2SLS and OLS in a spatial autoregressive model with equal spatial weights

Harry H. Kelejian*, Ingmar R. Prucha

Department of Economics, University of Maryland, College Park, MD 20742, USA

Received 14 November 2000; received in revised form 4 December 2001; accepted 4 January 2002

Abstract

The paper considers a Cliff–Ord type spatial model with a spatially lagged dependent variable and a row normalized weighting matrix with equal weights. We show that the 2SLS and OLS estimators are inconsistent unless panel data are available. The weighting matrix in question is one which would naturally be considered if all units are neighbors to each other, and there is no other reasonable or observable measure of distance between them. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Spatial autoregressive models; Row normalized and equal spatial weights; Ordinary least squares; Two stage least squares; Panel data

JEL classification: C0; C2

1. Introduction

We consider estimation issues in linear Cliff–Ord type spatial models which contain a spatially lagged dependent variable.1 We give both a ‘negative’ result that serves as a warning for a potential pitfall, as well as a ‘positive result’. More
specifically, we first show that under typical conditions the two-stage least squares (2SLS) and the ordinary least squares (OLS) estimators are inconsistent if the weighting matrix is row normalized and has equal weights, and data are only available for a single cross section. We then show that this problem typically does not occur if panel data are available. Ironically, if two or more panels of data are available, both the 2SLS and OLS estimator are not only consistent but are also efficient within the class of instrumental variable estimators. These results are important to note because the weighting matrix considered is one which would naturally suggest itself if all units are neighbors to each other and there is no other natural or observable measure of distance. Cases which may be consistent with this are ones in which all cross sectional units interact in a confined space. Such a matrix was considered by Splitstoser (1999) in a study of spatial interdependence involving the ideology of legislators. It was also considered by Case (1992) in a panel data study of the adoption of new technologies by farmers, and by Lee (1999b) in a study of the properties of least squares estimators in linear spatial models.

The paper is organized as follows. In Section 2 we specify the considered Cliff–Ord type model for a single cross section, and demonstrate the inconsistency of 2SLS and OLS within this setting. Section 3 extends the framework to panel data and demonstrates that in contrast to the previous case, the model parameters can now be estimated consistently by 2SLS and OLS. We also derive the limiting distribution of the estimators. All proofs are relegated to Appendix A.

The following notation and conventions are useful: let \( v \) and \( P \) be, respectively, a vector and matrix. Then we denote the \( i \)th element of \( v \) as \( v_i \) and the \( i,j \)th element of \( P \) as \( p_{ij} \). The same convention holds for vectors and matrices which do not involve the index \( N \), in which case the index \( N \) is suppressed. We will say that the elements of the sequence of matrices \( P \) are uniformly bounded in absolute value if

\[ |p_{ij,N}| \leq C < \infty \]

for all \( 1 \leq i, j \leq N; N \geq 1 \), where the constant \( C \) does not depend on any of the indices.

2. 2SLS and OLS in the case of a single cross section

2.1. The model

Consider the following cross sectional spatial model \( (N \geq 1) \)

---

\(^7\)The OLS estimator may, of course, be viewed as a special case of an instrumental variable estimator.

\(^8\)Other relevant cases may relate to student performance in a school, certain interactions of workers in a firm, etc.
\[ y_N = \alpha e_N + X_N \beta + \lambda W_N y_N + e_N \]
\[ = Z_N \gamma + e_N \]  
(1)

where \( Z_N = (e_N, X_N, W_N y_N) \) and \( \gamma = (\alpha, \beta', \lambda)' \). Here \( y_N \) is the \( N \times 1 \) vector of observations on the dependent variable, \( e_N \) is the \( N \times 1 \) vector of unit elements, \( X_N \) is an \( N \times k \) matrix of observations on \( k \) exogenous variables, \( W_N \) is an \( N \times N \) weighting matrix, \( e_N \) is the disturbance vector, \( \alpha \) and \( \lambda \) are scalar parameters, and \( \beta \) is a \( k \times 1 \) vector of parameters. Our analysis is conditional on the realized values of the exogenous variables and so the matrices \( X_N \) will be taken as matrices of constants.

Our assumptions for (1) are the standard ones, except for the specification of the weighting matrix. Let \( D_N = (e_N, X_N) \). Then, we assume

**Assumption 1.** \( |\lambda| < 1 \). Also, for all \( 1 \leq i \leq N, \ N \geq 1 \) the error terms \( e_{i,N} \) are identically distributed with mean zero and finite variance \( \sigma^2 \). In addition, for each \( N \) the error terms \( e_{1,N}, \ldots, e_{N,N} \) are independently distributed.

**Assumption 2.** The elements of the sequence of matrices \( X_N \) are uniformly bounded in absolute value, and \( Q_{DD} = \lim_{N \to \infty} N^{-1} D_N^t D_N \) is finite and nonsingular.

It proves convenient to decompose \( Q_{DD} \) as
\[ Q_{DD} = \begin{bmatrix} 1 & \mu_X' \\ \mu_X & Q_{XX} \end{bmatrix}. \]

Clearly \( \mu_X \) is a \( k \times 1 \) vector, and \( Q_{XX} \) is a \( k \times k \) matrix. Let \( \bar{x}_N = N^{-1} X_N' e_N \) be the vector of sample means of the exogenous regressors. Then for future reference note that given the above assumption \( \bar{x}_N \to \mu_X \) and \( N^{-1} X_N' e_N \to Q_{XX} \).

**Assumption 3.** \( W_N = (1/(N-1)) J_N - \mu_X e' \) \( J_N \geq 1 \), where \( J_N = e_N e'_N \).

This assumption implies that
\[ W_{ij,N} = \begin{cases} 1/(N-1) & \text{for } i \neq j \\ 0 & \text{for } i = j \end{cases} \]

and thus, among other things, the model in (1) relates each element of \( y_N \) to the average of the other \( N - 1 \) elements.

**Remark.** It seems of interest to briefly compare the above spatial autoregressive model with the widely used class of mixed models that combines fixed and random effects. Following, e.g., Harville (1988), a mixed model is defined as
\[ y_N = F_N \delta + R_N s_N + e_N \]
where \( y_n \) denotes the vector of observations on the dependent variable, \( F_n \) and \( R_n \) are known nonstochastic matrices (e.g., observations on exogenous regressors), \( \delta \) is a vector of unknown parameters, \( s_n \) is a vector of unobserved random effects with zero mean, \( e_n \) is the unobserved disturbance vector, and \( s_n \) and \( e_n \) are uncorrelated. In (1) \( AW_n y_n \) can be expressed as \( W_n (\lambda y_n) \). Therefore, if we define \( F_n = (e_n, X_n) \), \( \delta = (\alpha, \beta')' \), \( R_n = W_n \) and \( s_n = \lambda y_n \) it may seem, at first glance, that we can express the spatial autoregressive model (1) in the above form of a mixed model. However, upon closer inspection we see that the two models are substantially different. Among other things, in (1) the ‘random effects’ \( s_n = \lambda y_n \) are observed (up to a common scalar), have nonzero mean, and clearly \( s_n = \lambda y_n \) and \( e_n \) are correlated. This is reflective of the fact that the spatial autoregressive model represents a simultaneous system, while the above mixed model does not. The subsequent discussion pertains to the spatial autoregressive model as defined before this paragraph.

2.2. Some useful results

For future reference we note that, as is easily verified,

\[
(I - \lambda W_n)^{-1} = \delta_{1N}J_N + \delta_{2N}I_N
\]

where

\[
\delta_{1N} = \frac{\lambda}{(N - 1 + \lambda)(1 - \lambda)} \quad \text{and} \quad \delta_{2N} = \frac{N - 1}{N - 1 + \lambda}.
\]

The result in (2) can be used to obtain the reduced form for the spatial lag of the dependent variable in (1). Specifically, (1) implies that \( W_n y_n = W_n (I - \lambda W_n)^{-1} [\alpha e_n + X_n \beta + \varepsilon_n] \) and so, in light of (2),

\[
W_n y_n = c_n e_n + \phi_{2N} X_n \beta + \phi_{2N} \varepsilon_n
\]

where

\[
c_n = N \phi_{1N} [\alpha + X_n' \beta + \varepsilon_n] + \alpha \phi_{2N},
\]

\[
\phi_{1N} = \delta_{1N} + \frac{\delta_{2N}}{N - 1}, \quad \phi_{2N} = -\frac{\delta_{2N}}{N - 1}.
\]

and \( \varepsilon_n = N^{-1} e_{n}' e_n \). From the expressions in (3) we see that

\[
\phi_{1N} \to 0, \quad \phi_{2N} \to 0 \quad \text{and} \quad N \phi_{1N} \to 1/(1 - \lambda)
\]

as \( N \to \infty \).

Since \( \varepsilon_n \) has zero mean and variance \( \sigma^2/N \), it follows from Chebyshev’s inequality that \( \lim_{N \to \infty} \varepsilon_n = 0 \). Therefore, again for future reference, we note that

\[
p \lim_{N \to \infty} c_n = \frac{\alpha + \mu_n' \beta}{1 - \lambda}.
\]
The results in (4)–(6) suggest that there may be a basic estimation problem in that, asymptotically, the spatial lag \( W_Ny_N \) is proportional to the unit vector, and thus collinear with the intercept. In the following sections we discuss this problem and its effect on the 2SLS and OLS estimator in more detail.

2.3. 2SLS estimation

Since \( W_Ny_N \) is correlated with the disturbance term in (1), see, e.g., Anselin (1988), and Kelejian and Prucha (1998), one suggested method of estimation for \( \gamma \) is 2SLS. Let \( H_N = (e_N, X_N, G_N) \) be the matrix of instruments, where \( G_N \) is an \( N \times r \) matrix of nonstochastic variables, \( r \geq 1 \). We maintain the following assumption concerning those instruments.

**Assumption 4.** The elements of the sequence of matrices \( G_N \) are uniformly bounded in absolute value, and \( Q_{HH} = \lim_{N \to \infty} N^{-1}H_N'X_NH_N \) is finite and nonsingular.

One suggestion for the columns of \( G_N \) given by Kelejian and Prucha (1998) and by Rey and Boarnet (1998) are the spatial lags of \( X_N \) of various orders. We note, however, that the subsequent discussion does not postulate any particular structure for \( G_N \) apart from what is maintained in Assumption 4.

Let \( P_{H_N} = H_N(H_N'H_N)^{-1}H_N \) and note that since \( H_N \) contains \( e_N \) and \( X_N \), \( P_{H_N}e_N = e_N \) and \( P_{H_N}X_N = X_N \). Let

\[
\hat{Z}_N = P_{H_N}Z_N = (e_N, X_N, P_{H_N}W_Ny_N).
\]

The 2SLS estimator of \( \gamma \) is then given by

\[
\hat{\gamma}_{2SLS} = (\hat{Z}_N'\hat{Z}_N)^{-1}\hat{Z}_N'y_N.
\]

The following lemma, whose proof is given in Appendix A, establishes that the 2SLS estimator is inconsistent under the assumptions maintained above.

**Lemma 1.** Given the model in (1), Assumptions 1, 2, 3 and 4,

\[
p \lim_{N \to \infty} \hat{\gamma}_{2SLS,N} \neq \gamma.
\]

On an intuitive level the above lemma can be motivated from the following observations: results given in Amemiya (1985, pp. 245–255) imply that the optimal instrument matrix is \( E(Z_N) \). In terms of our model, however, \( E[\hat{e}_N] = 0 \) and so the results in (4) imply

\[
E(Z_N) = [e_N, X_N, E(c_N)e_N + \phi_{2N}X_N\beta]
\]

where

\[
E(c_N) = \alpha N\phi_{1N} + N\phi_{1N}\bar{e}_N'\beta + \alpha\phi_{2N}.
\]
Since \( E(c_N) \), \( \phi_1 \), and \( \phi_2 \) are scalars, the last column of \( E(Z_N) \) is clearly multicollinear with the first \( k + 1 \), and so \( E(Z_N) \) does not have full column rank.

We also note that the above lemma does not contradict 'positive' results given in Kelejian and Prucha (1998) concerning the consistency and asymptotic normality of the 2SLS estimator in linear spatial models containing a spatially lagged dependent variable. For instance, Eqs. (4)–(6) imply that

\[
Q_{HZ} = p \lim_{N \to \infty} N^{-1}H_N'Z_N = p \lim_{N \to \infty} N^{-1}H_N' \left[ e_N, X_N, \left( \frac{\alpha + \mu_1' \beta}{1 - \lambda} \right) e_N \right],
\]

which shows that \( Q_{HZ} \) has less than full column rank. This violates Assumption 7 maintained in Kelejian and Prucha (1998). We note further that as a consequence

\[
p \lim_{N \to \infty} N^{-1}(\hat{Z}_N' \hat{Z}_N) = p \lim_{N \to \infty} [N^{-1}Z_N'X_N][N^{-1}(H_N'X_N)^{-1}][N^{-1}H_N'Z_N]
= Q_{HZ}Z_N X_N
\]

is singular.

### 2.4. OLS estimation

Lee (1999b) has shown that under certain conditions the OLS estimator of the parameters of a linear spatial model containing a spatially lagged dependent variable is consistent and asymptotically normal. One of the conditions assumed in Lee’s model implies that each element of the spatial weights matrix limits to zero as the sample size \( N \to \infty \). Since that condition is satisfied by our model, one might also consider the least squares estimator of \( \gamma \) in (1), i.e. \( \hat{\gamma}_{OLS,N} = (Z_N'Z_N)^{-1}Z_N'y_N \). However, as seen from the subsequent lemma, under the maintained assumptions the OLS estimator is also inconsistent.

**Lemma 2.** Given the model in (1), and Assumptions 1, 2, 3,

\[
p \lim_{N \to \infty} \hat{\gamma}_{OLS,N} \neq \gamma.
\]

We note that the above lemma does not contradict the results given in Lee (1999b). While at first glance the spatial weights matrix defined in Assumption 3 seems to satisfy Lee’s assumptions, one of Lee’s conditions for the consistency of the OLS estimator is that \( Q_{Zz} = \lim_{N \to \infty} N^{-1}E(Z_N)E(Z_N) \) is a finite nonsingular matrix. Under the assumptions maintained here this condition does not hold. Using (4)–(6) and (9) it is readily seen that

\[
Q_{Zz} = \begin{bmatrix}
1 & \mu_1' & c_x \\
\mu_1 & Q_{xx} & c_x \mu_2 \\
c_x & c_x \mu_1' & c_x^2
\end{bmatrix}
\]
where \( c_5 = (\alpha + \mu_i'\beta)/(1 - \lambda) \), and so \( Q_{zz} \) is singular since its first and third columns are proportional.

### 2.5. Limiting log-likelihood function

As was observed above, the root cause for the inconsistency of the 2SLS and OLS estimator of \( \lambda \) is that asymptotically \( W_{ny} \) is proportional to the unit vector, and thus collinear with the intercept. This suggests that the problem encountered with the 2SLS and OLS estimator is not specific to those estimators. To shed more light on the issue it seems of interest to investigate the limiting log-likelihood function, properly normalized by the sample size. Under normality the normalized log-likelihood function at parameter values \( \alpha, \beta, \lambda, \sigma^2 \) is given by

\[
L_n(\alpha, \beta, \lambda, \sigma^2) = -\frac{1}{2} \ln (2\pi) + L_1n(\alpha, \beta, \lambda, \sigma^2) + L_2n(\alpha, \beta, \lambda, \sigma^2)
\]

where

\[
L_1n(\alpha, \beta, \lambda, \sigma^2) = \frac{1}{N} \ln \|I_N - AW_n\| \\
= \frac{N - 1}{N} \ln \left(1 + \frac{\lambda}{N - 1}\right) + \frac{1}{N} \ln (1 - \lambda),
\]

\[
L_2n(\alpha, \beta, \lambda, \sigma^2) = -\frac{1}{2} \ln (\sigma^2) - \frac{1}{2\sigma^2} \frac{1}{N} \left[ (y - Z\gamma)'[y - Z\gamma] \right]
\]

where \( \gamma = (\alpha', \beta', \lambda)' \). The term \( L_1n(\alpha, \beta, \lambda, \sigma^2) \) represents the log of the absolute value of the determinant of the Jacobian normalized by the sample size. We thank one of the referees for supplying us with the second expression for \( L_1n(\alpha, \beta, \lambda, \sigma^2) \), which makes the maximum likelihood estimator easily computable. From this expression it is also readily seen that

\[
\lim_{N \to \infty} L_1n(\alpha, \beta, \lambda, \sigma^2) = 0.
\]

Thus this term drops out from the normalized log-likelihood function as the sample size tends to infinity. The term \( L_2n(\alpha, \beta, \lambda, \sigma^2) \) is in essence the objective function of the OLS estimator. Observe that in light of (1) and the results in (4)–(6) we have

---

This normalization ensures that the terms involved converge to finite limits; compare, e.g., Gallant (1987) and Pötscher and Prucha (1997) who provide a general analysis of the limiting behavior of extremum estimators.
Thus the limiting normalized log-likelihood function is given by

$$p \lim_{N \to \infty} L_n(\alpha, \beta, \lambda, \sigma^2) = -\frac{1}{2} \ln (2\pi) - \frac{1}{2} \ln (\sigma^2) - \frac{1}{2\sigma^2}(\sigma^2 + (\gamma - \gamma)'Q_{zz}(\gamma - \gamma)).$$

As remarked above, the matrix $Q_{zz}$ is singular, and thus the limiting normalized log-likelihood function does not have a unique maximum at the true parameter values $\alpha, \beta, \lambda, \sigma^2$. In fact, as is readily seen, the function is maximized at any parameter value $\alpha, \beta, \lambda, \sigma^2$ with $\alpha - \alpha + c(\lambda - \lambda) = 0, \beta = \beta, \sigma^2 = \sigma^2$. That is, the limiting log-likelihood function is flat along the set of parameter values satisfying $\alpha = \alpha - c, \lambda + c, \lambda$. This implies that the typical identifiable uniqueness assumption maintained to prove consistency of an estimator does not hold at the true parameter values; compare, e.g., Gallant and White (1988, Chapter 3) and Pötscher and Prucha (1997, Chapter 3 and 4.6). This in turn suggests that also the maximum likelihood estimator, as well as any other estimator, will be inconsistent.

3. 2SLS and OLS in the case of panel data

3.1. The model

Our panel data generalization of the model in (1) is ($1 \leq t \leq T, N \geq 1$)

$$y_{iN} = \alpha e_N + X_{iN}\beta + \lambda W_N y_{iN} + e_{iN} = Z_{iN}\gamma + e_{iN} \quad (10)$$

where $Z_{iN} = (e_{iN}, X_{iN}, W_N y_{iN})$ and $\gamma = (\alpha, \beta', \lambda)'$. Here $y_{iN} = (y_{1iN}, \ldots, y_{NiN})'$ is the $N \times 1$ vector of observations on the dependent variable in the $t$-th ‘setting’, $X_{iN}$ is the $N \times k$ matrix of observations on the exogenous variables in the $t$-th ‘setting’, $e_{iN} = (e_{i1N}, \ldots, e_{iN,N})'$ is the $N \times 1$ vector of disturbance terms in the $t$-th ‘setting’, and $W_N$ is defined above in Assumption 3. These ‘settings’ could relate to time periods, or spatial entities for which there are multiple observations. Examples of such spatial entities would be villages, as in Case (1991), schools, etc. We again condition on the realized values of the exogenous variables, and so we will take the matrices $X_{iN}$ to be matrices of constants. We have assumed a balanced panel for ease of presentation; it will be clear that our
results do not require a balanced panel. In the following we assume the number of settings \( T \) is greater than one. In our asymptotic analysis \( N \) is assumed to tend to infinity.

The model in (10) can be expressed as

\[
y_N = Z_N \gamma + \varepsilon_N
\]

where

\[
y_N = [y_{1N}', \ldots, y_{TN}']',
Z_N = [Z_{1N}', \ldots, Z_{TN}']',
\varepsilon_N = [\varepsilon_{1N}', \ldots, \varepsilon_{TN}']'.
\]

Let \( D_{iN} = (e_{iN}, X_{iN}) \) for \( t = 1, \ldots, T \). Our assumptions for the panel data model are:

**Assumption 5.** \(|\lambda| < 1\). Also, for all \( 1 \leq t \leq T \), \( 1 \leq i \leq N \), and \( N \geq 1 \) the error terms \( \varepsilon_{i,t} \) are identically distributed with mean zero and finite variance \( \sigma^2 \). In addition, for each \( N \geq 1 \) the error terms \( \varepsilon_{11,N}, \ldots, \varepsilon_{i1,N}, \ldots, \varepsilon_{T1,N}, \ldots, \varepsilon_{TN,N} \) are independently distributed.

**Assumption 6.** For each \( t = 1, \ldots, T \) the elements of the sequence of matrices \( X_{iN} \) are uniformly bounded in absolute value and \( Q_{DD,t} = \lim_{N \to \infty} N^{-1} D_{iN}D_{iN} \) is finite and nonsingular.

It proves again convenient to decompose \( Q_{DD,t} \) as

\[
Q_{DD,t} = \begin{bmatrix} 1 & \mu_{s,t} \\ \mu_{s,t} & Q_{XX,t} \end{bmatrix}.
\]

Clearly \( \mu_{s,t} \) is a \( k \times 1 \) vector, and \( Q_{XX,t} \) is a \( k \times k \) matrix. Let \( \bar{x}_{iN} = N^{-1} X_{iN}' e_N \) be the vector of sample means of the exogenous regressors for the \( t \)-th setting. Then for future reference note that given the above assumption \( \bar{x}_{iN} \to \mu_{s,t} \) and \( N^{-1} X_{iN}' X_{iN} \to Q_{XX,t} \).

Observe that the panel data model in (10) corresponds to the model in (1) specified for each of the 'settings' \( t = 1, \ldots, T \). Applying (4)–(6) we see that within each setting the spatial lag \( W_N y_{iN} \) will be asymptotically proportional to the unit vector. However, unless \( \mu_{s,t} \) is the same for each setting, the magnitude of the proportionality factor will vary across settings, and thus the spatial lag in the stacked model (11), that is \( (I_T \otimes W_N) y_N \), will not be proportional to the (stacked) unit vector. This suggests that the availability of panel data should alleviate the basic estimation problem discussed in Section 2 in the case of a single cross section.\(^1\)

\(^1\)It should also be clear, however, that panel data will not alleviate the problem if the intercept in (10) is allowed to vary with the 'setting' \( t \), i.e. if the model has fixed effects.
\[ Q_{ZZ} = \lim_{N \to \infty} N^{-1}E(Z_N)'E(Z_N) \]
\[ = \sum_{t=1}^{T} \begin{bmatrix} \mu_{t,t} & c_{*,t} \\ c_{*,t} & \mu_{t,t}^2 & c_{*,t}^2 & \mu_{t,t}c_{*,t} \end{bmatrix} \]  
(12)

with \( c_{*,t} = (\alpha + \mu_{t,t}'\beta)/(1 - \lambda) \). Our remaining assumption is as follows:

**Assumption 7.** The matrix \( Q_{ZZ} \) is nonsingular.

Clearly, as observed in the previous section, for \( T = 1 \) this assumption would be violated. However, an inspection of the expression for \( Q_{ZZ} \) shows that in general this matrix will not be singular, except in very special cases if \( T > 1 \). Those cases are ruled out by Assumption 7. In particular, Assumption 7 rules out the case in which \( \beta = 0 \), as well as the case in which \( \mu_{t,t} = \mu \) for all \( t = 1, \ldots, T \).

### 3.2. 2SLS and OLS estimation

The form of the instruments we use for the 2SLS estimator is partially based on the results obtained in Section 2. As remarked, the panel data model in (10) corresponds to the model in (1) specified for each of the ‘settings’ \( t = 1, \ldots, T \). Applying (4) to the \( t \)-th ‘setting’ we see that the only deterministic regressors in the reduced form for \( W_NY_{IN} \) are \( e \) and \( X \). The coefficient corresponding to \( e \) in the model solution for \( W_NY_{IN} \) will clearly involve \( x \), and therefore, consistent with our remarks above, will typically not be the same for each ‘setting’. This suggests that the list of instruments underlying the 2SLS procedure should contain \( T \) dummy variables instead of just one constant term. For the purposes of generality we also interact these dummy variables with the regressor matrices \( X_{IN} \), \( t = 1, \ldots, T \). More specifically, the instrument matrix we consider is

\[
H_N = \begin{bmatrix}
D_{IN} & 0 & \cdots & 0 \\
0 & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & 0 \\
0 & \cdots & 0 & D_{TN}
\end{bmatrix}
\]  
(13)

Then, the ‘second’ stage regressor matrix is \( \hat{\mathbf{Z}}_N = P_{H_N}Z_N \), where \( P_{H_N} = H_N(H_N'H_N)^{-1}H_N' \). It is not difficult to see that \( \hat{\mathbf{Z}}_N \) can be expressed as

\[
\hat{\mathbf{Z}}_N = [\hat{Z}_{1N}', \ldots, \hat{Z}_{TN}']
\]  
(14)

with \( \hat{Z}_{IN} = P_{D_{IN}}Z_{IN} \) and \( P_{D_{IN}} = D_{IN}(D_{IN}'D_{IN})^{-1}D_{IN}' \). Since \( D_{IN} \) includes both \( e_N \) and \( X_{IN} \), it follows that \( P_{D_{IN}}e_N = e_N \) and \( P_{D_{IN}}X_{IN} = X_{IN} \). Therefore

\[
\hat{Z}_{IN} = (e_N, X_{IN}, P_{D_{IN}}W_{N'}Y_{IN}).
\]  
(15)
The 2SLS estimator of $\gamma$ based on the panel data model in (14) is now given by
\[ \hat{\gamma}_{2SLS,N} = (\tilde{Z}_N'\tilde{Z}_N)^{-1}\tilde{Z}_N'y_N. \] (16)

The corresponding least squares estimator of $\gamma$ is
\[ \hat{\gamma}_{OLS,N} = (Z_N'Z_N)^{-1}Z_N'y_N. \] (17)

The next theorem, whose proof is given in Appendix A, establishes the asymptotic properties of the 2SLS and OLS estimator.

**Theorem 1.** Given the model (10) and Assumptions 3, 5–7

(a) $N^{1/2}(\hat{\gamma}_{2SLS,N} - \gamma) \overset{D}{\to} N(0, \sigma^2 Q_{ZZ}^{-1})$ as $N \to \infty$.
(b) $p \lim_{N \to \infty} N^{-1}(\tilde{Z}_N'\tilde{Z}_N) = \lim_{N \to \infty} N^{-1}E(Z_N'E(Z_N) = Q_{ZZ}$.
(c) $p \lim_{N \to \infty} N^{1/2}(\hat{\gamma}_{2SLS,N} - \hat{\gamma}_{OLS,N}) = 0$.
(d) $p \lim_{N \to \infty} \hat{\sigma}_N^2 = \sigma^2$ where
\[ \hat{\sigma}_N^2 = (NT)^{-1}(y_N - Z_N\hat{\gamma}_{2SLS,N})'(y_N - Z_N\hat{\gamma}_{2SLS,N}). \]

A number of points should be noted about Theorem 1. Firstly, part (a) implies that the 2SLS estimator is consistent and asymptotically normal. Part (b) implies that the 2SLS estimator is the efficient instrumental variable estimator, since the limiting variance–covariance matrix is the inverse of the limit of the second moment of the mean of the regressor matrix. Part (c) implies that the 2SLS and OLS estimators are asymptotically equivalent. We note that in proving this result we establish, among other things, that
\[ p \lim_{N \to \infty} N^{-1}(\tilde{Z}_N'\tilde{Z}_N) = p \lim_{N \to \infty} N^{-1}Z_NZ_N. \] (18)

Part (d) implies that the typical estimator of the variance of the disturbance term is consistent.

Theorem 1 suggests that finite sample inferences concerning $\gamma$ can be based on the approximation
\[ \hat{\gamma}_{2SLS,N} \equiv N(\gamma, \hat{\sigma}_N^2(\tilde{Z}_N'\tilde{Z}_N)^{-1}), \]

or via (18)
\[ \hat{\gamma}_{2SLS,N} \equiv N(\gamma, \hat{\sigma}_N^2(Z_N'Z_N)^{-1}). \]

In further research it may be of interest to analyze the small sample properties of the 2SLS and OLS estimator, and the quality of the approximate distribution derived from large sample theory, via a Monte Carlo study.
Acknowledgements

We thank the referees for helpful suggestions. We also gratefully acknowledge financial support from the National Science Foundation through grant SES-0001780.

Appendix A

Proof of Lemma 1

Let \( \hat{\beta}_{2SLS,N} = (\hat{\alpha}_{2SLS,N}, \hat{\beta}_{2SLS,N}, \hat{\lambda}_{2SLS,N})' \). To prove Lemma 1 it suffices to show that \( p \lim_{n \to \infty} \hat{\lambda}_{2SLS,N} \neq \lambda \). Let \( P_{D_\lambda} = D_\lambda (D_\lambda' D_\lambda)^{-1} D_\lambda' \). Then, by partitioned matrix inversion, \( \hat{\lambda}_{2SLS,N} \) can be expressed as

\[
\hat{\lambda}_{2SLS,N} = \left( [I_n - P_{D_\lambda}] P_{H_\lambda} W_n y_N \right)' \left( [I_n - P_{D_\lambda}] P_{H_\lambda} W_n y_N \right)^{-1} \times \left( [I_n - P_{D_\lambda}] P_{H_\lambda} W_n y_N \right) y_N. \tag{A.1}
\]

Note that \( [I_n - P_{D_\lambda}] D_N = 0 \); also, since \( H_N \) contains the columns of \( D_N \), \( P_{H_N} D_N = D_N \) and so

\[
P_{H_N} P_{D_\lambda} = P_{D_\lambda}. \tag{A.2}
\]

Therefore, replacing \( y_N \) by its model expression in (1), and then using (A.2), it follows from (A.1) that

\[
\hat{\lambda}_{2SLS,N} = \lambda + \left( [W_N y_N]' (P_{H_N} - P_{D_\lambda}) [W_N y_N] \right)^{-1} \times \left( [W_N y_N]' (P_{H_N} - P_{D_\lambda}) e_N \right) = \lambda + \frac{\left( [W_N y_N]' (P_{H_N} - P_{D_\lambda}) e_N \right)}{\left( [W_N y_N]' (P_{H_N} - P_{D_\lambda}) [W_N y_N] \right)} \tag{A.3}
\]

In light of (4), \( W_N y_N \) is linear in \( e_N, X_N, \) and \( e_N \). Since both \( H_N \) and \( D_N \) contain \( e_N \) and \( X_N \) it follows that \( P_{H_N} e_N = e_N, P_{H_N} X_N = X_N, P_{D_N} e_N = e_N, \) and \( P_{D_N} X_N = X_N \). Therefore, from (4)

\[
P_{H_N} W_N y_N = c_N e_N + \phi_{2N} X_N \beta + \phi_{2N} P_{H_N} e_N \tag{A.4}
\]

and so the numerator in (A.3) is

\[
(W_N y_N)' (P_{H_N} - P_{D_\lambda}) e_N = \phi_{2N} e_N P_{H_N} e_N = e_N P_{H_N} e_N \tag{A.5}
\]

The result in (4) and (A.4) imply that the denominator in (A.3) is expressible as

\[
(W_N y_N)' (P_{H_N} - P_{D_\lambda}) (W_N y_N) = \phi_{2N} (W_N y_N)' (P_{H_N} - P_{D_\lambda}) e_N = \phi_{2N}^2 e_N P_{H_N} e_N = e_N P_{D_N} e_N \tag{A.6}
\]
It follows from (A.3), (A.5) and (A.6) that
\[ \hat{\lambda}_{2SLS,N} = \lambda + \frac{1}{\phi_{2N}} \]  
and so via (5)
\[ p \lim_{N \to \infty} |\hat{\lambda}_{2SLS,N}| = \infty \neq |\lambda|. \]  

Proof of Lemma 2

Let \( \gamma'_{OLS,N} = (\hat{\beta}_{OLS,N}', \hat{\lambda}_{OLS,N}) \). To prove Lemma 2 it suffices to show that \( p \lim_{N \to \infty} \hat{\lambda}_{OLS,N} \neq \lambda \). By partitioned matrix inversion the model in (1) implies
\[ \hat{\lambda}_{OLS,N} = \left( \left( I_N - P_{D_N} W_N Y_N \right) \left( I_N - P_{D_N} W_N Y_N \right) \right)^{-1} \times \left( W_N Y_N \right) \]
\[ = \lambda + \frac{(W_N Y_N)' (I_N - P_{D_N}) e_N}{(W_N Y_N)' (I_N - P_{D_N}) W_N Y_N}. \]  
Recalling that \( P_{D_N} e_N = e_N \) and \( P_{D_N} X_N = X_N \), the results in (4) and (A.4) imply
\[ \hat{\lambda}_{OLS,N} = \lambda + \frac{\phi_{2N}(e_N' e_N - e_N' P_{D_N} e_N)}{\phi_{2N}(e_N' e_N - e_N' P_{D_N} e_N)} = \lambda + \frac{1}{\phi_{2N}} \]  
and so via (5)
\[ p \lim_{N \to \infty} |\hat{\lambda}_{OLS,N}| = \infty \neq |\lambda|. \]  

As a point of interest we note that in light of (A.7) and (A.10) we have \( \hat{\lambda}_{2SLS,N} = \hat{\lambda}_{OLS,N} \). The equality between the 2SLS and OLS estimators for \( \lambda \) does not extend, however, to all components of the parameter vector \( \gamma \).

Proof of Theorem 1

Parts (a) and (b): Since \( \hat{Z}_N' \hat{Z}_N = \hat{Z}_N' \hat{Z}_N \), the model in (11) implies
\[ N^{1/2} (\hat{\gamma}_{2SLS,N} - \gamma) = (N^{-1} \hat{Z}_N' \hat{Z}_N)^{-1} N^{-1/2} \hat{Z}_N' e_N \]
\[ = \left( N^{-1} \sum_{i=1}^{T} \hat{Z}_i' \hat{Z}_i \right)^{-1} N^{-1/2} \sum_{i=1}^{T} \hat{Z}_i' e_i. \]  
\[ \text{(A.11)} \]
Consider the first term on the r.h.s. of (A.11). Extending the result in (4) to the panel data model in (11), and noting that \( P_{D_N} D_N = D_N \) it follows from (15) that
\[ \hat{Z}_N = (D_{N_N} c_{NN} e_N + \Pi_{NN}) \]  
\[ \text{(A.12)} \]
where
The results in (A.12) imply
\[
N^{-1} \hat{Z}_{in}^2 = \begin{bmatrix}
N^{-1}D_{in}^2 \Sigma_{in}^{-1} + c_{in}N^{-1}D_{in}^2 \epsilon_{in} + N^{-1}D_{in}^2 \Pi_{in} \\
c_{in}N^{-1} \epsilon_{in}^2 D_{in} + N^{-1} \Pi_{in}^2 D_{in} \\
\end{bmatrix}.
\]

(A.13)

Consider the elements of the matrix in (A.13). As a preliminary observation we note that applying the results (4)–(6) to the \(t\)-th setting gives
\[
c_{in} N^{-1} \epsilon_{in} \rightarrow c_{*,t} = \frac{\alpha + \mu_{*,t} \beta}{1 - \lambda} < \infty
\]
as \(N \rightarrow \infty\). Now let \(\Psi_{in} = N^{-1}D_{in}^2 \epsilon_{in}.\) Then, Assumptions 5 and 6 imply \(E(\Psi_{in}) = 0\) and \(E(\Psi_{in}^2) = \sigma^2 N^{-1}D_{in}^2 \epsilon_{in} \rightarrow 0\) as \(N \rightarrow \infty\), and so, via Chebyshev’s inequality, \(p \lim_{N \rightarrow \infty} N^{-1}D_{in}^2 \epsilon_{in} = 0\). As a consequence
\[
N^{-1}D_{in}^2 \Pi_{in} = \phi_{2n}(N^{-1}D_{in}^2 X_{in})\beta + \phi_{2n}(N^{-1}D_{in}^2 \epsilon_{in}) \overset{p}{\rightarrow} 0
\]
\[
N^{-1} \Pi_{in}^2 D_{in} = \phi_{2n}^2 \beta^2(1 - \lambda)
\]
\[
+ \phi_{2n}^2(N^{-1} \epsilon_{in}^2 D_{in}^2)(N^{-1}D_{in}^2 D_{in}^2)^{-1}(N^{-1}D_{in}^2 \epsilon_{in})
\]
\[
+ 2\phi_{2n}^2 \beta^2(N^{-1}X_{in} D_{in})(N^{-1}D_{in}^2 D_{in}^2)^{-1}(N^{-1}D_{in}^2 \epsilon_{in}) \overset{p}{\rightarrow} 0
\]
(A.14)
as \(N \rightarrow \infty\). Since \(D_{in}\) contains \(\epsilon_{in}\) we have furthermore
\[
c_{in} N^{-1} \epsilon_{in} \Pi_{in} \overset{p}{\rightarrow} 0
\]
(A.15)
as \(N \rightarrow \infty\).

The results in (A.13)–(A.15), and Assumption 6 imply that for \(t = 1, \ldots, T\)
\[
p \lim_{N \rightarrow \infty} N^{-1} \hat{Z}_{in}^2 = p \lim_{N \rightarrow \infty} \begin{bmatrix}
N^{-1}D_{in}^2 \Pi_{in} \\
c_{in}N^{-1}D_{in}^2 \epsilon_{in}
\end{bmatrix}
\]
\[
= \begin{bmatrix}
1 & c_{*,t} \\
\mu_{*,t}^2 & c_{*,t}^2
\end{bmatrix}
\]
(A.16)
and consequently
In light of Assumptions 6 and 7 this matrix is finite and nonsingular.

Now consider the second term on the r.h.s. of (A.11). First note from (A.12) that
\[
N^{-1/2} \Pi'_{IN} e_{IN} = (N^{1/2} \phi_{2N}) \beta'(N^{-1} X'_{IN} e_{IN}) \\
+ (N^{1/2} \phi_{2N}) (N^{-1} \epsilon'_{IN} D_{IN}) (N^{-1} D'_{IN} D_{IN})^{-1} (N^{-1} D'_{IN} e_{IN}).
\] (A.18)

From (3) and (4) observe that \( N^{1/2} \phi_{2N} \to 0 \). Since \( p \lim_{N \to \infty} N^{-1} D'_{IN} e_{IN} = 0 \), and utilizing Assumption 6 it follows that
\[
p \lim_{N \to \infty} N^{-1/2} \Pi'_{IN} e_{IN} = 0, \quad t = 1, \ldots, T.
\] (A.19)

Also note that
\[
p \lim_{N \to \infty} (c_{N,t} - c_*) N^{-1/2} e'_{IN} e_{IN} = 0, \quad t = 1, \ldots, T
\] (A.20)
since \( p \lim_{N \to \infty} (c_{N,t} - c_*) = 0 \) and \( N^{-1/2} e'_{IN} e_{IN} \to \mathcal{N}(0, \sigma^2_e) \) by the Lindeberg–Lévy central limit theorem.

Now define
\[
\bar{Z}_N = (\bar{Z}'_{1N}, \ldots, \bar{Z}'_{T_N})' 
\]
where
\[
\bar{Z}_{tN} = (D_{tN}, c_{t*, t} e_{tN}), \quad t = 1, \ldots, T.
\]
Then clearly
\[
\lim_{N \to \infty} N^{-1/2} \bar{Z}^t_{N} \bar{Z}_N = Q_{ZZ},
\] (A.21)
which is, as remarked above, finite and nonsingular. Assumptions 6 and 7 imply that the elements of \( Z_N \) are uniformly bounded in absolute value (observing that \( T \) is fixed). In light of Assumption 5 all conditions for the central limit theorem given in Kelejian and Prucha (1998, p. 112) are thus seen to be satisfied. Applying this theorem yields
\[
N^{-1/2} \bar{Z}'_{N} e_{N} \overset{D}{\to} \mathcal{N}(0, \sigma^2 Q_{ZZ}).
\] (A.22)

From (A.12), (A.19) and (A.20) it is readily seen that
\[
p \lim_{N \to \infty} (N^{-1/2} \bar{Z}'_{N} e_{N} - N^{-1/2} \bar{Z}'_{N} e_{N}) = 0.
\] (A.23)

Consequently we also have
\[ N^{-1/2} \gamma_{Z_N}^{'} e_N \xrightarrow{D} N(0, \sigma^2 Q_{ZZ}) \]  
(A.24)

as \( N \to \infty \). Parts (a) and (b) of Theorem 1 now follows from (A.11), (A.17), (A.24) and (12).

**Part (c):** From (11) and (17) we have
\[ N^{1/2}(\tilde{\gamma}_{OLS,N} - \gamma) = (N^{-1}Z_N^{'}Z_N)^{-1}N^{-1/2}Z_N^{'}e_N. \]  
(A.25)

Consider \( N^{-1}(Z_N^{'}Z_N) \). Note again from the extension of (4)–(6) that
\[ Z_N = (D_N, c_N, e_N + \phi_{2N}X_N^N \beta + \phi_{2N}e_{Nc}). \]

Since Assumptions 5 and 6 imply that \( \lim_{N \to \infty} \phi_{2N}N^{-1}D_N^{'}D_N = 0 \) and \( \lim_{N \to \infty} N^{-1}D_N^{'}e_{Nc} = 0 \) it should be clear that
\[ \lim_{N \to \infty} N^{-1}\tilde{Z}_N^{'}\tilde{Z}_N = \lim_{N \to \infty} N^{-1}Z_N^{'}Z_N = Q_{ZZ}. \]  
(A.26)

Therefore the first factors in (A.11) and (A.25) have the same probability limit.

Now consider the second factors in (A.11) and (A.25). Given (A.19), recalling that \( N^{1/2} \phi_{2N} \to 0 \) as \( N \to \infty \), and then noting that Assumptions 5 and 6 imply
\[ \lim_{N \to \infty} (N^{1/2} \phi_{2N})N^{-1}e_{Nc} = 0, \]
\[ \lim_{N \to \infty} (N^{1/2} \phi_{2N})N^{-1}X_N^N e_{Nc} = 0, \]  
(A.27)

it follows that
\[ \lim_{N \to \infty} (N^{-1/2}Z_N^{'}e_N - N^{-1/2}\tilde{Z}_N^{'}e_N) = 0. \]  
(A.28)

Part (c) follows from (A.26) and (A.28).

**Part (d):** Let
\[ \tilde{e}_N = y_N - Z_N \tilde{\gamma}_{2SLS,N} = e_N - Z_N \Delta_N, \]
where \( \Delta_N = \tilde{\gamma}_{2SLS,N} - \gamma. \) Then
\[ \tilde{\sigma}_N^2 = (NT)^{-1}e_N^{'}e_N - 2(NT)^{-1}e_N^{'}Z_N \Delta_N + \Delta_N^{'}(NT)^{-1}Z_N^{'}Z_N \Delta_N. \]  
(A.29)

For a given value of \( T \), Assumption 5 and Khintchine’s law of large numbers, imply
\[ \lim_{N \to \infty} (NT)^{-1}e_N^{'}e_N = \sigma^2. \]  
(A.30)

The consistency of \( \tilde{\gamma}_{2SLS,N} \) implies \( \lim_{N \to \infty} \Delta_N = 0 \). Given this, (A.24), (A.26) and (A.28)
\[ p \lim_{N \to \infty} (NT)^{-1} \varepsilon_t^\prime Z_N \Delta_N = 0, \]
\[ p \lim_{N \to \infty} (NT)^{-1} \Delta_t^\prime Z_N^\prime Z_N \Delta_N = 0. \]  
(A.31)

Part (d) then follows from (A.30) and (A.31).

References

Regional Science and Urban Economics 26, 77–104.
Papers in Regional Science 73, 135–152.