JOURNAL OF REGIONAL SCIENCE, VOL. 46, NO. 3, 2006, pp. 507–515

ESTIMATION PROBLEMS IN MODELS WITH SPATIAL WEIGHTING MATRICES WHICH HAVE BLOCKS OF EQUAL ELEMENTS*

Harry H. Kelejian

Department of Economics, University of Maryland, College Park, MD 20742.
E-mail: kelejian@econ.umd.edu

Ingmar R. Prucha

Department of Economics, University of Maryland, College Park, MD 20742.
E-mail: prucha@econ.umd.edu

Yevgeny Yuzefovich

Fannie Mae, 3900 Wisconsin Avenue, Mailstop 1H-GR/04B, Washington, DC.
E-mail: yevgeny_yuzefovich@fanniemae.com

ABSTRACT. Spatial models whose weighting matrices have blocks of equal elements might be considered if units are viewed as equally distant within certain neighborhoods, but unrelated between neighborhoods. We give exact small sample results for such models that contain a spatially lagged-dependent variable. We consider cases in which the data relate to one or more panels, for example, villages, schools, etc. Our results are consistent with large sample results given in Kelejian and Prucha (2002) but indicate a variety of issues they did not consider.

1. INTRODUCTION

Spatial models whose weighting matrices have blocks of equal elements might be considered if units can reasonably be viewed as equally distant within certain neighborhoods, but unrelated between neighborhoods.[1] Examples of this would be studies in which the data relate to schools, villages, etc.[2] In an

*We would like to thank Marlon Boarnet and three referees for helpful comments. Kelejian and Prucha gratefully acknowledge support from the National Science Foundation through grant SES-0001780.

Received: July 2004; Revised: July 2005; Accepted: August 2005.

[1]Classic references on spatial models are Cliff and Ord (1973, 1981), Anselin (1988), and Cressie (1993). Some recent applications of spatial modeling are Audretsch and Felmann (1996), Bell and Bockstael (2000), Pinkse, Slade, and Brett (2002), Yuzefovich (2003), and Kapoor (2003).

[2]Among others, such a weighting matrix was considered by Splitstoser (2000) in a study of spatial interdependence involving the ideology of legislators, by Case (1992) in a panel data study of the adoption of new technologies by farmers, and by Lee (2002) in a study of the properties of least squares estimators in linear spatial models.

Blackwell Publishing, Inc., 350 Main Street, Malden, MA 02148, USA and 9600 Garsington Road, Oxford, OX4 2DQ, UK.

earlier study Kelejian and Prucha (2002) considered such models and demonstrated that if the model contains a spatially lagged-dependent variable, both the ordinary least squares (OLS) as well as two-stage least squares (2SLS) estimators are inconsistent if only one panel is available. They also demonstrate that if two or more, but a finite number, of panels are available both the OLS and the 2SLS estimators are not only consistent but both are also efficient within the class of instrumental variable (IV) estimators. These results were given for the case where no fixed effects are present.

In this paper we consider the model of Kelejian and Prucha (2002) but give exact small sample results for the OLS and the 2SLS estimators relating to both the single as well as multiple panel cases. Our results are consistent with the large sample results in Kelejian and Prucha (2002). However, we demonstrate that if fixed effects are considered in the multiple panel framework of Kelejian and Prucha (2002), as they very well might be, both the OLS and the 2SLS estimators are inconsistent. Our results suggest that for cases in which these estimators are inconsistent, typical tests of significance will be based on test statistics that are undefined in that they require division by zero. The implication is that results obtained in practice will, most likely, be determined entirely by rounding errors. We also show that if the model contains a spatially lagged-dependent variable and if the weighting matrix is not known and, therefore, is parameterized and its parameters are estimated along with the regression parameters by an IV procedure, the results will be inconsistent for a wide class of parameterizations of the weighting matrix. As somewhat of a corollary, we also indicate certain biases that result when the specification of a weighting matrix is selected on the basis of a measure of fit.

We specify the single panel data model in Section 2, and give our main results relating to that model in Section 3. Panel data extensions of that model are given in Section 4, along with corresponding results. Conclusions are given in Section 5.

2. MODEL SPECIFICATION

Consider the following Cliff–Ord type spatial model:

$$\begin{aligned} \mathbf{y}_N &= \mathbf{e}_N\alpha + \mathbf{X}_N\beta + \lambda\mathbf{W}_N\mathbf{y}_N + \varepsilon_N \\ &= \mathbf{Z}_N\gamma + \varepsilon_N \qquad (1) \\ \mathbf{Z}_N &= (\mathbf{e}_N, \mathbf{X}_N, \mathbf{W}_N\mathbf{y}_N), \quad \gamma' = (\alpha, \beta', \lambda) \end{aligned}$$

where $\mathbf{y}_N$ is the $N \times 1$ vector of observations on the dependent variable, $\mathbf{e}_N$ is an $N \times 1$ vector of unit elements, $\mathbf{W}_N$ is an $N \times N$ weighting matrix which is nonstochastic and observed, $\mathbf{X}_N$ is a full column rank $N \times k$ regressor matrix which is viewed as exogenous and which does not contain the intercept term, α is the intercept parameter, β is the parameter vector corresponding to $\mathbf{X}_N$, λ is the spatial autoregressive parameter corresponding to the spatial lag $\mathbf{W}_N\mathbf{y}_N$, and ε_N is the $N \times 1$ disturbance vector. Although evident from the specification, we note for future reference that the model under consideration contains both an intercept and a spatial lag of the dependent variable.

Suppose the researcher assumes, as would often be the case, that $E(\varepsilon_N) = \mathbf{0}$ and $E(\varepsilon_N \varepsilon_N') = \sigma^2 I_N$. Then, given $I_N - \lambda \mathbf{W}_N$ is nonsingular, we have $\mathbf{y}_N = (\mathbf{I}_N - \lambda \mathbf{W}_N)^{-1}[\alpha \mathbf{e}_N + \mathbf{X}_N \beta + \varepsilon_N]$, and so $\mathbf{W}_N \mathbf{y}_N = \mathbf{W}_N (\mathbf{I}_N - \lambda \mathbf{W}_N)^{-1} \times [\alpha \mathbf{e}_N + \mathbf{X}_N \beta + \varepsilon_N]$. Therefore

$$E(\mathbf{W}_N \mathbf{y}_N \varepsilon_N') = \sigma^2 \mathbf{W}_N (\mathbf{I}_N - \lambda \mathbf{W}_N)^{-1} \neq 0 \tag{2}$$

that is, in general the spatial lag $\mathbf{W}_N \mathbf{y}_N$ will be correlated with the disturbance vector ε_N. Given this endogeneity of $\mathbf{W}_N \mathbf{y}_N$ the researcher might attempt to estimate model (1) by the 2SLS procedure.[3]

3. BASIC RESULTS

Suppose the model in (1) is indeed estimated by 2SLS in terms of the full column rank $N \times (1 + k + r)$ matrix of instruments $\mathbf{H}_N = (\mathbf{e}_N, \mathbf{X}_N, \mathbf{G}_N)$ where, of course, $\mathbf{G}_N$ is an $N \times r$ matrix and $r \geq 1$. Given results in Kelejian and Prucha (1998), $\mathbf{G}_N$ could be taken to be the linearly independent columns of $(\mathbf{W}_N \mathbf{X}_N, \mathbf{W}_N^2 \mathbf{X}_N, \ldots, \mathbf{W}_N^q \mathbf{X}_N)$, where typically $q \leq 2$. Let $\mathbf{P}_{\mathbf{H}_N} = \mathbf{H}_N (\mathbf{H}_N' \mathbf{H}_N)^{-1} \mathbf{H}_N'$ and $\hat{\mathbf{Z}}_N = \mathbf{P}_{\mathbf{H}_N} \mathbf{Z}_N$. Then, assuming that $\hat{\mathbf{Z}}_N$ has full column rank, the 2SLS estimator of γ in (1) is

$$\hat{\gamma}_N = (\hat{\alpha}_N, \hat{\beta}_N', \hat{\lambda}_N)' = (\hat{\mathbf{Z}}_N' \hat{\mathbf{Z}}_N)^{-1} \hat{\mathbf{Z}}_N' \mathbf{y}_N \tag{3}$$

Our main result in this section is given in Theorem 1. Its implications are given in the remarks that follow.

THEOREM 1: *Assume the model in (1), and that $\hat{\mathbf{Z}}_N$ has full column rank so that $\hat{\gamma}_N$ can be calculated.*[4] Let $\bar{y}_N = \mathbf{e}_N' \mathbf{y}_N / N$ *denote the sample mean of* $\mathbf{y}_N$. *If*

$$\mathbf{W}_N = a_N[\mathbf{e}_N \mathbf{e}_N' - \mathbf{I}_N] = \begin{bmatrix} 0 & a_N & \ldots & a_N & a_N \\ a_N & 0 & \ldots & a_N & a_N \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_N & a_N & \ldots & 0 & a_N \\ a_N & a_N & \ldots & a_N & 0 \end{bmatrix} \tag{4}$$

where a_N is a constant whose value could depend upon the sample size, N, then

(a) $\hat{\gamma}_N = (\hat{\alpha}_N, \hat{\beta}_N', \hat{\lambda}_N)' = (N\bar{y}_N, \mathbf{0}, -1/a_N)$,

(b) $\hat{\varepsilon}_N = \mathbf{y}_N - \mathbf{Z}_N \hat{\gamma}_N = \mathbf{0}$.

[3]Concerning 2SLS estimation of spatial models see, for example, Das, Kelejian, and Prucha (2003), Kelejian and Prucha (1998), Lee (2003), and Rey and Boarnet (2004).

[4]Note, no further assumptions concerning ε_N are needed for the results of Theorem 1.

Proof of Theorem 1. First note that if $\mathbf{W}_N$ is given by (4), then

$$\mathbf{W}_N\mathbf{y}_N = (Na_N\bar{y}_N)\mathbf{e}_N - a_N\mathbf{y}_N \tag{5}$$

which is *linear in the variable being explained, namely* $\mathbf{y}_N$. Given (5), the estimated residual vector $\hat{\varepsilon}_N = \mathbf{y}_N - \mathbf{Z}_N\hat{\gamma}_N$ can be written as

$$\begin{aligned}\hat{\varepsilon}_N &= \mathbf{y}_N - \mathbf{e}_N\hat{\alpha}_N - \mathbf{X}_N\hat{\beta}_N - \hat{\lambda}_N\mathbf{W}_N\mathbf{y}_N \\ &= \mathbf{y}_N(1+\hat{\lambda}_Na_N) - \mathbf{e}_N(\hat{\alpha}_N + N\hat{\lambda}_Na_N\bar{y}_N) - \mathbf{X}_N\hat{\beta}_N\end{aligned}$$

Substituting the expressions for $\hat{\alpha}_N$, $\hat{\beta}_N$, and $\hat{\lambda}_N$ given in part (a) of the theorem, it is then readily seen that $\hat{\varepsilon}_N = \mathbf{0}$. The 2SLS objective function is given by

$$\hat{\varepsilon}_N'\mathbf{H}_N(\mathbf{H}_N'\mathbf{H}_N)^{-1}\mathbf{H}_N'\hat{\varepsilon}_N$$

Since $\mathbf{H}_N(\mathbf{H}_N'\mathbf{H}_N)^{-1}\mathbf{H}_N'$ is positive semidefinite $\hat{\varepsilon}_N'\mathbf{H}_N(\mathbf{H}_N'\mathbf{H}_N)^{-1}\mathbf{H}_N'\hat{\varepsilon}_N \geq 0$. The 2SLS objective function is thus clearly minimized for $\hat{\varepsilon}_N = \mathbf{0}$. Since we have just shown that $\hat{\varepsilon}_N = \mathbf{0}$ for $\hat{\gamma}_N' = (N\bar{y}_N, \mathbf{0}, -1/a_N)$ it follows that $\hat{\gamma}_N$ is indeed the vector of 2SLS estimators.

REMARK 1: Since the diagonal elements of $\mathbf{W}_N$ are all zero, the nondiagonal elements are all equal, and the sample size is N, one would typically take $a_N = \frac{1}{N-1}$ in the above illustrative cases, for example, villages. We have specified $\mathbf{W}_N$ in terms of a_N for purposes of generality. Kelejian and Prucha (2002) demonstrated that if $a_N = \frac{1}{N-1}$, then $p\lim_{N\to\infty}|\hat{\lambda}_N| = \infty$, and hence they noted that the 2SLS estimator in this case is inconsistent.[5] Their result is clearly consistent with Theorem 1 since $|\hat{\lambda}_N| = 1/a_N = N-1$.

Theorem 1 extends the consistency result in Kelejian and Prucha (2002) by giving explicit finite sample results for the 2SLS estimators of the elements of the parameter vector $\gamma = (\alpha, \beta', \lambda)'$. Given the model in (1) and the weighting matrix in (4) it should be clear that any other estimators that are defined as a minimizer of a positive semidefinite quadratic form of the disturbances, for example, OLS, will be identical to the 2SLS estimators. Finally we note that the results of the above theorem do not contradict the consistency results for the 2SLS estimator given in Kelejian and Prucha (1998) because their assumptions rule out a weighting matrix such as the one considered in (4).

REMARK 2: Part (a) of Theorem 1 implies that the single panel data model in (1) with $\mathbf{W}_N$ specified as in (4)[6] is not a "useful" one, and indeed, should be

[5]We note that our model specification treats all parameters as unrestricted. Correspondingly, the 2SLS estimator defined in (3), and the one considered in Kelejian and Prucha (2002), do not incorporate parameter restrictions.

[6]We stress that Theorem 1 relates to a single panel data model because Kelejian and Prucha (2002) show that a panel data extension of the model with W_N specified as in (4) can be consistently estimated by 2SLS if fixed effects are not considered. See Section 4 of this paper for further clarifications and results.

avoided![7] This should also be clear from part (b), which implies that the usual estimator for σ^2 is given by $\hat{\sigma}_N^2 = N^{-1}\hat{\varepsilon}_N'\hat{\varepsilon}_N = 0$, and so typical test statistics are not defined because they require division by zero. The suggestion is that results relating to them obtained in practice will, most likely, be based on rounding errors. Finally, we note that part (b) implies that $R^2 = 1$.

REMARK 3: Theorem 1 also has implications concerning 2SLS estimation of model (1) for situations where the weighting matrix is not observed, but instead is parameterized in terms of observable variables and then its parameters are estimated by a nonlinear 2SLS procedure along with the regression parameters. Unfortunately, for a wide variety of parameterizations the results of such an estimation procedure would not be consistent. To see the issues involved, suppose for the moment that the (i, j)th element of the weighting matrix is specified as

$$w_{ii,N}(c) = 0; \quad w_{ij,N}(c) = \frac{1}{1 + d_{ij,N}^c}, \quad i \neq j \tag{6}$$

where $d_{ij,N} \geq 0$ is an observable distance measure between the (i)th and (j)th units, and $c \geq 0$ is a parameter to be estimated.[8] Let $\mathbf{W}_N(c)$ be the $N \times N$ weighting matrix for this case, and let $\mathbf{Z}_N(c) = (\mathbf{e}_N, \mathbf{X}_N, \mathbf{W}_N(c)\mathbf{y}_N)$ be the regressor matrix corresponding to this more general version of the model in (1). Let $\tilde{\varepsilon}_N(\tilde{c}_N) = \mathbf{y}_N - \mathbf{Z}_N(\tilde{c}_N)\tilde{\gamma}_N$ where $\tilde{\gamma}_N' = (\tilde{\alpha}_N, \tilde{\beta}_N', \tilde{\lambda}_N)$. Then the nonlinear 2SLS estimator for this model would minimize

$$\tilde{\varepsilon}_N'(\tilde{c}_N)\mathbf{H}_N(\mathbf{H}_N'\mathbf{H}_N)^{-1}\mathbf{H}_N'\tilde{\varepsilon}_N(\tilde{c}_N) \tag{7}$$

w.r.t. $\hat{\alpha}_N$, $\tilde{\beta}_N$, $\tilde{\lambda}_N$, and $\tilde{c}_N$. Unfortunately, as should be clear from Theorem 1, the results of the minimization will lead to

$$\tilde{c}_N = 0, \quad (\hat{\alpha}_N, \tilde{\beta}_N', \tilde{\lambda}_N) = (N\bar{y}, \mathbf{0}, -2)$$

since $c = 0$ implies uniform weights (in this case, $a_N = 1/2$) and this, in turn, implies via part (b) of Theorem 1 that the minimized value in (7) is zero. We note that this negative result would not be altered for other specifications of $w_{ij,N}$, as long as there are admissible parameter values such that all nondiagonal weights are equal.

In a sense there is a corollary to Remark 3. Specifically, suppose in a model such as (1) the weighting matrix is not known a priori and the researcher considers various observable specifications of it in terms of, say, various distance measures, for example, trade shares, geographic distance, etc. Remark 3 suggests that if that researcher then selects the specification of the weighting

[7] As a point of interest we note that the maximum likelihood estimator also yields "peculiar" results, even if the parameter space for λ is restricted; for example, see Kelejian and Prucha (2002).

[8] Among other things, the proofs given in Kelejian and Prucha (1998, 1999) require the elements of the weighting matrix to be uniformly bounded in absolute value. The specification in (6) is not taken as $w_{ij,N}(c) = 1/d_{ij,N}^c$ because of this condition.

matrix on the basis of the standard R^2 statistic, the results may be biased in the direction of the matrix with the "most uniform weights." Clearly, the suggestion is that the R^2 measure of fit should not be used to determine the weighting matrix.

Of course, if additional "identifying" information is available, then it may be possible to consistently estimate a model such as (1) with $\mathbf{W}_N$ specified parametrically. Such information could, for example, be parameter restrictions, or the availability of additional behavioral equations that contain some of the parameters of (1), or those defining $\mathbf{W}_N$.

4. A PANEL DATA EXTENSION

Kelejian and Prucha (2002) considered a panel data extension of the model in (1) and (4). Their extension did not consider fixed effects. Kelejian and Prucha showed that, in this case, the 2SLS estimator is consistent. In the following we demonstrate that this consistency result does not extend to panel data models if the specification includes fixed effects. This is important to note because fixed effects are often considered in panel data models.

In particular, consider the following balanced fixed effects panel data model:

$$\mathbf{y}_{t,N} = \mathbf{e}_N\alpha_t + \mathbf{X}_{t,N}\beta + \lambda\mathbf{W}_N\mathbf{y}_{t,N} + \varepsilon_{t,N}, \quad t = 1, \ldots, T \tag{8}$$

where $\mathbf{y}_{t,N}$ is the $N \times 1$ vector of observations on the dependent variable in the (t)th "panel" (in village t, in school t, etc.), $\mathbf{X}_{t,N}$ is the $N \times k$ matrix of observations on the exogenous variables in the (t)th "panel," $\varepsilon_{t,N}$ is the $N \times 1$ vector of disturbance terms in the (t)th "panel," $\mathbf{W}_N$ is defined above in (4), α_t, $t = 1, \ldots, T$, are a scalar "fixed effects" parameters which are defined with respect to the panels,[9] and λ, β, and $\mathbf{e}_N$ are defined as above. The model considered in Kelejian and Prucha (2002) corresponds to (8) with $\alpha_t = \alpha$, $t = 1, \ldots, T$.

In the following we assume that $T > 1$, but is finite. In order to express (8) in stacked form, let

$$\mathbf{y}_{NT} = [\mathbf{y}'_{1,N}, \ldots, \mathbf{y}'_{T,N}]'$$
$$\mathbf{X}_{NT} = [\mathbf{X}'_{1,N}, \ldots, \mathbf{X}'_{T,N}]'$$
$$\varepsilon_{NT} = [\varepsilon'_{1,N}, \ldots, \varepsilon'_{T,N}]'$$
$$\alpha' = (\alpha_1, \ldots, \alpha_T)$$

Given this notation, the model in (8) can be expressed as

$$\begin{aligned} \mathbf{y}_{NT} &= (\mathbf{I}_T \otimes \mathbf{e}_N)\alpha + \mathbf{X}_{NT}\beta + \lambda(\mathbf{I}_T \otimes \mathbf{W}_N)\mathbf{y}_{NT} + \boldsymbol{\varepsilon}_{NT} \\ &= \mathbf{Z}_{NT}\gamma + \varepsilon_N \\ \mathbf{Z}_{NT} &= [\mathbf{I}_T \otimes \mathbf{e}_N, \mathbf{X}_{NT}, (\mathbf{I}_T \otimes \mathbf{W}_N)\mathbf{y}_{NT}], \quad \gamma' = (\alpha', \beta', \lambda) \end{aligned} \tag{9}$$

[9] As an illustration, if t relates to villages, then the fixed effects are village effects.

By arguments analogous to those put forth above for the case of a single panel $(\mathbf{I}_T \otimes \mathbf{W}_N)\mathbf{y}_{NT}$ in (9) would typically be viewed as correlated with the innovations $\boldsymbol{\varepsilon}_{NT}$, and thus $(\mathbf{I}_T \otimes \mathbf{W}_N)\mathbf{y}_{NT}$ would be treated as endogenous. Given this the researcher might attempt to estimate model (9) by 2SLS based on some matrix of instruments, say, $\mathbf{H}_{NT} = [\mathbf{I}_T \otimes \mathbf{e}_N, \mathbf{X}_N, \mathbf{G}_{NT}]$, where $\mathbf{G}_{NT}$ is an $NT \times r$ matrix, $r \geq 1$, and $\mathbf{H}_{NT}$ has full column rank.[10] Let $\hat{\mathbf{Z}}_{NT} = \mathbf{P}_{NT}\mathbf{Z}_{NT}$ where $\mathbf{P}_{NT} = \mathbf{H}_{NT}(\mathbf{H}'_{NT}\mathbf{H}_{NT})^{-1}\mathbf{H}'_{NT}$. Then, assuming that $\hat{\mathbf{Z}}_{NT}$ has full column rank, the 2SLS estimator of γ in (9) is

$$\hat{\gamma}_N = (\hat{\alpha}_{1,N}, \ldots, \hat{\alpha}_{T,N}, \hat{\beta}'_N, \hat{\lambda}_N)' = (\hat{\mathbf{Z}}'_{NT}\hat{\mathbf{Z}}_{NT})^{-1}\hat{\mathbf{Z}}_{NT}\mathbf{y}_{NT} \tag{10}$$

Our main result in this section is given in Theorem 2. Its implications are discussed in the remarks that follow.

THEOREM 2: *Assume the model in (8) and its stacked form in(9), and that $\hat{\mathbf{Z}}_{NT}$ has full column rank so that $\hat{\gamma}_N$ in (10) can be calculated. Let $\bar{y}_{t,N} = \mathbf{e}'_N\mathbf{y}_{t,N}/N$ denote the sample mean of $\mathbf{y}_{t,N}$ in the (t)th panel. If the weighting matrix is of the form given in (4) then*

(a) $\hat{\gamma}_N = (\hat{\alpha}_{1,N}, \ldots, \hat{\alpha}_{T,N}, \hat{\beta}'_N, \hat{\lambda}_N)' = (N\bar{y}_{1,N}, \ldots, N\bar{y}_{T,N}, \mathbf{0}, -1/a_N)'$

(b) $\hat{\varepsilon}_{NT} = \mathbf{y}_{NT} - \mathbf{Z}_{NT}\hat{\gamma}_N = \mathbf{0}$.

Proof of Theorem 2. Analogous as in the proof of Theorem 1, it suffices to show that $\hat{\varepsilon}_{NT} = \mathbf{y}_{NT} - \mathbf{Z}_{NT}\hat{\gamma}_N = \mathbf{0}$ for $\hat{\gamma}_N = (N\bar{y}_{1,N}, \ldots, N\bar{y}_{T,N}, \mathbf{0}, -1/a_N)'$, since this also implies that $\hat{\gamma}_N$ is indeed the vector of 2SLS estimators. Clearly $\hat{\varepsilon}_{NT} = [\hat{\varepsilon}_{1,N}, \ldots, \hat{\varepsilon}_{T,N}]'$, where

$$\hat{\varepsilon}_{t,N} = \mathbf{y}_{t,N} - \mathbf{e}_N\hat{\alpha}_{t,N} - \mathbf{X}_{t,N}\hat{\beta}_N - \hat{\lambda}_N\mathbf{W}_N y_{t,N}$$

denotes the residuals corresponding to the tth panel. Substitution of expression (4) for $\mathbf{W}_N$ yields

$$\hat{\varepsilon}_{t,N} = \mathbf{y}_{t,N}(1 + \hat{\lambda}_N a_N) - \mathbf{e}_N(\hat{\alpha}_{t,N} + N\hat{\lambda}_N a_N \bar{y}_{t,N}) - \mathbf{X}_{t,N}\hat{\beta}_N$$

Upon substitution of the expressions for $\hat{\alpha}_{t,N} = N\bar{y}_{t,N}$, $\hat{\beta}_N = \mathbf{0}$, and $\hat{\lambda}_N = -1/a_N$ it is readily seen that indeed $\hat{\varepsilon}_{t,N} = \mathbf{0}$, and thus $\hat{\varepsilon}_{NT} = \mathbf{0}$.

REMARK 4: As in the single panel case, it should be clear that part (a) of Theorem 2 implies that the model in (8) and (9) with $\mathbf{W}_N$ specified as in (4) is not a "useful" one, and therefore should be avoided.

REMARK 5: Theorem 2 assumes that the panel is balanced. The results of Theorem 2 will not hold in the unbalanced panel case, for example, for the case in which the model in (8) is generalized to

$$\mathbf{y}_{t,N_t} = \alpha_t\mathbf{e}_{N_t} + \mathbf{X}_{t,N_t}\beta + \lambda\mathbf{W}_{t,N_t}\mathbf{y}_{t,N_t} + \varepsilon_{t,N_t}; \quad t = 1, \ldots, T$$

[10] One possible selection of $\mathbf{G}_{NT}$ would be the linearly independent columns of $((\mathbf{I}_T \otimes \mathbf{W}_N) \times \mathbf{X}_{NT}, (\mathbf{I}_T \otimes \mathbf{W}_N^2)\mathbf{X}_{NT}, \ldots, (\mathbf{I}_T \otimes \mathbf{W}_N^q)\mathbf{X}_{NT})$ where typically $q \leq 2$.

where $\mathbf{y}_{t,N_t}$ is the $N_t \times 1$ vector of observations on the dependent variable in the (t)th panel, $\mathbf{X}_{t,N_t}$ is the $N_t \times k$ matrix of observations on the exogenous variables in the (t)th panel, e_{N_t} is an $N_t \times 1$ vector of unit elements, ε_{t,N_t} is the corresponding disturbance vector, and

$$\mathbf{W}_{t,N} = a_{N_t}[\mathbf{e}_{N_t}\mathbf{e}'_{N_t} - \mathbf{I}_{N_t}]$$

where a_{N_t} is a scalar whose value would, in our setting, typically be $\frac{1}{N_t - 1}$. The reason for this is as follows. Let $\hat{\alpha}_{t,N}$, $\tilde{\beta}_N$, and $\tilde{\lambda}_N$ be estimators for α_t, β, and λ. Then the corresponding estimated disturbances in the (t)th panel are given by

$$\tilde{\varepsilon}_{t,N} = \mathbf{y}_{t,N}(1 + \tilde{\lambda}_N a_{N_t}) - \mathbf{e}_N(\tilde{\alpha}_{t,N} + N\tilde{\lambda}_N a_{N_t}\bar{y}_{t,N}) - \mathbf{X}_{t,N}\tilde{\beta}_N$$

Since $\tilde{\lambda}_N$ does not depend on t it is readily seen that it will not be possible to choose values for $\tilde{\lambda}_N$ and $\tilde{\alpha}_{t,N}$ such that $1 + \tilde{\lambda}_N a_{N_t} = 0$ and $\tilde{\alpha}_{t,N} + N\tilde{\lambda}_N a_{N_t}\bar{y}_{t,N} = 0$ for all t if a_{N_t} varies with t. Because of this it will generally not be possible to find estimators such that the estimated disturbances are all zero.

5. CONCLUSION

We have shown that estimation problems exist in spatial models containing a spatially lagged-dependent variable if the weighting matrix has uniform weights and if an intercept is present. One implication of this result is that serious estimation problems may arise in cases in which the weighting matrix is parameterized and its parameters are estimated along with the regression parameters by an IV technique. In contrast to the results in Kelejian and Prucha (2002), these problems exist even in the multiple panel case if fixed effects that relate to the panels are considered, and if there are the same number of observations in all of the panels, as they might be if the panels relate to time observations on a given unit—such as a village.

REFERENCES

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

Audretsch, David, and Maryann Feldmann. 1996. "R&D Spillovers and the Geography of Innovation and Production," *American Economic Review*, 86, 630–640.

Bell, Katheen P., and Nancy E. Bockstael. 2000. "Applying the Generalized Moments Estimation Approach to Spatial Problems Involving Micro Level Data," *Review of Economics and Statistics*, 82, 72–82.

Case, Ann. 1992. "Neighborhood Influence and Technological Change," *Regional Science and Urban Economics*, 22, 491–508.

Cliff, Andrew, and J. Keith Ord. 1973. *Spatial Autocorrelation*. London: Pion.

Cliff, Andrew, and J. Keith Ord. 1981. *Spatial Processes, Models and Applications*. London: Pion.

Cressie, Noel A. C. 1993. *Statistics of Spatial Data*. New York: Wiley.

Das, Debabrata, Harry H. Kelejian, and Ingmar R. Prucha. 2003. "Small Sample Properties of Estimators of Spatial Autoregressive Models with Autoregressive Disturbances," *Papers in Regional Science*, 82, 1–26.

Kapoor, Mudit. 2003. "Panel Data Models with Spatial Correlation: Estimation Theory and an Empirical Investigation of the US Wholesale Gasoline Industry," Ph.D. dissertation, Department of Economics, University of Maryland.

Kelejian, Harry H., and Ingmar R. Prucha. 1998. "A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances," *Journal of Real Estate Finance and Economics*, 17, 99–121.

———. 1999. "A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model," *International Economic Review*, 40, 509–533.

———. 2002. "2SLS and OLS in a Spatial Autoregressive Model with Equal Spatial Weights," *Regional Science and Urban Economics*, 32, 691–707.

Lee, Lung-Fei. 2002. "Consistency and Efficiency of Least Squares Estimation for Mixed Regressive, Spatial Autoregressive Models," *Econometric Theory*, 18, 252–277.

Lee, Lung-Fei. 2003. "Best Spatial Two-Stage Least Squares Estimators for a Spatial Autoregressive Model with Autoregressive Disturbances," *Econometric Reviews*, 22, 307–335.

Rey, Sergio J., and Marlon G. Boarnet. 2004. "A Taxonomy of Spatial Econometric Models for Simultaneous Equation Systems," in a L. Anselin, R. J. G. M. Florax, and S. J. Rey (eds.), *Advances in Spatial Econometrics*. New York: Springer, pp. 99–119.

Splitstoser, Gianna. 1999. "A Spatial Model for Legislator Ideology," Term Paper, Department of Economics, University of Maryland.

Pinkse, Joris, Margaret E. Slade, and Craig Brett. 2002. "Spatial Price Competition: A Semiparametric Approach," *Econometrica*, 70, 1111–1153.

Yuzefovich, Yevgeny. 2003. "Two Essays on Spatial Econometrics," Ph.D. dissertation. Department of Economics, University of Maryland.