$. These estimators will be denoted as $(\tilde{\rho}^{(1)}, \tilde{\sigma}_v^{2(1)}, \tilde{\sigma}_v^{2(1)})$, $(\hat{\rho}^{(1)}, \hat{\sigma}_v^{2(1)}, \hat{\sigma}_1^{2(1)})$, and $(\check{\rho}^{(1)}, \check{\sigma}_v^{2(1)}, \check{\sigma}_1^{2(1)})$.[13] In computing the matrices $\tilde{\Xi}_N$ and

---

[12] For simplicity of notation we drop the index corresponding to the sample size here and in the following.

[13] We have called the GM estimators $(\tilde{\rho}, \tilde{\sigma}_v^2, \tilde{\sigma}_v^2)$, which correspond to the first four unweighted moment conditions, the initial (unweighted) GM estimators. In slight abuse of language, we call the GM estimators $(\tilde{\rho}^{(1)}, \tilde{\sigma}_v^{2(1)}, \tilde{\sigma}_v^{2(1)})$, which correspond to the first four unweighted moment conditions based on feasible GLS disturbances corresponding to $\tilde{\beta}_{i,\mathrm{FGLS}}$, the once iterated initial (unweighted) GM estimators.

Table 2
RMSEs of the estimators of σ_v^2

Parameter values		RMSE						
J	ρ	$\hat{\sigma}_v^{2,\text{ML}}$	$\tilde{\sigma}_v^2$	$\hat{\sigma}_v^2$	$\breve{\sigma}_v^2$	$\tilde{\sigma}_v^{2(1)}$	$\hat{\sigma}_v^{2(1)}$	$\breve{\sigma}_v^{2(1)}$
2	−0.9	0.0761	0.1238	0.1046	0.1231	0.0846	0.0772	0.0846
2	−0.5	0.0734	0.0749	0.0728	0.0735	0.0736	0.0732	0.0736
2	−0.25	0.0730	0.0739	0.0730	0.0735	0.0737	0.0738	0.0737
2	0	0.0725	0.0720	0.0724	0.0719	0.0725	0.0723	0.0725
2	0.25	0.0726	0.0748	0.0731	0.0749	0.0748	0.0731	0.0748
2	0.5	0.0742	0.0777	0.0753	0.0758	0.0776	0.0751	0.0776
2	0.9	0.0756	0.0876	0.0798	0.0836	0.0864	0.0790	0.0864
6	−0.9	0.0774	0.0767	0.0743	0.0738	0.0764	0.0756	0.0762
6	−0.5	0.0743	0.0735	0.0733	0.0734	0.0755	0.0743	0.0738
6	−0.25	0.0737	0.0730	0.0727	0.0734	0.0746	0.0731	0.0739
6	0	0.0732	0.0739	0.0741	0.0739	0.0732	0.0731	0.0731
6	0.25	0.0733	0.0730	0.0734	0.0731	0.0733	0.0733	0.0735
6	0.5	0.0727	0.0722	0.0729	0.0729	0.0724	0.0726	0.0730
6	0.9	0.0740	0.0743	0.0735	0.0746	0.0734	0.0727	0.0747
10	−0.9	0.0739	0.0751	0.0755	0.0745	0.0751	0.0740	0.0758
10	−0.5	0.0732	0.0725	0.0732	0.0721	0.0738	0.0736	0.0736
10	−0.25	0.0730	0.0730	0.0728	0.0723	0.0745	0.0736	0.0733
10	0	0.0732	0.0739	0.0737	0.0735	0.0737	0.0733	0.0733
10	0.25	0.0734	0.0731	0.0736	0.0735	0.0729	0.0735	0.0734
10	0.5	0.0723	0.0714	0.0720	0.0722	0.0713	0.0719	0.0722
10	0.9	0.0727	0.0720	0.0731	0.0725	0.0723	0.0727	0.0729
Column averages		0.0737	0.0768	0.0752	0.0763	0.0750	0.0739	0.0750

$\hat{\sigma}_v^{2,\text{ML}}$ is the maximum likelihood estimator; $\tilde{\sigma}_v^2$ the initial (unweighted) GM estimator; $\hat{\sigma}_v^2$ the weighted GM estimator; $\breve{\sigma}_v^2$ the partially weighted GM estimator; $\tilde{\sigma}_v^{2(1)}$ the initial (unweighted) GM estimator, iterated once; $\hat{\sigma}_v^{2(1)}$ the weighted GM estimator, iterated once; and $\breve{\sigma}_v^{2(1)}$ the partially weighted GM estimator, iterated once. $N = 100$, $T = 5$, and $(\sigma_\mu^2, \sigma_v^2, \sigma_1^2, \beta_1, \beta_2) = (1, 1, 6, 1, 1)$.

$\tilde{\Upsilon}_N$ that enter into the objective function for the weighted and the partially weighted GM estimators we use $(\hat{\sigma}_v^2, \hat{\sigma}_1^2)$, and $(\breve{\sigma}_v^2, \breve{\sigma}_1^2)$. Using the (once) iterated initial (unweighted), the weighted, and the partially weighted GM estimators for $(\rho, \sigma_v^2, \sigma_1^2)$ we compute the (once) iterated feasible GLS estimators which we denote as $\tilde{\beta}_{i,\text{FGLS}}^{(1)}$, $\hat{\beta}_{i,\text{FGLS}}^{(1)}$ and $\breve{\beta}_{i,\text{FGLS}}^{(1)}$, $i = 1, 2$.

Table 1 contains results on a measure of dispersion relating to the small sample distributions of the GM estimators $\tilde{\rho}, \hat{\rho}, \breve{\rho}$, the corresponding iterated GM estimators, and the ML estimator, $\hat{\rho}^{\text{ML}}$, for each of 21 cases. Our adopted measure of dispersion is closely related to the standard measure of the root mean squared error (RMSE), but is based on quantiles rather than moments, because, unlike moments, quantiles are assured to exit. For ease of presentation, we also refer to our measure as the RMSE. It is defined as

$$RMSE = \left[bias^2 + \left[\frac{IQ}{1.35} \right]^2 \right]^{1/2},$$

where *bias* is the difference between the median and the true value of ρ, and IQ is the interquantile range defined as $c_1 - c_2$, where c_1 is the 0.75 quantile and c_2 is the 0.25

Table 3
RMSEs of the estimators of σ_1^2

Parameter values		RMSE						
J	ρ	$\hat{\sigma}_1^{2,\mathrm{ML}}$	$\tilde{\sigma}_1^2$	$\hat{\sigma}_1^2$	$\check{\sigma}_1^2$	$\tilde{\sigma}_1^{2(1)}$	$\hat{\sigma}_1^{2(1)}$	$\check{\sigma}_1^{2(1)}$
2	−0.9	0.8269	1.3474	1.2562	1.3364	0.8587	0.8217	0.8374
2	−0.5	0.8222	0.8340	0.8267	0.8269	0.8450	0.8168	0.8401
2	−0.25	0.8298	0.8445	0.8270	0.8431	0.8381	0.8307	0.8528
2	0	0.8357	0.8288	0.8298	0.8377	0.8267	0.8300	0.8454
2	0.25	0.8374	0.8336	0.8409	0.8494	0.8341	0.8353	0.8527
2	0.5	0.8411	0.8098	0.8322	0.8446	0.8312	0.8445	0.8679
2	0.9	0.8350	1.1895	1.1898	1.2240	0.8497	0.8320	0.8664
6	−0.9	0.8306	0.8455	0.8414	0.8501	0.8470	0.8265	0.8654
6	−0.5	0.8410	0.8457	0.8336	0.8344	0.8366	0.8413	0.8379
6	−0.25	0.8335	0.8406	0.8355	0.8175	0.8428	0.8353	0.8269
6	0	0.8268	0.8196	0.8183	0.8124	0.8199	0.8280	0.8221
6	0.25	0.8181	0.8073	0.8205	0.8070	0.8124	0.8193	0.8077
6	0.5	0.8265	0.8080	0.8244	0.8121	0.8096	0.8324	0.8192
6	0.9	0.8425	0.9532	0.9442	0.9475	0.8284	0.8368	0.8238
10	−0.9	0.8535	0.8522	0.8472	0.8536	0.8447	0.8493	0.8479
10	−0.5	0.8307	0.8301	0.8319	0.8269	0.8349	0.8426	0.8337
10	−0.25	0.8324	0.8243	0.8234	0.8189	0.8262	0.8305	0.8252
10	0	0.8202	0.8161	0.8241	0.8197	0.8267	0.8288	0.8221
10	0.25	0.8286	0.8169	0.8242	0.8254	0.8216	0.8288	0.8272
10	0.5	0.8305	0.8250	0.8312	0.8288	0.8339	0.8321	0.8297
10	0.9	0.8415	0.8854	0.8826	0.8771	0.8352	0.8348	0.8313
Column averages		0.8236	0.8788	0.8755	0.8806	0.8335	0.8323	0.8373

$\hat{\sigma}_1^{2,\mathrm{ML}}$ is the maximum likelihood estimator; $\tilde{\sigma}_1^2$ the initial (unweighted) GM estimator; $\hat{\sigma}_1^2$ the weighted GM estimator; $\check{\sigma}_1^2$ the partially weighted GM estimator; $\tilde{\sigma}_1^{2(1)}$ the initial (unweighted) GM estimator, iterated once; $\hat{\sigma}_1^{2(1)}$ the weighted GM estimator, iterated once; and $\check{\sigma}_1^{2(1)}$ the partially weighted GM estimator, iterated once. $N = 100$, $T = 5$, and $(\sigma_\mu^2, \sigma_v^2, \sigma_1^2, \beta_1, \beta_2) = (1, 1, 6, 1, 1)$.

quantile. Clearly, if the distribution is normal the median is the mean and, aside from a slight rounding error, $IQ/1.35$ is the standard deviation. In this case our measure reduces to the standard RMSE measure. The figures in Table 1 are Monte Carlo estimates of the RMSE based on a 1000 replications for each of the indicated 21 cases. The setup of Tables 2–5 is analogous.

A glance at Table 1 suggests that, on average, the RMSEs of the ML, the weighted GM, the partially weighted GM estimators, and of the iterated GM estimators are quite similar. The RMSEs of the initial (unweighted) GM estimator are, on average, approximately 17% and 14% larger, respectively, than those of the weighted and partially weighted GM estimators. The results relating to the weighted and partially weighted GM estimators are encouraging and suggest that the computational benefits associated with their use, which are considerable in large samples, does not appear to have much cost in terms of efficiency. For example, in terms of column averages, the RMSEs of the weighted and partially weighted GM estimator are, approximately, only 1% and 4% larger than those of the ML estimator. The RMSEs of the iterated GM estimators for ρ are on average very similar to those of the noniterated GM estimators.

Table 4
RMSEs of the feasible GLS estimators of β_1

Parameter values		RMSE						
J	ρ	$\widehat{\beta}_1^{\text{ML}}$	$\tilde{\beta}_1$	$\hat{\beta}_1$	$\breve{\beta}_1$	$\tilde{\beta}_1^{(1)}$	$\hat{\beta}_1^{(1)}$	$\breve{\beta}_1^{(1)}$
2	−0.9	0.2577	0.2600	0.2597	0.2600	0.2581	0.2577	0.2603
2	−0.5	0.2872	0.2885	0.2874	0.2875	0.2889	0.2873	0.2876
2	−0.25	0.3257	0.3248	0.3259	0.3255	0.3250	0.3249	0.3259
2	0	0.3546	0.3551	0.3559	0.3572	0.3542	0.3558	0.3569
2	0.25	0.3709	0.3715	0.3703	0.3731	0.3721	0.3709	0.3732
2	0.5	0.4262	0.4232	0.4234	0.4279	0.4243	0.4252	0.4252
2	0.9	1.1412	1.1330	1.1340	1.1365	1.1379	1.1412	1.1382
6	−0.9	0.2776	0.2729	0.2757	0.2596	0.2740	0.2764	0.2740
6	−0.5	0.3094	0.3104	0.3131	0.3101	0.3112	0.3114	0.3102
6	−0.25	0.3271	0.3321	0.3276	0.3278	0.3313	0.3268	0.3260
6	0	0.3545	0.3500	0.3559	0.3521	0.3534	0.3560	0.3522
6	0.25	0.3777	0.3822	0.3823	0.3811	0.3820	0.3820	0.3807
6	0.5	0.4339	0.4371	0.4363	0.4356	0.4371	0.4377	0.4359
6	0.9	1.1199	1.1176	1.1172	1.1241	1.1195	1.1179	1.1193
10	−0.9	0.2672	0.2730	0.2649	0.2689	0.2656	0.2679	0.2663
10	−0.5	0.3057	0.3084	0.3066	0.3090	0.3069	0.3065	0.3082
10	−0.25	0.3267	0.3287	0.3303	0.3278	0.3299	0.3291	0.3278
10	0	0.3611	0.3568	0.3561	0.3549	0.3600	0.3571	0.3584
10	0.25	0.3870	0.3825	0.3833	0.3847	0.3853	0.3837	0.3859
10	0.5	0.4276	0.4291	0.4287	0.4297	0.4293	0.4288	0.4292
10	0.9	1.1499	1.1475	1.1486	1.1490	1.1462	1.1465	1.1474
Column averages		0.4566	0.4564	0.4564	0.4566	0.4568	0.4567	0.4566

$\widehat{\beta}_1^{\text{ML}}$ is the maximum likelihood estimator; $\tilde{\beta}_1$ the feasible GLS estimator based on initial GM estimators; $\hat{\beta}_1$ the feasible GLS estimator based on weighted GM estimators; $\breve{\beta}_1$ the feasible GLS estimator based on partially weighted GM estimators; $\tilde{\beta}_1^{(1)}$ the feasible GLS estimator based on initial GM estimators, iterated once; $\hat{\beta}_1^{(1)}$ the feasible GLS estimator based on weighted GM estimators, iterated once; and $\breve{\beta}_1^{(1)}$ the feasible GLS estimator based on partially weighted GM estimators, iterated once.
$N = 100$, $T = 5$, and $(\sigma_\mu^2, \sigma_v^2, \sigma_1^2, \beta_1, \beta_2) = (1,1,6,1,1)$.

The similarity of the RMSEs of the weighted and partially weighted GM estimators suggests that, at least for the weighting matrices considered in our experiments, the covariance structure of the weighting variance covariance matrix in (26) is not "very important" in determining the efficiency of the corresponding GM estimator. The variance factors, however, are important. Results of experiments, which are not reported here, reinforce this conjecture. For example, we also explored the small sample properties of an unweighted GM estimator of ρ based on all six moments, and found that this estimator is much less efficient than any of the other estimators, including the initial unweighted GM estimator based only on the first three moments. Upon reflection the reason for this is evident. The variance factors in the variance covariance weighting matrix in (26) are $\sigma_v^4/(T-1)$ and σ_1^4, where $\sigma_1^2 = \sigma_v^2 + T\sigma_\mu^2$. For $T = 5$, the values of $\sigma_v^4/(T-1)$ and σ_1^4 corresponding to $(\sigma_\mu^2, \sigma_v^2) = (1, 1)$ are, respectively, 0.25 and 36, implying a ratio of $0.25/36 = 0.0069$. The unweighted GM estimator based on six equations gives equal weight to each of the six equations.

Table 5
RMSEs of feasible GLS estimators of β_2

Parameter values		RMSE						
J	ρ	$\widehat{\beta}_2^{\mathrm{ML}}$	$\tilde{\beta}_2$	$\hat{\beta}_2$	$\breve{\beta}_2$	$\tilde{\beta}_2^{(1)}$	$\hat{\beta}_2^{(1)}$	$\breve{\beta}_2^{(1)}$
2	-0.9	0.0112	0.0112	0.0112	0.0113	0.0112	0.0112	0.0111
2	-0.5	0.0129	0.0128	0.0128	0.0129	0.0128	0.0129	0.0129
2	-0.25	0.0137	0.0140	0.0137	0.0138	0.0139	0.0137	0.0138
2	0	0.0151	0.0151	0.0152	0.0152	0.0150	0.0152	0.0152
2	0.25	0.0162	0.0162	0.0162	0.0162	0.0162	0.0162	0.0161
2	0.5	0.0162	0.0163	0.0162	0.0161	0.0164	0.0164	0.0161
2	0.9	0.0154	0.0153	0.0153	0.0154	0.0154	0.0154	0.0154
6	-0.9	0.0122	0.0121	0.0122	0.0121	0.0122	0.0123	0.0122
6	-0.5	0.0133	0.0133	0.0133	0.0134	0.0134	0.0132	0.0132
6	-0.25	0.0142	0.0143	0.0143	0.0143	0.0143	0.0142	0.0142
6	0	0.0151	0.0151	0.0152	0.0153	0.0152	0.0152	0.0152
6	0.25	0.0156	0.0154	0.0156	0.0156	0.0154	0.0156	0.0156
6	0.5	0.0158	0.0161	0.0160	0.0159	0.0160	0.0160	0.0160
6	0.9	0.0163	0.0162	0.0163	0.0162	0.0162	0.0163	0.0162
10	-0.9	0.0121	0.0116	0.0120	0.0120	0.0120	0.0121	0.0121
10	-0.5	0.0132	0.0134	0.0133	0.0132	0.0133	0.0133	0.0134
10	-0.25	0.0142	0.0143	0.0142	0.0142	0.0143	0.0142	0.0141
10	0	0.0154	0.0152	0.0151	0.0153	0.0153	0.0153	0.0153
10	0.25	0.0161	0.0159	0.0159	0.0160	0.0160	0.0161	0.0161
10	0.5	0.0170	0.0168	0.0171	0.0170	0.0169	0.0171	0.0171
10	0.9	0.0171	0.0169	0.0171	0.0170	0.0169	0.0171	0.0170
Column averages		0.0147	0.0146	0.0147	0.0147	0.0147	0.0147	0.0147

$\widehat{\beta}_2^{\mathrm{ML}}$ is the maximum likelihood estimator; $\tilde{\beta}_2$ the feasible GLS estimator based on initial GM estimators; $\hat{\beta}_2$ the feasible GLS estimator based on weighted GM estimators; $\breve{\beta}_2$ the feasible GLS estimator based on partially weighted GM estimators; $\tilde{\beta}_2^{(1)}$ the feasible GLS estimator based on initial GM estimators, iterated once; $\hat{\beta}_2^{(1)}$ the feasible GLS estimator based on weighted GM estimators, iterated once; and $\breve{\beta}_2^{(1)}$ the feasible GLS estimator based on partially weighted GM estimator, iterated once.
Based on $N = 100$, $T = 5$, and $(\sigma_\mu^2, \sigma_v^2, \sigma_1^2, \beta_1, \beta_2) = (1, 1, 6, 1, 1)$.

As remarked, our GM estimators for ρ are not confined to the interval $(-1, 1)$. However, in our Monte Carlo study we only observed such outliers in less that 1%. As described above, our RMSE measure is based on quantiles, and hence it is not effected by those infrequent outliers.

In Tables 2 and 3 we report RMSEs for the GM estimators $\tilde{\sigma}_v^2$, $\hat{\sigma}_v^2$, $\breve{\sigma}_v^2$ and $\breve{\sigma}_1^2$, $\tilde{\sigma}_1^2$, $\hat{\sigma}_1^2$, the corresponding iterated GM estimators, and the ML estimators, $\hat{\sigma}_v^{2,\mathrm{ML}}$ and $\hat{\sigma}_1^{2,\mathrm{ML}}$, respectively. The RMSEs for the initial GM estimators are again larger than those for the weighted and partially weighted GM estimators. In contrast to the GM estimators for ρ, the iterated GM estimators for σ_v^2 and σ_1^2 exhibit smaller RMSEs than their noniterated counter parts. The RMSEs of the weighted GM estimators $\hat{\sigma}_v^2$ and $\tilde{\sigma}_1^2$ are on average 2% and 6% higher than the RMSEs of the ML estimators. The difference between the average RMSEs of the iterated weighted GM estimators and the RMSEs of the ML estimators is less than 1%. The difference between the average RMSEs of the iterated partially weighted GM estimators and the RMSEs of the ML estimators is less than 2%.

In Tables 4 and 5 we report RMSEs for the feasible GLS estimators $\tilde{\beta}_{i,\text{FGLS}}$, $\hat{\beta}_{i,\text{FGLS}}$ and $\breve{\beta}_{i,\text{FGLS}}$, $i = 1, 2$, respectively, and iterated versions thereof. Those tables reveal that the RMSEs of the various feasible GLS estimators and the ML estimators for the regression parameters β_1 and β_2, respectively, are very similar. These results and those given for the GM estimators of ρ, σ_v^2 and σ_1^2 are encouraging, given the computational simplicity of our suggested estimators relative to the ML estimator.

6. Future research

Several suggestions for future research come to mind. First, on a theoretical level, it would be of interest to also consider fixed effects specifications. Furthermore, it would be of interest to extend the results of this paper to models containing spatially lagged dependent variables, as well as to systems of equations. In doing this it would certainly be of interest to consider higher order spatial lags. Another area of interest would be to extend our results to nonlinear models. Secondly, a Monte Carlo study relating to a wider set of experiments, as well as corresponding estimators, than those considered in this paper would be of interest.

Acknowledgments

Appendix A

Remark A1. In the following we make repeated use of the properties of the matrices $Q_{0,N}$ and $Q_{1,N}$ discussed after (8). In particular those matrices are symmetric, idempotent and orthogonal, and

$$Q_{0,N} + Q_{1,N} = I_{NT},$$

$$tr(Q_{0,N}) = N(T-1), \quad tr(Q_{1,N}) = N,$$

$$Q_{0,N}(e_T \otimes I_N) = 0, \quad Q_{1,N}(e_T \otimes I_N) = (e_T \otimes I_N). \tag{A.1}$$

Furthermore, let R_N be any $N \times N$ matrix, then it is readily seen that

$$(I_T \otimes R_N)Q_{0,N} = Q_{0,N}(I_T \otimes R_N), \quad (I_T \otimes R_N)Q_{1,N} = Q_{1,N}(I_T \otimes R_N),$$

$$tr(Q_{0,N}(I_T \otimes R_N)) = (T-1)\,tr(R_N), \quad tr(Q_{1,N}(I_T \otimes R_N)) = tr(R_N), \tag{A.2}$$

observing that for any $T \times T$ matrix S_T we have $tr(S_T \otimes R_N) = tr(S_T)\,tr(R_T)$.

Remark A2. In the following we shall make repeated use of the following observations:[14]
(a) Let R_N be a (sequence of) $N \times N$ matrices whose row and column sums are bounded uniformly in absolute value, and let S be some $k \times k$ matrix (with $k \geqslant 1$ fixed). Then the row and column sums of $S \otimes R_N$ are bounded uniformly in absolute value.

---

[14]The observations are readily verified. For some explicit proofs see Kelejian and Prucha (1999).

(b) If A_N and B_N are (sequences of) $kN \times kN$ matrices (with $k \geqslant 1$ fixed), whose row and column sums are bounded uniformly in absolute value, then so are the row and column sums of $A_N B_N$ and $A_N + B_N$. If Z_N is a (sequence of) $kN \times p$ matrices whose elements are uniformly bounded in absolute value, then so are the elements of $A_N Z_N$ and $(kN)^{-1} Z'_N A_N Z_N$.

Derivation of the moment conditions in (13): Given the error component specification (5) for ε_N and utilizing the properties of $Q_{0,N}$ and $Q_{1,N}$ described in Remark A1 it follows that:

$$Q_{0,N}\varepsilon_N = Q_{0,N}v_N,$$

$$Q_{0,N}\bar{\varepsilon}_N = (I_T \otimes W_N)Q_{0,N}v_N,$$

$$Q_{1,N}\varepsilon_N = (e_T \otimes I_N)\mu_N + Q_{1,N}v_N,$$

$$Q_{1,N}\bar{\varepsilon}_N = (e_T \otimes W_N)\mu_N + (I_T \otimes W_N)Q_{1,N}v_N. \tag{A.3}$$

Recall that by Assumption 1, $\mathrm{E}\mu_N\mu'_N = \sigma_\mu^2 I_N$, $\mathrm{E}v_N v'_N = \sigma_v^2 I_{NT}$, and $\mathrm{E}\mu_N v'_N = 0$. Also recall that for any random vector η and conformable matrix A we have $\mathrm{E}(\eta'A\eta) = tr(A\mathrm{E}\eta\eta')$. Given this, Remark A1, and (A.3) we have

$$\begin{aligned} \mathrm{E}[\varepsilon'_N Q_{0,N}\varepsilon_N] &= \mathrm{E}[v'_N Q_{0,N}v_N] \\ &= \sigma_v^2 \, tr(Q_{0,N}) = \sigma_v^2 N(T-1), \end{aligned}$$

$$\begin{aligned} \mathrm{E}[\bar{\varepsilon}'_N Q_{0,N}\bar{\varepsilon}_N] &= \mathrm{E}[v'_N Q_{0,N}(I_T \otimes W'_N W_N)Q_{0,N}v_N] \\ &= \sigma_v^2 \, tr[Q_{0,N}(I_T \otimes W'_N W_N)] = \sigma_v^2 (T-1)\, tr(W'_N W_N), \end{aligned}$$

$$\begin{aligned} \mathrm{E}[\bar{\varepsilon}'_N Q_{0,N}\varepsilon_N] &= \mathrm{E}[v'_N Q_{0,N}(I_T \otimes W'_N)Q_{0,N}v_N] \\ &= \sigma_v^2 \, tr[Q_{0,N}(I_T \otimes W'_N)] = \sigma_v^2 (T-1)\, tr(W'_N) = 0, \end{aligned} \tag{A.4}$$

and

$$\begin{aligned} \mathrm{E}[\varepsilon'_N Q_{1,N}\varepsilon_N] &= \mathrm{E}[\mu'_N(e'_T e_T \otimes I_N)\mu_N] + \mathrm{E}[v'_N Q_{1,N}v_N] \\ &= \sigma_\mu^2 \, tr(e'_T e_T \otimes I_N) + \sigma_v^2 \, tr(Q_{1,N}) \\ &= NT\sigma_\mu^2 + N\sigma_v^2 = N\sigma_1^2, \end{aligned}$$

$$\begin{aligned} \mathrm{E}[\bar{\varepsilon}'_N Q_{1,N}\bar{\varepsilon}_N] &= \mathrm{E}[\mu'_N(e'_T e_T \otimes W'_N W_N)\mu_N] + \mathrm{E}[v'_N Q_{1,N}(I_T \otimes W'_N W_N)Q_{1,N}v_N] \\ &= \sigma_\mu^2 \, tr(e'_T e_T \otimes W'_N W_N) + \sigma_v^2 \, tr[Q_{1,N}(I_T \otimes W'_N W_N)] \\ &= T\sigma_\mu^2 \, tr(W'_N W_N) + \sigma_v^2 \, tr(W'_N W_N) = \sigma_1^2 \, tr(W'_N W_N), \end{aligned}$$

$$\begin{aligned} \mathrm{E}[\bar{\varepsilon}'_N Q_{1,N}\varepsilon_N] &= \mathrm{E}[\mu'_N(e'_T e_T \otimes W'_N)\mu_N] + \mathrm{E}[v'_N Q_{1,N}(I_T \otimes W'_N)Q_{1,N}v_N] \\ &= \sigma_\mu^2 \, tr(e'_T e_T \otimes W'_N) + \sigma_v^2 \, tr[Q_{1,N}(I_T \otimes W'_N)] \\ &= T\sigma_\mu^2 \, tr(W'_N) + \sigma_v^2 \, tr(W'_N) = 0, \end{aligned} \tag{A.5}$$

with $\sigma_1^2 = T\sigma_\mu^2 + \sigma_v^2$. The moment equations given in (13) now follow immediately from (A.4) and (A.5).

We now give a sequence of lemmata which are needed for the proof of Theorems 1–3.

Lemma A1. *Let S_T be some $T \times T$ matrix (with T fixed), and let R_N be some $N \times N$ matrix whose row and column sums are bounded uniformly in absolute value. Let $\varepsilon_N = (e_T \otimes I_N)\mu_N + v_N$ where μ_N and v_N satisfy Assumption 1. Consider the quadratic form*

$$\varphi_N = N^{-1}\varepsilon_N'(S_T \otimes R_N)\varepsilon_N.$$

Then $\mathrm{E}\varphi_N = \mathrm{O}(1)$ and $var(\varphi_N) = \mathrm{o}(1)$, and as a consequence

$$\varphi_N - \mathrm{E}\varphi_N \xrightarrow{P} 0 \quad as \ N \to \infty.$$

Proof. Let $\xi_N = (\xi_{1,N}, \ldots, \xi_{(T+1)N,N})' = (\mu_N', v_N')'$ so that $\varepsilon_N = [(e_T \otimes I_N), I_{NT}]\xi_N$ and $\varphi_N = N^{-1}\xi_N' C_N \xi_N$ with

$$C_N = \begin{bmatrix} e_T' S_T e_T & e_T' S_T \\ S_T e_T & S_T \end{bmatrix} \otimes R_N. \tag{A.6}$$

Since the first matrix of the Kronecker product in (A.6) does not depend on N it follows immediately from the maintained assumption concerning R_N and Remark A2 that the row and column sums (and hence the elements) of C_N are bounded uniformly in absolute value by, say $k_c < \infty$. Next observe that in light of Assumption 1 the $(T+1)N \times 1$ vector ξ_N has mean zero, variance covariance matrix

$$\Omega_\xi = \mathrm{E}\xi_N \xi_N' = \begin{bmatrix} \sigma_\mu^2 I_N & 0 \\ 0 & \sigma_v^2 I_{NT} \end{bmatrix},$$

and finite fourth moments. In the following let $1 < k_\eta < \infty$ denote an upper bound for the variances and fourth moments of the elements of μ_N and v_N, and thus of the elements of ξ_N. Then $|\mathrm{E}(\varphi_N)| = |N^{-1} tr(C_N \Omega_\xi)| \leqslant N^{-1}(\sum_{i=1}^{(T+1)N}|c_{ii,N}|var(\xi_{i,N}) \leqslant (T+1)k_\eta k_c < \infty$, and thus $\mathrm{E}\varphi_N = \mathrm{O}(1)$. Using the expression for the variance of quadratic forms given, e.g., in Kelejian and Prucha (2001), we have

$$var(\varphi_N) = \frac{1}{2N^2}\left\{ tr[(C_N + C_N')\Omega_\xi(C_N + C_N')\Omega_\xi] + \sum_{i=1}^{(T+1)N} c_{ii,N}^2[\mathrm{E}\xi_{i,N}^4 - 3 var^2(\xi_{i,N})]\right\}.$$

Given Remark A2, it is clear that the row and column sums of the absolute values of the matrix $(C_N + C_N')\Omega_\xi(C_N + C_N')\Omega_\xi$ are bounded uniformly by $4k_c^2 k_\eta^2$. Hence, using the triangle inequality, we have

$$var(\varphi_N) \leqslant \frac{1}{2N^2}\{(T+1)N4k_c^2 k_\eta^2 + (T+1)N4k_c^2 k_\eta^2\} \to 0 \quad as \ N \to \infty,$$

which shows that $var(\varphi_N) = \mathrm{o}(1)$. The last claim follows immediately from Chebyshev's inequality. $\square$

Lemma A2. *Let G_N^* and g_N^* be identical to Γ_N and γ_N in (15) except that the expectations operator is dropped. Suppose Assumptions 1, 2, and 4 hold. Then $\Gamma_N = \mathrm{O}(1)$, $\gamma_N = \mathrm{O}(1)$ and*

$$G_N^* - \Gamma_N \xrightarrow{P} 0 \quad and \quad g_N^* - \gamma_N \xrightarrow{P} 0 \ as \ N \to \infty. \tag{A.7}$$

Proof. Note from (4) that

$$u_N = [I_T \otimes (I_N - \rho W_N)^{-1}]\varepsilon_N = [I_T \otimes P_N]\varepsilon_N,$$
$$\bar{u}_N = (I_T \otimes W_N)u_N = (I_T \otimes W_N P_N)\varepsilon_N,$$
$$\bar{\bar{u}}_N = (I_T \otimes W_N)\bar{u}_N = (I_T \otimes W_N^2 P_N)\varepsilon_N. \tag{A.8}$$

Define

$$S_{0,T} = \frac{1}{T-1}\left(I_T - \frac{J_T}{T}\right),$$

$$S_{1,T} = \frac{J_T}{T}.$$

Then, recalling the definition of $Q_{0,N}$ and $Q_{1,N}$ in (8) it is not difficult to verify that the respective quadratic forms in u_N, $\bar{u}_N$, and $\bar{\bar{u}}_N$ involved in G_N^* and g_N^* are, apart from a constant, expressible as

$$\varphi_{1j,N} = \frac{1}{N(T-1)^{1-j}}u_N' Q_{j,N} u_N = \frac{1}{N}u_N'(S_{j,T} \otimes I_N)u_N$$
$$= \frac{1}{N}\varepsilon_N'(S_{j,T} \otimes R_{1,N})\varepsilon_N, \quad R_{1,N} = P_N' P_N,$$

$$\varphi_{2j,N} = \frac{1}{N(T-1)^{1-j}}u_N' Q_{j,N} \bar{u}_N = \frac{1}{N}u_N'(S_{j,T} \otimes W_N)u_N$$
$$= \frac{1}{N}\varepsilon_N'(S_{j,T} \otimes R_{2,N})\varepsilon_N, \quad R_{2,N} = P_N' W_N P_N,$$

$$\varphi_{3j,N} = \frac{1}{N(T-1)^{1-j}}\bar{u}_N' Q_{j,N} \bar{u}_N = \frac{1}{N}u_N'(S_{j,T} \otimes W_N' W_N)u_N$$
$$= \frac{1}{N}\varepsilon_N'(S_{j,T} \otimes R_{3,N})\varepsilon_N, \quad R_{3,N} = P_N' W_N' W_N P_N,$$

$$\varphi_{4j,N} = \frac{1}{N(T-1)^{1-j}}\bar{\bar{u}}_N' Q_{j,N} \bar{u}_N = \frac{1}{N}u_N'(S_{j,T} \otimes (W_N')^2 W_N)u_N$$
$$= \frac{1}{N}\varepsilon_N'(S_{j,T} \otimes R_{4,N})\varepsilon_N, \quad R_{4,N} = P_N' (W_N')^2 W_N P_N,$$

$$\varphi_{5j,N} = \frac{1}{N(T-1)^{1-j}}\bar{\bar{u}}_N' Q_{j,N} \bar{\bar{u}}_N = \frac{1}{N}u_N'(S_{j,T} \otimes (W_N')^2 W_N^2)u_N$$
$$= \frac{1}{N}\varepsilon_N'(S_{j,T} \otimes R_{5,N})\varepsilon_N, \quad R_{5,N} = P_N' (W_N')^2 W_N^2 P_N,$$

$$\varphi_{6j,N} = \frac{1}{N(T-1)^{1-j}}u_N' Q_{j,N} \bar{\bar{u}}_N = \frac{1}{N}u_N'(S_{j,T} \otimes W_N^2)u_N$$
$$= \frac{1}{N}\varepsilon_N'(S_{j,T} \otimes R_{6,N})\varepsilon_N, \quad R_{6,N} = P_N' W_N^2 P_N, \tag{A.9}$$

with $j = 0, 1$. In light of Assumptions 2 and 4 the row and column sums of W_N and P_N, and hence those of the matrices $R_{i,N}$ $(i = 1, \ldots, 6)$, are bounded uniformly in absolute

value; compare Remark A2. The lemma now follows by applying Lemma A1 to each of the quadratic forms in (A.9) which compose G_N^* and g_N^*. $\square$

Lemma A3. *Let* G_N^* *and* g_N^* *be as defined in Lemma* A2. *Then, given Assumptions* 1–4

$$G_N - G_N^* \xrightarrow{P} 0 \quad and \quad g_N - g_N^* \xrightarrow{P} 0 \ as \ N \to \infty \tag{A.10}$$

provided $\tilde{\beta}_N \xrightarrow{P} \beta$ *as* $N \to \infty$.

Proof. The quadratic forms composing the elements of G_N^* and g_N^* have been collected in (A.9) and are seen to be of the form ($i = 1, \ldots, 6, j = 1, 2$):

$$\varphi_{ij,N} = N^{-1} u_N' C_{ij,N} u_N, \tag{A.11}$$

where the $C_{ij,N}$ are nonstochastic $NT \times NT$ matrices. Since the row and column sums of the elements of W_N are uniformly bounded in absolute value it follows that also the row and columns sums of the matrices $C_{ij,N}$ have that property; compare Remark A2. The quadratic forms composing the elements of G_N and g_N defined in (17) are given by

$$\tilde{\varphi}_{ij,N} = N^{-1} \tilde{u}_N' C_{ij,N} \tilde{u}_N. \tag{A.12}$$

To proof the lemma we now show that $\tilde{\varphi}_{ij,N} - \varphi_{ij,N} \xrightarrow{P} 0$ as $N \to \infty$. Clearly

$$\tilde{u}_N = y_N - X_N \tilde{\beta}_N = u_N - X_N \Delta_N, \tag{A.13}$$

where $\Delta_N = \tilde{\beta}_N - \beta$. Since $\tilde{\beta}_N$ is consistent, $\Delta_N \xrightarrow{P} 0$. Substituting (A.13) into (A.12) yields

$$\tilde{\varphi}_{ij,N} - \varphi_{ij,N} = \Delta_N'(N^{-1} X_N' C_{ij,N} X_N)\Delta_N - 2\Delta_N'(N^{-1} X_N' C_{ij,N} u_N). \tag{A.14}$$

Consider the first term on the right-hand side of (A.14). Since the row and column sums of $C_{ij,N}$ are bounded uniformly in absolute value and since the elements of X_N are uniformly bounded in absolute value it follows—see Remark A2—that all K^2 elements of $N^{-1} X_N' C_{ij,N} X_N$ are O(1). Therefore, the first term on the right-hand side of (A.14) converges to zero in probability since $\Delta_N \xrightarrow{P} 0$.

Now consider the second term on the right-hand side of (A.14). In particular, consider the vector $\zeta_N = N^{-1} X_N' C_{ij,N} u_N$. Clearly, the mean of ζ_N is zero and its variance covariance matrix is given by

$$N^{-1}(N^{-1} X_N' C_{ij,N}' \Omega_{u,N} C_{ij,N} X_N), \tag{A.15}$$

where $\Omega_{u,N}$ is given by (7) and (11). Given the maintained assumptions the row and column sums of $\Omega_{u,N}$ are uniformly bounded in absolute value, and therefore so are those of $C_{ij,N}' \Omega_{u,N} C_{ij,N}$. Observing that the elements of X_N are uniformly bounded in absolute value, it follows again from the observations in Remark A2 that all K^2 elements of $N^{-1} X_N' C_{ij,N}' \Omega_{u,N} C_{ij,N} X_N$ are O(1). This implies that the variance covariance matrix of ζ_N given in (A.15) converges to zero, and hence ζ_N converges to zero in probability. This establishes that also the second term on the right-hand side of (A.14) converges to zero in probability. Thus $\tilde{\varphi}_{ij,N} - \varphi_{ij,N} \xrightarrow{P} 0$ as $N \to \infty$, which completes the proof. $\square$

Proof of Theorem 1. Observe that by combining Lemmata A2 and A3 we have

$$G_N - \Gamma_N \xrightarrow{P} 0 \quad and \quad g_N - \gamma_N \xrightarrow{P} 0 \ as \ N \to \infty. \tag{A.16}$$

We are now ready for the final step in the proof of Theorem 1. We first demonstrate the consistency of $\tilde{\rho}_N$ and $\tilde{\sigma}_{v,N}^2$ defined in (24).[15] The existence and measurability of $\tilde{\rho}_N$ and $\tilde{\sigma}_{v,N}^2$ are ensured by, for example, Lemma 2 in Jennrich (1969). To establish the consistency of $\tilde{\rho}_N$ and $\tilde{\sigma}_{v,N}^2$ we show that the conditions of Lemma 3.1 in Pötscher and Prucha (1997) are satisfied for the problem at hand. Let $\theta = (\rho, \sigma_v^2)$ and $\underline{\theta} = (\underline{\rho}, \underline{\sigma}_v^2)$. We first show that the true parameter vector θ is identifiably unique. The objective function of the nonlinear least squares estimator and its corresponding nonstochastic counterpart are given by, respectively,

$$R_N^0(\underline{\theta}) = [G_N^0[\underline{\rho}, \underline{\rho}^2, \underline{\sigma}_v^2]' - g_N^0]'[G_N^0[\underline{\rho}, \underline{\rho}^2, \underline{\sigma}_v^2]' - g_N^0],$$

$$\overline{R}_N^0(\underline{\theta}) = [\Gamma_N^0[\underline{\rho}, \underline{\rho}^2, \underline{\sigma}_v^2]' - \gamma_N^0]'[\Gamma_N^0[\underline{\rho}, \underline{\rho}^2, \underline{\sigma}_v^2]' - \gamma_N^0].$$

Observe that in light of (15) or (18) we have $\overline{R}_N^0(\theta) = 0$, i.e., $\overline{R}_N^0(\underline{\theta})$ is zero at $\underline{\theta} = \theta$. Then

$$
\begin{aligned}
\overline{R}_N^0(\underline{\theta}) - \overline{R}_N^0(\theta) &= [\underline{\rho} - \rho, \underline{\rho}^2 - \rho^2, \underline{\sigma}_v^2 - \sigma_v^2] \Gamma_N^{0\prime} \Gamma_N^0 [\underline{\rho} - \rho, \underline{\rho}^2 - \rho^2, \underline{\sigma}_v^2 - \sigma_v^2]' \\
&\geqslant \lambda_{\min}(\Gamma_N^{0\prime}\Gamma_N^0)[\underline{\rho} - \rho, \underline{\rho}^2 - \rho^2, \underline{\sigma}_v^2 - \sigma_v^2][\underline{\rho} - \rho, \underline{\rho}^2 - \rho^2, \underline{\sigma}_v^2 - \sigma_v^2]' \\
&\geqslant \lambda_* \| \underline{\theta} - \theta \|^2
\end{aligned}
$$

in light of Assumption 5. Hence, for every $\varepsilon > 0$ and any N,

$$\inf_{\{\underline{\theta}: \| \underline{\theta} - \theta \| \geqslant \varepsilon\}} [\overline{R}_N^0(\underline{\theta}) - \overline{R}_N^0(\theta)] \geqslant \inf_{\{\underline{\theta}: \| \underline{\theta} - \theta \| \geqslant \varepsilon\}} \lambda_* \| \underline{\theta} - \theta \|^2 = \lambda_* \varepsilon^2 > 0,$$

which proves that θ is identifiably unique. Next, let $F_N^0 = [G_N^0, -g_N^0]$ and $\Phi_N^0 = [\Gamma_N^0, -\gamma_N^0]$, then

$$R_N^0(\underline{\theta}) = [\underline{\rho}, \underline{\rho}^2, \underline{\sigma}_v^2, 1] F_N^{0\prime} F_N^0 [\underline{\rho}, \underline{\rho}^2, \underline{\sigma}_v^2, 1],$$

$$\overline{R}_N^0(\underline{\theta}) = [\underline{\rho}, \underline{\rho}^2, \underline{\sigma}_v^2, 1]' \Phi_N^{0\prime} \Phi_N^0 [\underline{\rho}, \underline{\rho}^2, \underline{\sigma}_v^2, 1].$$

Hence for $\rho \in [-a, a]$ and $\sigma_v^2 \in [0, b]$

$$
\begin{aligned}
|R_N^0(\underline{\theta}) - \overline{R}_N^0(\underline{\theta})| &= |[\underline{\rho}, \underline{\rho}^2, \underline{\sigma}_v^2, 1][F_N^{0\prime} F_N^0 - \Phi_N^{0\prime} \Phi_N^0][\underline{\rho}, \underline{\rho}^2, \underline{\sigma}_v^2, 1]| \\
&\leqslant \| F_N^{0\prime} F_N^0 - \Phi_N^{0\prime} \Phi_N^0 \| \| \underline{\rho}, \underline{\rho}^2, \underline{\sigma}_v^2, 1 \|^2 \\
&\leqslant \| F_N^{0\prime} F_N^0 - \Phi_N^{0\prime} \Phi_N^0 \| [1 + a^2 + a^4 + b^2].
\end{aligned}
$$

Given (A.16) we have $F_N^0 - \Phi_N^0 \xrightarrow{P} 0$. Observing that by Lemma A2 the elements of Φ_N^0 are $O(1)$ it follows that $\| F_N^{0\prime} F_N^0 - \Phi_N^{0\prime} \Phi_N^0 \| \xrightarrow{P} 0$, and consequently that $R_N^0(\underline{\theta}) - \overline{R}_N^0(\underline{\theta})$ converges to zero uniformly over the (extended) parameter space, that is,

$$
\sup_{\rho \in [-a, a], \sigma_v^2 \in [0, b]} |R_N^0(\underline{\theta}) - \overline{R}_N^0(\underline{\theta})|
$$

$$
\leqslant \| F_N^{0\prime} F_N^0 - \Phi_N^{0\prime} \Phi_N^0 \| [1 + a^2 + a^4 + b^2] \xrightarrow{P} 0
$$

---

[15]This step is analogous to that taken by Kelejian and Prucha (1999) in their consistency proof. We adopt the following notation: Let A be some vector or matrix, then $\|A\| = [tr(A'A)]^{1/2}$. We note that this norm is submultiplicative, i.e., $\|AB\| \leqslant \|A\| \|B\|$.

as $N \to \infty$. The consistency of $\tilde{\rho}_N$ and $\tilde{\sigma}^2_{v,N}$ now follows directly from Lemma 3.1 in Pötscher and Prucha (1997).

Next consider $\tilde{\sigma}^2_{1,N}$ defined by (25). In light of (15) we have

$$
\tilde{\sigma}^2_{1,N} - \sigma^2_1 = g^1_{1,N} - \gamma^1_{1,N} - (g^1_{11,N} - \gamma^1_{11,N})\tilde{\rho}_N - (g^1_{12,N} - \gamma^1_{12,N})\tilde{\rho}^2_N \\
- \gamma^1_{11,N}(\tilde{\rho}_N - \rho) - \gamma^1_{12,N}(\tilde{\rho}^2_N - \rho^2).
$$

Observing again that $F^0_N - \Phi^0_N \xrightarrow{P} 0$ and that the elements of Φ^0_N are O(1) it follows from the just established consistency of $\tilde{\rho}_N$ that $\tilde{\sigma}^2_{1,N} - \sigma^2_1 \xrightarrow{P} 0$ as $N \to \infty$. $\square$

Proof of Consistency of OLS. The least squares estimator of β based on (3) is $(X'_N X_N)^{-1} X'_N y_N$. Under the maintained Assumptions 1–4 the estimator is clearly unbiased and its variance covariance matrix is given by

$$
(NT)^{-1}[(NT)^{-1}X'_N X_N]^{-1}(NT)^{-1}X'_N \Omega_{u,N} X_N[(NT)^{-1}X'_N X_N]^{-1}. \tag{A.17}
$$

By Assumption 3, $(NT)^{-1}X'_N X_N$ converges to a finite positive definite matrix. As a special case of the discussion surrounding (A.15) it is seen that the K^2 elements of $(NT)^{-1}X'_N \Omega_{u,N} X_N$ are O(1). Therefore, the variance covariance matrix in (A.17) converges to zero as $N \to \infty$, and the consistency of the least squares estimator follows from Chebyshev's inequality.

Derivation of Variance Covariance Matrix in (26): Assume for the moment that $\varepsilon_N \sim N(0, \Omega_{\varepsilon,N})$. The derivation of Ξ_N in (26) is straight forward by using the following result (compare, e.g., Amemiya, 1973, Lemma): If A_N and B_N are nonnegative definite symmetric matrices, then

$$
cov(\varepsilon'_N A_N \varepsilon_N, \varepsilon'_N B_N \varepsilon_N) = 2\, tr(A_N \Omega_{\varepsilon,N} B_N \Omega_{\varepsilon,N}).
$$

As an illustration we compute the (2,1)-element of Ξ_N. Recall that $\bar{\varepsilon}_N = (I_T \otimes W_N)\varepsilon_N$ and $\Omega_{\varepsilon,N} = \sigma^2_v Q_{0,N} + \sigma^2_1 Q_{1,N}$. In light of Remark A1 it is not difficult to see that this element can be computed as follows:

$$
N\, cov\left[\frac{1}{N(T-1)}\bar{\varepsilon}'_N Q_{0,N} \bar{\varepsilon}_N, \frac{1}{N(T-1)}\varepsilon'_N Q_{0,N}\varepsilon_N\right]
$$

$$
= \frac{2}{N(T-1)^2} tr[(I_T \otimes W'_N)Q_{0,N}(I_T \otimes W_N)\Omega_{\varepsilon,N} Q_0 \Omega_{\varepsilon,N}]
$$

$$
= \frac{2}{N(T-1)^2} tr[(I_T \otimes W'_N W_N)Q_{0,N}\Omega_{\varepsilon,N} Q_0 \Omega_{\varepsilon,N}]
$$

$$
= \frac{2\sigma^4_v}{N(T-1)^2} tr[(I_T \otimes W'_N W_N)Q_{0,N}] = \frac{2\sigma^4_v}{(T-1)} tr\left[\frac{W'_N W_N}{N}\right].
$$

The derivation of the remaining elements of Ξ_N is analogous. $\square$

Proof of Theorem 2. The existence and measurability of the weighted GM estimators $\hat{\rho}_N$, $\hat{\sigma}^2_{v,N}$, and $\hat{\sigma}^2_{1,N}$ defined by (27) are again ensured by Lemma 2 in Jennrich (1969). Let $\theta = [\rho, \sigma^2_v, \sigma^2_1]$ and let $\underline{\theta} = [\underline{\rho}, \underline{\sigma}^2_v, \underline{\sigma}^2_1]$. The objective function of the weighted GM estimator and its corresponding nonstochastic counterpart are then given by, respectively,

$$
R_N(\underline{\theta}) = [G_N[\underline{\rho}, \underline{\rho}^2, \underline{\sigma}^2_v, \underline{\sigma}^2_1]' - g_N]'\tilde{\Xi}^{-1}_N[G_N[\underline{\rho}, \underline{\rho}^2, \underline{\sigma}^2_v, \underline{\sigma}^2_1]' - g_N],
$$

$$
\overline{R}_N(\underline{\theta}) = [\Gamma_N[\underline{\rho}, \underline{\rho}^2, \underline{\sigma}^2_v, \underline{\sigma}^2_1]' - \gamma_N]'\Xi^{-1}_N[\Gamma_N[\underline{\rho}, \underline{\rho}^2, \underline{\sigma}^2_v, \underline{\sigma}^2_1]' - \gamma_N].
$$

We now establish two preliminary results implied by the assumptions. First we show that $\lambda_{\min}(\Gamma_N' \Xi_N^{-1} \Gamma_N) \geqslant \lambda_\circ$ for some $\lambda_\circ > 0$. To see this let $A = (a_{ij}) = (\Gamma_N^0)'(\Gamma_N^0)$ and $B = (b_{ij}) = (\Gamma_N^1)'(\Gamma_N^1)$. Then in light of (15)

$$\Gamma_N' \Gamma_N = \begin{bmatrix} a_{11} & a_{12} & a_{13} & 0 \\ a_{21} & a_{22} & a_{23} & 0 \\ a_{31} & a_{32} & a_{33} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & 0 & b_{13} \\ b_{21} & b_{22} & 0 & b_{23} \\ 0 & 0 & 0 & 0 \\ b_{31} & b_{32} & 0 & b_{33} \end{bmatrix}.$$

Hence, utilizing Assumption 5,

$$
\begin{aligned}
x' \Gamma_N' \Gamma_N x &= [x_1, x_2, x_3] A [x_1, x_2, x_3]' + [x_1, x_2, x_4] B [x_1, x_2, x_4]' \\
&\geqslant \lambda_{\min}(A)[x_1, x_2, x_3][x_1, x_2, x_3]' + \lambda_{\min}(B)[x_1, x_2, x_4][x_1, x_2, x_4]' \\
&\geqslant \lambda_* x' x
\end{aligned}
$$

for any $x = [x_1, x_2, x_3, x_4]'$. Thus in light of, e.g., Rao (1973, p. 62)

$$\lambda_{\min}(\Gamma_N' \Gamma_N) = \inf_x \frac{x' \Gamma_N' \Gamma_N x}{x' x} \geqslant \lambda_* > 0.$$

Next observe that

$$
\begin{aligned}
\lambda_{\min}(\Gamma_N' \Xi_N^{-1} \Gamma_N) &= \inf_x \frac{x' \Gamma_N' \Xi_N^{-1} \Gamma_N x}{x' x} \geqslant \lambda_{\min}(\Xi_N^{-1}) \inf_x \frac{x' \Gamma_N' \Gamma_N x}{x' x} \\
&= \lambda_{\min}(\Xi_N^{-1}) \lambda_{\min}(\Gamma_N' \Gamma_N) \geqslant \lambda_\circ > 0,
\end{aligned}
$$

with $\lambda_\circ = \bar{\lambda}_* \lambda_*$, since $\lambda_{\min}(\Xi_N^{-1}) \geqslant \bar{\lambda}_* > 0$ by assumption.

Second, we show that $\Xi_N^{-1} = \mathrm{O}(1)$. For notational convenience let $S_N = \Xi_N^{-1}$. To verify the claim we need to show that $|s_{ij,N}| \leqslant k < \infty$ for some constant k that does not depend on N. Again in light of, e.g., Rao (1973, p. 62),

$$\lambda_{\min}(S_N) = \inf_x \frac{x' S_N x}{x' x}, \quad \lambda_{\max}(S_N) = \max_x \frac{x' S_N x}{x' x}.$$

Hence it follows from the maintained assumptions concerning the smallest and largest eigenvalues of $S_N = \Xi_N^{-1}$ that

$$0 < \bar{\lambda}_* \leqslant \frac{x' S_N x}{x' x} \leqslant \bar{\lambda}_{**} < \infty.$$

Taking x to be the vector that has a one in the ith and jth positions and zeros elsewhere we have

$$
\begin{aligned}
0 &\leqslant s_{ii,N} \leqslant \bar{\lambda}_{**} < \infty, \quad i = j, \\
0 &\leqslant (s_{ii,N} + s_{jj,N} + 2 s_{ij,N})/2 \leqslant \bar{\lambda}_{**} < \infty, \quad i \neq j.
\end{aligned}
$$

From this it is readily seen that $|s_{ij,N}| \leqslant \bar{\lambda}_{**}$ for all i, j, which proves the claim.

Analogous as in the proof of Theorem 1 observe that $\overline{R}_N(\underline{\theta})$ is zero at $\underline{\theta} = \theta$, and hence

$$
\begin{aligned}
\overline{R}_N(\underline{\theta}) &- \overline{R}_N(\theta) \\
&= [\underline{\rho} - \rho, \underline{\rho}^2 - \rho^2, \underline{\sigma}_v^2 - \sigma_v^2, \underline{\sigma}_1^2 - \sigma_1^2] \Gamma_N' \Xi_N^{-1} \Gamma_N [\underline{\rho} - \rho, \underline{\rho}^2 - \rho^2, \underline{\sigma}_v^2 - \sigma_v^2, \underline{\sigma}_1^2 - \sigma_1^2]'.
\end{aligned}
$$

Next, let $F_N = [G_N, -g_N]$ and $\Phi_N = [\Gamma_N, -\gamma_N]$, then

$$R_N(\underline{\theta}) = [\underline{\rho}, \underline{\rho}^2, \underline{\sigma}_v^2, \underline{\sigma}_1^2, 1]F_N' \tilde{\Xi}_N^{-1} F_N [\underline{\rho}, \underline{\rho}^2, \underline{\sigma}_v^2, \underline{\sigma}_1^2, 1],$$

$$\overline{R}_N(\underline{\theta}) = [\underline{\rho}, \underline{\rho}^2, \underline{\sigma}_v^2, \underline{\sigma}_1^2, 1]' \Phi_N' \Xi_N^{-1} \Phi_N [\underline{\rho}, \underline{\rho}^2, \underline{\sigma}_v^2, \underline{\sigma}_1^2, 1].$$

The remainder of the proof of Theorem 2 is now analogous to that of Theorem 1. $\quad\square$

Proof of Theorem 3. Let $\bar{\lambda}_* = \min\{(T-1)\sigma_v^{-4}, \sigma_1^{-4}\}$ and $\bar{\lambda}_{**} = \max\{(T-1)\sigma_v^{-4}, \sigma_1^{-4}\}$. Then $0 < \bar{\lambda}_* \leqslant \lambda_{\min}(\Upsilon_N^{-1}) \leqslant \lambda_{\max}(\Upsilon_N^{-1}) \leqslant \bar{\lambda}_{**} < \infty$. The proof of Theorem 3 is now analogous to that of Theorem 2 with Ξ_N and $\tilde{\Xi}_N$ replaced by Υ_N and $\tilde{\Upsilon}_N$, respectively. $\quad\square$

Proof of Theorem 4. *Part* (a): We prove this part by verifying the conditions of the central limit theorem given in Pötscher and Prucha (2001) as Theorem 30.[16] First observe that (30)–(32) imply

$$
\begin{aligned}
(NT)^{1/2}&[\hat{\beta}_{\text{GLS},N} - \beta] \\
&= [(NT)^{-1} X_N' \Omega_{u,N}^{-1} X_N]^{-1} (NT)^{-1/2} X_N' \Omega_{u,N}^{-1} u_N \\
&= [(NT)^{-1} X_N^*(\rho)' \Omega_{\varepsilon,N}^{-1} X_N^*(\rho)]^{-1} (NT)^{-1/2} X_N^*(\rho)' \Omega_{\varepsilon,N}^{-1} \varepsilon_N.
\end{aligned} \tag{A.18}
$$

Consider the first factor in the last expression in (A.18). Recalling that $\Omega_{\varepsilon,N}^{-1} = \sigma_v^{-2} Q_{0,N} + \sigma_1^{-2} Q_{1,N}$ it follows from Assumption 3 and the discussion thereafter that

$$\lim_{N\to\infty} (NT)^{-1} X_N^*(\rho)' \Omega_{\varepsilon,N}^{-1} X_N^*(\rho) = \sigma_v^{-2} M_{xx}^0 + \sigma_1^{-2} M_{xx}^1, \tag{A.19}$$

which is a finite nonsingular matrix. Now consider the second factor in the last expression in (A.18). The results in (A.3) imply

$$
\begin{aligned}
(NT)^{-1/2} &X_N^*(\rho)' \Omega_{\varepsilon,N}^{-1} \varepsilon_N \\
&= (NT)^{-1/2} X_N^*(\rho)' [\sigma_v^{-2} Q_{0,N} + \sigma_1^{-2} Q_{1,N}] \varepsilon_N \\
&= (NT)^{-1/2} Z_{1,N}' \mu_N + (NT)^{-1/2} Z_{2,N}' v_N,
\end{aligned} \tag{A.20}
$$

where

$$Z_{1,N}' = X_N^*(\rho)'(\sigma_1^{-2}(e_T \otimes I_N)) \quad \text{and} \quad Z_{2,N}' = X_N^*(\rho)'(\sigma_v^{-2} Q_{0,N} + \sigma_1^{-2} Q_{1,N}).$$

Given Assumptions 3 and 4 it follows from the observations in Remark A2 that the elements of $Z_{1,N}$ and $Z_{2,N}$ are uniformly bounded in absolute value, and furthermore

$$
\begin{aligned}
\lim_{N\to\infty} (NT)^{-1} Z_{1,N}' Z_{1,N} &= \lim_{N\to\infty} (NT)^{-1} T \sigma_1^{-4} X_N^*(\rho)' Q_{1,N} X_N^*(\rho) \\
&= T \sigma_1^{-4} M_{xx}^1, \\
\lim_{N\to\infty} (NT)^{-1} Z_{2,N}' Z_{2,N} &= \lim_{N\to\infty} (NT)^{-1} [X_N^*(\rho)'(\sigma_v^{-4} Q_{0,N} + \sigma_1^{-4} Q_{1,N}) X_N^*(\rho)] \\
&= \sigma_v^{-4} M_{xx}^0 + \sigma_1^{-4} M_{xx}^1.
\end{aligned} \tag{A.21}
$$

---

[16]An inspection of the proof of that theorem shows that the theorem also holds if the innovations are allowed to form a triangular array and the sample index n is replaced by r_n.

The matrices on the right-hand side of (A.21) are finite. Given Assumption 1 it now follows from Theorem 30 in Pötscher and Prucha (2001) that

$$(NT)^{-1/2}Z'_{1,N}\mu_N \overset{D}{\to} N\{0, T\sigma^2_\mu\sigma^{-4}_1 M^1_{xx}\},$$
$$(NT)^{-1/2}Z'_{2,N}v_N \overset{D}{\to} N\{0, \sigma^2_v[\sigma^{-4}_v M^0_{xx} + \sigma^{-4}_1 M^1_{xx}]\}. \tag{A.22}$$

Since $\sigma^2_1 = \sigma^2_v + T\sigma^2_\mu$ and, by Assumption 1, the two processes $\{\mu_{i,N}\}$ and $\{v_{it,N}\}$ are independent, it follows furthermore from (A.20) and (A.22) that

$$(NT)^{-1/2}X^*_N(\rho)'\Omega^{-1}_{\varepsilon,N}\varepsilon_N \overset{D}{\to} N(0, \sigma^{-2}_v M^0_{xx} + \sigma^{-2}_1 M^1_{xx}) \tag{A.23}$$

as $N \to \infty$. Part (a) of Theorem 4 follows from the results in (A.18), (A.19) and (A.23).

  Part (b): To prove this part it suffices to show that[17]

$$\Delta_{1,N} = (NT)^{-1}X^*_N(\check{\rho}_N)'\check{\Omega}^{-1}_{\varepsilon,N}X^*_N(\check{\rho}_N) - (NT)^{-1}X^*_N(\rho)'\Omega^{-1}_{\varepsilon,N}X^*_N(\rho) \overset{P}{\to} 0, \tag{A.24}$$

and

$$\Delta_{2,N} = (NT)^{-1/2}X^*_N(\check{\rho}_N)'\check{\Omega}^{-1}_{\varepsilon,N}u^*_N(\check{\rho}_N) - (NT)^{-1/2}X^*_N(\rho)'\Omega^{-1}_{\varepsilon,N}u^*_N(\rho) \overset{P}{\to} 0, \tag{A.25}$$

where $\check{\Omega}^{-1}_{\varepsilon,N} = \Omega^{-1}_{\varepsilon,N}(\check{\sigma}^2_{v,N}, \check{\sigma}^2_{1,N})$ and $u^*_N(\check{\rho}_N) = (I_N - \check{\rho}_N W_N)u_N$.

  We first demonstrate (A.24). To this effect first substitute the expression for $X^*_N(\check{\rho}_N)$ given in (31) into (A.24). Then, observing that in light of (9) and Remark A1

$$\Omega^{-1}_{\varepsilon,N}[I_T \otimes (I_N - \rho W_N)] = [I_T \otimes (I_N - \rho W_N)]\Omega^{-1}_{\varepsilon,N},$$
$$\check{\Omega}^{-1}_{\varepsilon,N}[I_T \otimes (I_N - \check{\rho}_N W_N)] = [I_T \otimes (I_N - \check{\rho}_N W_N)]\check{\Omega}^{-1}_{\varepsilon,N}, \tag{A.26}$$

it is readily seen that

$$\begin{aligned}
\Delta_{1,N} &= [\check{\rho}_N - \rho]\check{\sigma}^{-2}_{v,N}(NT)^{-1}X'_N A_{1,N}X_N + [\check{\rho}_N - \rho]\check{\sigma}^{-2}_{1,N}(NT)^{-1}X'_N A_{2,N}X_N \\
&\quad + [\check{\rho}^2_N - \rho^2]\check{\sigma}^{-2}_{v,N}(NT)^{-1}X'_N A_{3,N}X_N + [\check{\rho}^2_N - \rho^2]\check{\sigma}^{-2}_{1,N}(NT)^{-1}X'_N A_{4,N}X_N \\
&\quad + [\check{\sigma}^{-2}_{v,N} - \sigma^{-2}_v](NT)^{-1}X'_N A_{5,N}X_N + [\check{\sigma}^{-2}_{1,N} - \sigma^{-2}_1](NT)^{-1}X'_N A_{6,N}X_N,
\end{aligned} \tag{A.27}$$

where

$$\begin{aligned}
A_{1,N} &= [I_T \otimes (W'_N + W_N)]Q_{0,N}, \\
A_{2,N} &= [I_T \otimes (W'_N + W_N)]Q_{1,N}, \\
A_{3,N} &= [I_T \otimes (W'_N W_N)]Q_{0,N}], \\
A_{4,N} &= [I_T \otimes (W'_N W_N)]Q_{1,N}, \\
A_{5,N} &= [I_T \otimes (I_N - \rho W'_N)(I_N - \rho W_N)]Q_{0,N}, \\
A_{6,N} &= [I_T \otimes (I_N - \rho W'_N)(I_N - \rho W_N)]Q_{1,N}.
\end{aligned} \tag{A.28}$$

In light of Remark A2 and Assumptions 3 and 4, the elements of the six $K \times K$ matrices $(NT)^{-1}X'_N A_{i,N}X_N$, $i = 1, \ldots, 6$, are seen to be uniformly bounded in absolute value. Since $\check{\rho}_N$, $\check{\sigma}^2_{v,N}$, and $\check{\sigma}^2_{1,N}$ are consistent estimators it follows that $\Delta_{1,N} \overset{P}{\to} 0$ as $N \to \infty$.

---

[17]See, e.g., Schmidt (1976, p. 71).

We next demonstrate (A.25). Using similar manipulations we have

$$
\begin{aligned}
\Delta_{2,N} ={}& [\check{\rho}_N - \rho]\check{\sigma}_{v,N}^{-2}(NT)^{-1/2}X_N'A_{1,N}u_N + [\check{\rho}_N - \rho]\check{\sigma}_{1,N}^{-2}(NT)^{-1/2}X_N'A_{2,N}u_N \\
&+ [\check{\rho}_N^2 - \rho^2]\check{\sigma}_{v,N}^{-2}(NT)^{-1/2}X_N'A_{3,N}u_N + [\check{\rho}_N^2 - \rho^2]\check{\sigma}_{1,N}^{-2}(NT)^{-1/2}X_N'A_{4,N}u_N \\
&+ [\check{\sigma}_{v,N}^{-2} - \sigma_v^{-2}](NT)^{-1/2}X_N'A_{5,N}u_N + [\check{\sigma}_{1,N}^{-2} - \sigma_1^{-2}](NT)^{-1/2}X_N'A_{6,N}u_N. \quad \text{(A.29)}
\end{aligned}
$$

Since X_N and $A_{i,N}$ are nonstochastic matrices the expected values of the $K \times 1$ vectors $(NT)^{-1/2}X_N'A_{i,N}u_N$ are zero and their variance covariance matrices are given by $(i = 1, \ldots, 6)$

$$
(NT)^{-1}X_N'A_{i,N}\Omega_{u,N}A_{i,N}'X_N,
$$

where $\Omega_{u,N}$ is defined via (7) and (11). Given Assumptions 3 and 4 and Remarks A1 and A2 the row and column sums of $\Omega_{u,N}$, and furthermore those of $A_{i,N}\Omega_{u,N}A_{i,N}'$, are uniformly bounded in absolute value. Hence, the elements of the $K \times K$ matrices $(NT)^{-1}X_N' A_{i,N}\Omega_{u,N}A_{i,N}'X_N$ are uniformly bounded in absolute value. Therefore $(NT)^{-1/2}X_N'A_{i,N}u_N = O_P(1)$ and thus $\Delta_{2,N} \xrightarrow{P} 0$ as $N \to \infty$, since $\check{\rho}_N$, $\check{\sigma}_{v,N}^2$, and $\check{\sigma}_{1,N}^2$ are consistent estimators.

Part (c): This part follows immediately from (A.19) and (A.24). $\quad\square$

References

Amemiya, T., 1971. The estimation of the variances in a variance-component model. International Economic Review 12, 1–13.

Amemiya, T., 1985. Advanced Econometrics. Harvard University Press, Cambridge, MA.

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

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

Balestra, P., Nerlove, M., 1966. Pooling cross section and time series data in the estimation of a dynamic model: the demand for natural gas. Econometrica 34, 585–612.

Baltagi, B.H., 1980. On seemingly unrelated regressions with error components. Econometrica 48, 1547–1551.

Baltagi, B.H., 1981. Simultaneous equations with error components. Journal of Econometrics 17, 189–200.

Baltagi, B.H., 2001. Econometric Analysis of Panel Data. Wiley, New York.

Baltagi, B.H., Li, D., 2001a. Double length artificial regressions for testing spatial dependence. Econometric Reviews 20, 31–40.

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

Baltagi, B.H., Li, D., 2004. Prediction in the panel data model with spatial correlation. In: Anselin, L., Florax, R.J.G.M., Rey, S.J. (Eds.), Advances in Spatial Econometrics: Methodology, Tools and Applications. Springer, Berlin, pp. 283–295.

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

Bernat Jr., G., 1996. Does manufacturing matter? A spatial econometric view of Kaldor's laws. Journal of Regional Science 36, 463–477.

Besley, T., Case, A., 1995. Incumbent behavior: vote-seeking, tax-setting, and yardstick competition. American Economic Review 85, 25–45.

Bollinger, C., Ihlanfeldt, K., 1997. The impact of rapid rail transit on economic development: the case of Atlanta's Marta. Journal of Urban Economics 42, 179–204.

Buettner, T., 1999. The effect of unemployment, aggregate wages, and spatial contiguity on local wages: an investigation with German district level data. Papers in Regional Science 78, 47–67.

Case, A., 1991. Spatial patterns in household demand. Econometrica 59, 953–966.

Case, A., Hines Jr., J., Rosen, H., 1993. Budget spillovers and fiscal policy independence: evidence from the States. Journal of Public Economics 52, 285–307.

Cliff, A., Ord, J., 1973. Spatial Autocorrelation. Pion, London.

Cliff, A., Ord, J., 1981. Spatial Processes, Models and Applications. Pion, London.

Conley, T., 1999. GMM estimation with cross sectional dependence. Journal of Econometrics 92, 1–45.

Cressie, N.A.C., 1993. Statistics of Spatial Data. Wiley, New York.

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

Dowd, M.R., LeSage, J.P., 1997. Analysis of spatial contiguity influences on state price level formation. International Journal of Forecasting 13, 245–253.

Gallant, A.R., White, H., 1988. A Unified Theory of Estimation and Inference in Nonlinear Dynamic Models. Basil Blackwell, New York.

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

Horn, R., Johnson, C., 1985. Matrix Analysis. Cambridge University Press, Cambridge.

Hsiao, C., 2003. Analysis for Panel Data. Cambridge University Press, Cambridge.

Jennrich, R., 1969. Asymptotic properties of non-linear least squares estimators. The Annals of Mathematical Statistics 40, 633–643.

Kelejian, H.H., Prucha, I.R., 1997. Estimation of spatial regression models with autoregressive errors by two-stage least squares procedures: a serious problem. International Regional Science Review 20, 103–111.

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

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

Kelejian, H.H., Prucha, I.R., 2001. On the asymptotic distribution of the Moran I test statistic with applications. Journal of Econometrics 104, 219–257.

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

Kelejian, H.H., Robinson, D., 1997. Infrastructure productivity estimation and its underlying econometric specifications. Papers in Regional Science 76, 115–131.

Kelejian, H.H., Robinson, D., 2000. Returns to investment in navigation infrastructure: an equilibrium approach. Annals of Regional Science 34, 83–108.

Kelejian, H.H., Prucha, I.R., Yuzefovich, E., 2004, Instrumental variable estimation of a spatial autorgressive model with autoregressive disturbances: large and small sample results. In: LeSage, J., Pace, R.K. (Eds.), Spatial and Spatiotemporal Econometrics, Advances in Econometrics, Vol. 18. Elsevier, New York, pp. 163–198.

Lee, L.-F., 2001a. Generalized method of moments estimation of spatial autoregressive processes. Mimeo, Department of Economics, Ohio State University.

Lee, L.-F., 2001b. GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. Mimeo, Department of Economics, Ohio State University.

Lee, L.F., 2002. Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models. Econometric Theory 18, 252–277.

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

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

LeSage, J.P., 1997. Bayesian estimation of spatial autoregressive models. International Regional Science Review 20, 113–129.

LeSage, J.P., 1999. A spatial econometric analysis of China's economic growth. Journal of Geographic Information Sciences 5, 143–153.

LeSage, J.P., 2000. Bayesian estimation of limited dependent variable spatial autoregressive models. Geographic Analysis 32, 19–35.

Nerlove, M., 1971. A note on error component models. Econometrica 39, 383–396.

Pace, R., Barry, R., 1997. Sparse spatial autoregressions. Statistics and Probability Letters 33, 291–297.

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

Pinkse, J., Slade, M.E., Brett, C., 2002. Spatial price competition: a semiparametric approach. Econometrica 70, 1111–1153.

Pötscher, B.M., Prucha, I.R., 1997. Dynamic Nonlinear Econometric Models, Asymptotic Theory. Springer, New York.

Pötscher, B.M., Prucha, I.R., 2001. Basic elements of asymptotic theory. In: Baltagi, B.H. (Ed.), A Companion to Theoretical Econometrics. Blackwell, Oxford, pp. 201–229.

Prucha, I.R., 1984. On the asymptotic efficiency of feasible Aitken estimators for seemingly unrelated regression models with error components. Econometrica 52, 203–207.

Prucha, I.R., 1985. Maximum likelihood and instrumental variable estimation in simultaneous equation systems with error components. International Economic Review 26, 491–506.

Rao, C.R., 1973. Linear Statistical Inference and Its Applications. Wiley, New York.

Rey, S.J., Boarnet, M.G., 2004. A taxonomy of spatial econometric models for simultaneous equation systems. In: Anselin, L., Florax, R.J.G.M., Rey, S.J. (Eds.), Advances in Spatial Econometrics: Methodology Tools and Applications. Springer, Berlin, pp. 99–119.

Schmidt, P., 1976. Econometrics. Marcel Dekker, New York.

Shroder, M., 1995. Games the states don't play: welfare benefits and the theory of fiscal federalism. Review of Economics and Statistics 77, 183–191.

Vigil, R., 1998. Interactions among municipalities in the provision of police services: a spatial econometric approach, Ph.D. Thesis, University of Maryland.

Wallace, T.D., Hussain, A., 1969. The use of error component models in combining cross section with time series data. Econometrica 37, 55–72.