Econometric Theory, **18**, 2002, 252–277. Printed in the United States of America.
DOI: 10.1017.S0266466602182028

CONSISTENCY AND EFFICIENCY OF LEAST SQUARES ESTIMATION FOR MIXED REGRESSIVE, SPATIAL AUTOREGRESSIVE MODELS

LUNG-FEI LEE
The Ohio State University

Least squares estimation has casually been dismissed as an inconsistent estimation method for mixed regressive, spatial autoregressive models with or without spatial correlated disturbances. Although this statement is correct for a wide class of models, we show that, in economic spatial environments where each unit can be influenced aggregately by a significant portion of units in the population, least squares estimators can be consistent. Indeed, they can even be asymptotically efficient relative to some other estimators. Their computations are easier than alternative instrumental variables and maximum likelihood approaches.

1. INTRODUCTION

Spatial econometrics utilizes econometric techniques dealing with empirical economic problems caused by spatial autocorrelation in cross-sectional and/or panel data. Spatial dependence issues have recently been receiving more attention in both empirical and theoretical econometrics. These studies have focused on spatial dependence in urban, real estate, regional, public, agricultural, environmental, and industrial organizational economics (see for references, e.g., the recent survey by Anselin and Bera, 1998). There are a few textbooks that deal with spatial econometrics issues (Paelinck and Klaassen, 1979; Anselin, 1988; Cressie, 1993) and several collections of spatial econometrics papers published in special journals on regional and urban economics (Anselin, 1992; Anselin and Florax, 1995; Anselin and Rey, 1997).

To capture spatial dependence in cross-sectional sampling units, the approach in spatial econometrics is to impose structure on the specification of a

I appreciate financial support from the Research Grant Council of Hong Kong for my research (under grant HKUST595/96H to the Hong Kong University of Science and Technology). This paper was completed when I was a faculty member at the Hong Kong University of Science and Technology. Valuable comments and suggestions from two anonymous referees have substantially improved the presentation of this paper. An earlier version of this paper was presented at the Midwest Econometric Group Annual Meeting at Iowa State University in October 1999 and econometrics seminars at University of Southern California, Penn. State University, University of Illinois at Urbana, and Ohio State University. I appreciate comments from participants at the meeting and seminars. Address correspondence to: Lung-fei Lee, Department of Economics, Ohio State University, 410 Arps Hall, 1945 N. High Street, Columbus, OH 43210-1172, USA; e-mail: lflee@econ.ohio-state.edu.

 © 2002 Cambridge University Press 0266-4666/02 $9.50

https://doi.org/10.1017/S0266466602182028 Published online by Cambridge University Press

model. The spatial autoregressive models were popularized by Whittle (1954), Mead (1967), Cliff and Ord (1973), and Ord (1975). With explanatory variables in addition to a spatial autoregressive process incorporated in a linear model, the resulting model is termed a mixed regressive, spatial autoregressive model (Anselin, 1988). Whereas the spatial autoregression extends autocorrelation in a time series to spatial dimensions, the spatial aspect of a model has its own distinctive features that are not presented in the time domain. The spatial autocorrelation model of Whittle (1954) has the distinguishing feature of simultaneity in econometric equilibrium models.

Because spatial lagged variables can be correlated with disturbances in a spatial model, it has been noted in the literature (e.g., Whittle, 1954; Ord, 1975; Anselin, 1988; Kelejian and Prucha, 1998; Anselin and Bera, 1998) that the ordinary least squares (OLS) estimator (OLSE) can be inconsistent for estimating the mixed regressive, spatial autoregressive model. Alternative estimation methods such as the maximum likelihood (ML) method and the instrumental variables (IV) method have been suggested (see Anselin, 1988). Kelejian and Prucha (1998) provide an in-depth asymptotic analysis of the IV estimator (IVE). Kelejian and Prucha (1997) clarify the possible misuse of the two-stage least squares procedure for the estimation of spatial models with spatial autoregressive disturbances after a Cochrane–Orcutt transformation. From their analyses, it is clear that asymptotic properties of an IVE for spatial autoregressive models depend on the properties of the spatial weights matrices used. In particular, regularity conditions on functions of spatial weights are needed to limit the degree of correlations of spatial units. However, Kelejian and Prucha's analyses have concentrated only on IVE's.

In this paper, we investigate in detail the OLS approach in some economic spatial environments. For economic problems, economic characteristics rather than physical locations of economic units may form the proper space of interest (see, e.g., Case, Rosen, and Hines, 1993). Economic activities of a unit may have a weak influence on neighboring units. But, each unit can be influenced aggregately by a significant portion of neighboring units in the economy. Units can be compactly allocated in their characteristic economic spaces. Under these scenarios, we can show that the OLSE for the mixed regressive, spatial autoregressive model can be consistent regardless of whether or not the disturbances are spatially correlated. Indeed, it can be shown that the OLSE may even possess some relatively efficient properties.

This paper is organized as follows. In Section 2, we consider the OLS estimation of the mixed regressive, spatial autoregressive model. Regularity conditions are specified for the model. The asymptotic properties of the OLSE are compared with those of IVE and the ML estimator (MLE) in Section 3. In Section 4, we analyze possible asymptotic biases of the OLSE under some circumstances. An asymptotic expansion provides information on possible finite sample bias of the OLSE. In Section 5, mixed regressive models with spatial correlated disturbances are considered. The OLS and (feasible) generalized least

https://doi.org/10.1017/S0266466602182028 Published online by Cambridge University Press

squares (GLS) methods are studied. Section 6 briefly reports some findings on estimation of the pure spatial autoregressive process. Final conclusions are drawn in Section 7. Some useful lemmas for our analysis and proofs for the results presented in the text are found in Appendixes A and B.

2. MIXED REGRESSIVE, SPATIAL AUTOREGRESSIVE MODELS AND OLS ESTIMATION

The first-order mixed regressive, spatial autoregressive model is specified as

$$Y_n = X_n\beta + \lambda W_{n,n} Y_n + V_n, \tag{2.1}$$

where n is the total number of cross-sectional units, Y_n is an n-dimensional vector of sample observations, X_n is an $n \times k$-dimensional matrix of regressors, $W_{n,n}$ is an $n \times n$ matrix of spatial weights, λ is a scalar coefficient with $|\lambda| < 1$, and V_n is an n-dimensional vector of independent and identically distributed (i.i.d.) disturbances with zero mean and a constant variance σ^2.[1] This and related models originated from Whittle (1954), Mead (1967), Cliff and Ord (1973, 1981), and Ord (1975). The spatial weights matrix $W_{n,n}$ is usually row normalized (or row standardized; see Haining, 1990, p. 82) such that its ith row is

$$w_{i,n} = [d_{i1}, d_{i2}, \ldots, d_{in}] \Big/ \sum_{j=1}^{n} d_{ij}, \tag{2.2}$$

where d_{ij} represents a function of the spatial distance of the ith and jth units in some (characteristic) space. In geographic analysis, weights are functions of distance. In sociometrics, the weights reflect the social networks of individuals (Doreian, 1980). In economic applications, the weights may be based on "economic" distance such as income or racial composition (Case et al., 1993). As a normalization, it is understood that $d_{ii} = 0$ for all i. It is a common practice in empirical studies to have weights matrices row normalized (see, e.g., Kelejian and Robinson, 1993; Anselin and Rey, 1991). This ensures that all weights are between 0 and 1 and weighting operations can be interpreted as an average of the neighboring values. A row-standardized weights matrix, with its largest eigenvalue always 1, facilitates the interpretation of the autoregressive coefficient λ (Ord, 1975; Anselin and Bera, 1998, p. 243). In this model, the presence of regressors, X_n, plays an important role in the estimation. More explanations will be provided later.

Let $S_n = I_n - \lambda W_{n,n}$ where I_n is the $n \times n$-dimensional identity matrix. Equation (2.1) implies that $Y_n = S_n^{-1} X_n \beta + S_n^{-1} V_n$, and, therefore, $E(Y_n) = S_n^{-1} X_n \beta$ (under the maintained assumption that X_n, for all n, are constants). Let $Z_n = (X_n, W_{n,n} Y_n)$ be the regressor matrix and $\delta = (\beta', \lambda)'$ be the vector of coefficients in (2.1). The OLSE $\hat{\delta}_{LS}$ of δ is

$$\hat{\delta}_{LS} = (Z_n' Z_n)^{-1} Z_n' Y_n. \tag{2.3}$$

https://doi.org/10.1017/S0266466602182028 Published online by Cambridge University Press

It follows that $\hat{\delta}_{LS} = \delta + (Z_n'Z_n)^{-1}Z_n'V_n$. The consistency of $\hat{\delta}_{LS}$ may depend on the limiting behavior of $(1/n)(W_{n,n}Y_n)'V_n$, which is a component of $(1/n)Z_n'V_n$.[2] The expectation of $(1/n)(W_{n,n}Y_n)'V_n$ is

$$E\left(\frac{1}{n}(W_{n,n}Y_n)'V_n\right) = \frac{\sigma^2}{n}\operatorname{tr}(G_n), \tag{2.4}$$

where $G_n = W_{n,n}S_n^{-1}$. In general, $\operatorname{tr}(G_n)$ is not zero, if $\lambda \neq 0$.[3] Because of this nonzero expectation, in the event that it does not vanish as n goes to infinity, the OLSE is inconsistent (e.g., Whittle, 1954; Ord, 1975; Anselin, 1988; Kelejian and Prucha, 1998; Anselin and Bera, 1998). In this paper, we investigate cases where the OLSE can be consistent. Specifically, we consider the OLSE for the mixed regressive-autoregressive model where the spatial weights matrix is row normalized and has the property that $\sum_{j=1}^{n} d_{ij}$ diverges to infinity for each unit i. Under this scenario, $(1/n)\operatorname{tr}(G_n)$ can become smaller and converge to zero as n tends to infinity. For such a case, although the correlation in (2.4) may create finite sample bias for $\hat{\delta}_{LS}$, $\hat{\delta}_{LS}$ can still be consistent.

A rigorous justification of the asymptotic properties of $\hat{\delta}_{LS}$ assumes that model (2.1) satisfies some regularity conditions. First, we give definitions of sequences of constants and matrices to establish some useful terminologies.

DEFINITION 1. *Let* $\{b_{ni}\}$, $i = 1, \cdots, n$, *be a triangular array of constants.*

(1) $\{b_{ni}\}$ *are* at most of order $(1/h_n)$ *(denoted by* $O(1/h_n)$*)* uniformly in i *if there exists a finite constant* c *independent of* i *and* n *such that* $|b_{ni}| \leq c/h_n$ *for all* i *and* n.

(2) $\{b_{ni}\}$ *are* bounded away from zero uniformly in i at the rate of h_n *if there exists a positive sequence* $\{h_n\}$ *and a constant* $c > 0$ *independent of* i *and* n *such that* $c \leq |b_{ni}|/h_n$ *for all* i *for sufficiently large* n.

DEFINITION 2. *Let* $\{A_n\}$ *be a sequence of n-dimensional square matrices, where* $A_n = [a_{n,ij}]$.

(1) The column sums of $\{A_n\}$ are uniformly bounded *(in absolute value) if there exists a finite constant* c *that does not depend on* n *such that* $\max_{1\leq j\leq n}\sum_{i=1}^{n}|a_{n,ij}| \leq c$.

(2) The row sums of $\{A_n\}$ are uniformly bounded *(in absolute value) if there exists a finite constant* c *that does not depend on* n *such that* $\max_{1\leq i\leq n}\sum_{j=1}^{n}|a_{n,ij}| \leq c$.

The order of a sequence of constants is a familiar concept in econometrics. Uniformity is relevant for sequences of constants. Definition 1 provides a specific setting for sequences bounded away from zero. If sequences are divergent, this definition describes how fast the sequences tend to infinity. The uniform boundedness of a sequence of matrices in Definition 2 plays an important role for estimators of spatial regression models. The importance of the uniform boundedness of spatial weights and related matrices has been emphasized by Kelejian and Prucha (1998, 1999a, 1999b). It limits the correlation of ele-

https://doi.org/10.1017/S0266466602182028 Published online by Cambridge University Press

ments of Y_n across spatial units to a manageable degree. It is easily shown that if A_n and B_n are matrices uniformly bounded in row sums (resp. column sums), their product A_nB_n is also uniformly bounded in row sums (resp. column sums).

Let v_i denote the ith element of V_n. We assume the following basic structure throughout the paper.

Assumption 1. v_i's are i.i.d. with zero mean and finite fourth-order moments and $|\lambda| < 1$.

In addition to Assumption 1, model (2.1) is assumed to satisfy the following regularity conditions.

Assumption 2. The elements, d_{ij}, in $W_{n,n}$ are nonnegative constants and are uniformly bounded. The sums, $\sum_{j=1}^{n} d_{ij}$, are uniformly bounded away from zero at the rate h_n, where $\lim_{n\to\infty} h_n = \infty$.

Assumption 3. $\{W_{n,n}\}$ and $\{S_n^{-1}\}$ are uniformly bounded in both row and column sums.

Assumption 4. The regressors x_i in X_n are (row) vectors of constants and are uniformly bounded.

Assumption 5. The limit Γ of $(1/n)E(Z_n')E(Z_n)$ exists and is a nonsingular matrix.

The condition that $\sum_{j=1}^{n} d_{ij}$ is bounded away from zero uniformly in i at a certain rate h_n in Assumption 2 plays the most important role in the consistency of the OLSE. This assumption implies that $\sum_{j=1}^{n} d_{ij}$ goes to infinity and has effectively ruled out cases where each unit has only a (fixed) finite number of neighbors even when the total number of units increases to infinity. For example, it rules out the case where units correspond to counties and neighbors are defined as counties with a contiguous border. For the latter case, it is easy to construct models where the OLSE is inconsistent.

The validity of Assumption 2 requires that each unit in the limit has infinitely many neighbors. For economic applications where either the neighbors of any unit are dense in a relevant compact characteristic space or each unit is influenced by many of its neighboring units (which represents a significant portion of the total population of units), it is likely that $\sum_{j=1}^{n} d_{ij}$ will diverge and $(1/n)\sum_{j=1}^{n} d_{ij}$ will converge as n becomes large. For example, in an empirical study of state expenditures in Case et al. (1993), neighbors are defined with various economic characteristics. The most relevant weights matrix used in this study has $d_{ij} = 1/|r_i - r_j|$, where r_i is the proportion of state i's population that is of African descent. As no state in the United States has a zero proportion of African-Americans in its population, d_{ij} would be positive. As there is a distribution of the population of African-Americans across regions of U.S. states, $(1/n)\sum_{j=1}^{n} d_{ij}$ will be bounded away from zero and $\sum_{j=1}^{n} d_{ij}$ will be likely to possess the n rate of divergence in this example.[4] Other related examples are

https://doi.org/10.1017/S0266466602182028 Published online by Cambridge University Press

the studies of Case (1991, 1992). For those examples, $\sum_{j=1}^{n} d_{ij}$ may be divergent with a rate h_n slower than n. In the 1992 study, Case examines the possible spatial effect of neighbors on a farmer's adoption of new technologies, and in her 1991 study, Case looks at the household consumption of rice. In these studies, "neighbors" refer to farmers who live in the same district (in rural Java, Indonesia). This characterization of neighbors makes the $W_{n,n}$ matrix block diagonal. The only nonzero elements appear as a block of households in the same district. Suppose that there are r districts and there are m farmers in each district (for simplicity). The sample size is $n = mr$. Case assumes that in a district, each neighbor of a farmer is given equal weight. In that case, $W_{n,n} = \text{diag}(B_{m,m}^{(1)}, \ldots, B_{m,m}^{(r)})$ is a block diagonal matrix with $B_{m,m}^{(k)} = (1/(m-1))(l_m l_m' - I_m)$ for each $k = 1, \ldots, r$, where l_m is a m-dimensional column vector of ones. For this example, $h_n = (m-1)$ and $h_n/n = (m-1)/mr = O(1/r)$. If the increase of sample size, n, is partly generated by the increases of m, then h_n goes to infinity. If r is finite, h_n is divergent at the n rate; but if r goes to infinity, h_n is divergent at a rate slower than the n rate.

The reduced form equation of (2.1) is well defined only if S_n is nonsingular. The latter is guaranteed as $W_{n,n}$ is row normalized and $|\lambda| < 1$ (Horn and Johnson, 1985). To limit the correlation of sample observations across different units to a manageable degree, Kelejian and Prucha (1998) assume that the row and column sums of $W_{n,n}$ and S_n^{-1} are uniformly bounded in absolute value, which is Assumption 3. As $W_{n,n}$ is row normalized and its entries are nonnegative, the row sums of $W_{n,n}$ must be uniformly bounded. As $W_{n,n}$ is row normalized and $|\lambda| < 1$, S_n^{-1} exists and has the series expansion $S_n^{-1} = \sum_{i=0}^{\infty} \lambda^i W_{n,n}^i$ (Horn and Johnson, 1985, pp. 296–301; Kelejian and Prucha, 1998, footnote 7). With the series expansion, it is easy to check that the row sums of S_n^{-1} are uniformly bounded. For our analysis in the general case, the uniform boundedness of the column sums of $W_{n,n}$ and S_n^{-1} is also needed. For the case when $h_n = n$, Assumption 3 is its direct consequence as shown by the following lemma. Let $w_{n,ij}$ be the (i,j)th entry of $W_{n,n}$.

LEMMA 1. *Suppose that the entries $w_{n,ij}$ of a row-normalized spatial weights matrix $W_{n,n}$ are all nonnegative and are of order $O(1/n)$ uniformly in i and j. Then, the row and column sums of $W_{n,n}$ and S_n^{-1} are uniformly bounded in absolute value.*

For $W_{n,n}$ with $w_{i,n} = ((d_{i1}/\sum_{l=1}^{n} d_{il}), \ldots, (d_{in}/\sum_{l=1}^{n} d_{il}))$ and d_{il} being nonnegative for all i and l, if d_{il} are uniformly bounded for all i and l, and $\sum_{l=1}^{n} d_{il}$ are bounded away from zero uniformly in i at the rate n, then $w_{n,ij}$ will be of order $O(1/n)$ uniformly in i and j. The preceding lemma will guarantee that Assumption 3 is satisfied. If h_n is not the rate n, the uniform boundedness of column sums of $W_{n,n}$ may still be possible under some circumstances. If $W_{n,n}$ is symmetric, the uniform boundedness of column sums is an immediate consequence of $W_{n,n}$ being row normalized. The following lemma allows some asymmetry.

https://doi.org/10.1017/S0266466602182028 Published online by Cambridge University Press

LEMMA 2. *If $\sum_{j=1}^{n} d_{ij}$ are uniformly bounded away from zero at the rate h_n uniformly in i and $\sum_{i=1}^{n} d_{ij}$ are at most of order h_n uniformly in j, then $\{W_{n,n}\}$ are uniformly bounded in both row and column sums.*

Assumptions 2 and 3 are satisfied for the empirical models of Case (1991, 1992) and Case et al. (1993). The $W_{n,n}$ and S_n of Case (1991, 1992) are symmetric. The row-normalization of $W_{n,n}$ guarantees that Assumption 3 is satisfied. Assumption 2 is appropriate as there are many farmers in each district. In Case et al. (1993), as economic characteristics are measures of spatial distance and the United States is a closed region, Lemma 1 can be satisfied along with Assumptions 2 and 3.

For the OLS estimation, as $\mathrm{tr}(G_n) = \sum_{i=1}^{n} w_{i,n} S_n^{-1} e_{i,n}'$ where $e_{i,n}$ is the ith unit row vector of dimension n, whether the value in (2.4) will converge to zero or not will depend on the order of $w_{i,n} S_n^{-1} e_{i,n}'$. The following proposition summarizes orders of products of related weights vectors and matrices.

PROPOSITION 1. *Suppose that the column sums of $\{A_n\}$ are uniformly bounded in absolute value, $\sum_{l=1}^{n} d_{il}$ is bounded away from zero uniformly in i at the rate h_n, and d_{il} are nonnegative and uniformly bounded for all i and l. Then $|w_{i,n} A_n b_n'| = O(k_n/h_n)$ uniformly in i for any vector $b_n = (b_{n1}, \ldots, b_{nn})$ such that $\sum_{i=1}^{n} |b_{ni}| = O(k_n)$.*

Under Assumptions 2 and 3, an immediate consequence of Proposition 1 is that $w_{i,n} S_n^{-1} e_{i,n}' = O(1/h_n)$ uniformly in i. Therefore, the correlation of $w_{i,n} Y_n$ and v_i is of order $O(1/h_n)$ that goes to zero as $\lim_{n\to\infty} h_n = \infty$. Another useful implication is that $w_{i,n} S_n^{-1} w_{j,n}' = O(1/h_n)$ because $\sum_{k=1}^{n} w_{n,jk} = 1$.

Under Assumption 4, $\{X_n \beta\}$ are bounded. As $\{S_n^{-1}\}$ are bounded in row sums, one has $w_{i,n} S_n^{-1} X_n \beta = O(1)$ uniformly in i and n. Assumption 5 is related to an identification condition for estimation in the least squares or IV estimation frameworks. A necessary condition for Assumption 5 is that some regression coefficients (excluding the intercept) are not zero. The fifth assumption also rules out possible multicolinearities among regressors of X_n and $W_{n,n} S_n^{-1} X_n \beta$ for large n. It is shown in Appendix B that $(1/n) Z_n' Z_n = (1/n) E(Z_n') E(Z_n) + o_P(1)$. Assumption 5 guarantees that the probability limit of $(1/n) Z_n' Z_n$ exists and is nonsingular. In the IV estimation framework, Kelejian and Prucha (1998) point out the importance of the presence of relevant regressors rather than the constant term. If x_i consists of only a constant term, then $X_n = l_n$ and $W_{n,n} S_n^{-1} X_n = l_n/(1 - \lambda)$ where $l_n = (1, \ldots, 1)'$ is the constant vector of one, and hence multicolinearity occurs. This is so because $S_n^{-1} = \sum_{i=0}^{\infty} \lambda^i W_{n,n}^i$ and $W_{n,n} l_n = l_n$. For the least squares estimation, one would encounter the same issue in the limit if x_i were a constant term. The least squares method will not work well for the estimation of spatial autoregressive processes.[5] It is for this reason that we focus on the estimation of the mixed regressive model. The consistency and asymptotic distribution of $\hat{\delta}_{LS}$ are in the following theorem.

https://doi.org/10.1017/S0266466602182028 Published online by Cambridge University Press

THEOREM 1. *Under the assumed regularity conditions,*

1. $\hat{\delta}_{LS}$ *is a consistent estimator of* δ, *where* $\delta = (\beta', \lambda)$ *of model (2.1);*
2. *if* $\lim_{n\to\infty}(h_n/\sqrt{n}) = \infty$, $\sqrt{n}(\hat{\delta}_{LS} - \delta) = [(1/n)E(Z_n')E(Z_n)]^{-1}(1/\sqrt{n}) \times E(Z_n')V_n + o_P(1) \xrightarrow{D} N(0, \sigma^2\Gamma^{-1})$.

The consistency of $\hat{\delta}_{LS}$ depends crucially on $\lim_{n\to\infty} h_n = \infty$. For the cases with $\{h_n\}$ finite and bounded, $\hat{\delta}_{LS}$ is in general inconsistent as the limit of (2.4) does not vanish. Theorem 1 indicates that as long as h_n tends to infinity at a rate faster than the $\sqrt{n}$ rate, $\hat{\delta}_{LS}$ will converge in probability to δ at the usual $\sqrt{n}$ rate and its limiting distribution will be normal. It has the same limiting distribution of the OLSE of regressing Y_n on X_n and $W_{n,n}E(Y_n)$. The cases when h_n goes to infinity at a rate equal to $\sqrt{n}$ or slower are also interesting but will be investigated later.

The variance σ^2 of the disturbances can be estimated by $\hat{\sigma}^2 = (1/n)\hat{V}_n'\hat{V}_n$, where $\hat{V}_n = Y_n - Z_n\hat{\delta}_{LS}$ is the least squares residual vector. This estimator is consistent because

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{n}(V_n - Z_n(\hat{\delta}_{LS} - \delta))'(V_n - Z_n(\hat{\delta}_{LS} - \delta)) \\ &= \frac{1}{n}V_n'V_n - 2(\hat{\delta}_{LS} - \delta)'\frac{1}{n}Z_n'V_n + (\hat{\delta}_{LS} - \delta)'\frac{1}{n}Z_n'Z_n(\hat{\delta}_{LS} - \delta) \\ &= \sigma^2 + o_P(1),\end{aligned}$$

as $\hat{\delta}_{LS} - \delta = o_P(1)$, $(1/n)Z_n'V_n = o_P(1)$, and $(1/n)Z_n'Z_n = O_P(1)$ as shown in Appendix B in the proof of Theorem 1. It follows that $\hat{\sigma}^2(Z_n'Z_n)^{-1}$ provides a consistent estimate of the asymptotic variance of $\hat{\delta}_{LS}$.

3. IVE, MLE, AND OLSE

As v_i and $w_{i,n}Y_n$ in model (2.1) are correlated, alternative estimation methods instead of the method of least squares have been proposed. The IV method has been suggested by Anselin (1988) and Kelejian and Prucha (1998), among others. Lee (1999a) derives an optimal IVE. Suppose H_n is an $n \times (k+1)$ matrix of instrumental variables, the IVE of δ with H_n is

$$\hat{\delta}_{IV} = (H_n'Z_n)^{-1}H_n'Y_n. \tag{3.1}$$

The IVE $\hat{\delta}_{IV}$ was shown to be consistent and asymptotically normal (Kelejian and Prucha, 1998). Specifically, one has $\sqrt{n}(\hat{\delta}_{IV} - \delta) \xrightarrow{D} N(0, \Delta)$ where

$$\Delta = \sigma^2\left(\lim_{n\to\infty}\frac{1}{n}H_n'E(Z_n)\right)^{-1}\left(\lim_{n\to\infty}\frac{1}{n}H_n'H_n\right)\left(\lim_{n\to\infty}\frac{1}{n}E(Z_n')H_n\right)^{-1}. \tag{3.2}$$

For an optimal IVE, we note that $E(Z_n')E(Z_n) \geq E(Z_n')H_n(H_n'H_n)^{-1}H_n'E(Z_n)$ by the Cauchy–Schwarz inequality. It follows that if $H_n^* = E(Z_n) = (X_n, W_{n,n}E(Y_n))$ were used as instrumental variables, H_n^* would provide the op-

https://doi.org/10.1017/S0266466602182028 Published online by Cambridge University Press

timal IVE. The corresponding limiting variance matrix Δ^* of $\sqrt{n}(\hat{\delta}_{IV} - \delta)$ for the optimal IVE would be $\Delta^* = \sigma^2[\lim_{n\to\infty}(1/n)E(Z_n')E(Z_n)]^{-1}$. By comparing Δ^* with the limiting variance of $\hat{\delta}_{LS}$ in Theorem 1, we draw the following conclusion.

THEOREM 2. *The OLSE $\hat{\delta}_{LS}$ has the same limiting distribution of the optimal IVE of model (2.1) if* $\lim_{n\to\infty}(h_n/\sqrt{n}) = \infty$.

The construction of H_n^* for the optimal IVE is feasible, but it will require some consistent estimates of δ to begin with, because $E(Y_n) = S_n^{-1}X_n\beta$ involves the unknown parameters β and λ. The least squares estimation has the simplicity of avoiding the construction of instrumental variables, and it does not require any first-stage parameter estimates. It also bypasses the computation of the inverse of S_n for the optimal IVE. The computational simplicity of the OLSE should be appreciated, especially, when samples are large. However, the consistency and asymptotic normality of the IVE and optimal IVE do not require any condition on the behavior of h_n and they have broad applicability. The condition that $\lim_{n\to\infty}(h_n/\sqrt{n}) = \infty$ in Theorem 2 is needed only for $\hat{\delta}_{LS}$ in comparison.

When the disturbances v_i, $i = 1,\ldots,n$, are normally distributed, the ML method has been suggested by Ord (1975), among others. The log likelihood function of model (2.1) is

$$\ln L(\theta) = -\frac{n}{2}\ln(2\pi\sigma^2) + \ln|I_n - \lambda W_{n,n}| - \frac{1}{2\sigma^2}(Y_n - Z_n\delta)'(Y_n - Z_n\delta), \tag{3.3}$$

where $\theta = (\beta', \lambda, \sigma^2)'$. The asymptotic variance of the MLE $\hat{\theta}_M$ has been derived by Ord (1975) (see also Anselin and Bera, 1998, p. 256), and it is

$$\text{AsyVar}(\hat{\theta}_M) = \begin{pmatrix} \frac{X_n'X_n}{\sigma^2} & \frac{X_n'G_nX_n\beta}{\sigma^2} & 0 \\ * & \frac{(G_nX_n\beta)'(G_nX_n\beta)}{\sigma^2} + \text{tr}(G_n^2) + \text{tr}(G_n'G_n) & \frac{\text{tr}G_n}{\sigma^2} \\ * & * & \frac{n}{2\sigma^4} \end{pmatrix}^{-1}.$$

The consistency and asymptotic normality of the MLE for the mixed regressive model do not require any condition on h_n. With specific conditions on h_n, the asymptotic variance of $\hat{\delta}_{LS}$ can be compared with that of the MLE. The following corollary shows that $\hat{\delta}_{LS}$ can have the same limiting distribution as the MLE of δ.[6]

https://doi.org/10.1017/S0266466602182028 Published online by Cambridge University Press

THEOREM 3. *Under normally distributed disturbances, as* $\lim_{n\to\infty} h_n = \infty$, *the* MLE *and the optimal* IVE *of* δ *of model (2.1) have the same limiting distribution. When* $\lim_{n\to\infty}(h_n/\sqrt{n}) = \infty$, *the* OLSE $\hat{\delta}_{LS}$ *has the same limiting distribution as that of the* MLE *of* δ.

The ML approach is computationally much more expensive than the least squares or IV methods. The evaluation of $\ln L$ in (3.3) requires the determinant $|I_n - \lambda W_{n,n}|$. To ease the computational burden, Ord (1975) suggested computing the eigenvalues μ_i, $i = 1,\ldots,n$, of $W_{n,n}$. Once the eigenvalues have been derived, the log likelihood function can be rewritten as

$$\ln L(\theta) = -\frac{n}{2}\ln(2\pi\sigma^2) + \sum_{i=1}^{n}\ln|1-\lambda\mu_i| - \frac{1}{2\sigma^2}(Y_n - Z_n\delta)'(Y_n - Z_n\delta),$$

which can be evaluated at each possible θ. Ord's suggestion is extremely valuable for large samples. It reduces the computational burden in evaluating the determinant of the Jacobian transformation at each value of λ during the search for the MLE with iterative optimization routines. Ord (1975) points out that the μ_i's are real if the spatial matrix $[d_{ij}]$ of $W_{n,n}$, i.e., the spatial weights matrix before row normalization, is symmetric. For cases with a nonsymmetric matrix $[d_{ij}]$, μ_i will, in general, be complex.[7] When $W_{n,n}$ is a sparse matrix, Pace and Barry (1997) point out the benefit of sparse matrix techniques in handling storage space and algebraic operations by a computer. As the ML approach requires an iterative algorithm but the least squares does not, the computation of the least squares estimator is surely much simpler.

4. ASYMPTOTIC BIAS AND ASYMPTOTIC EXPANSION OF OLSE

The term $\hat{\delta}_{LS}$ can be asymptotically efficient for the case when h_n tends to infinity at a rate faster than the $\sqrt{n}$ rate. It remains for us to consider the cases when h_n goes to infinity at the $\sqrt{n}$ rate or a rate slower than $\sqrt{n}$. Case (1991, 1992) provides concrete examples. In Case (1992), "neighbors" refers to households in the same district, and each neighbor is given equal weight. With $n = mr$ where r is the number of districts in the sample and m is the number of households in a district, $h_n = (m-1)$. Thus, $h_n/\sqrt{n} = (m-1)/\sqrt{mr} = ((m-1)/m)(m/r)^{1/2}$. The ratio m/r will determine the rate of h_n relative to the $\sqrt{n}$ rate. If the number of households, m, in a district is relatively much larger than the number of districts, r, in a sample, one might regard h_n as divergent at a rate faster than $\sqrt{n}$. If r is relatively much larger than m, it would be reasonable to assume that h_n increases at a slower rate than $\sqrt{n}$. If the increase of sample size n corresponds to the increases of m and r at a constant proportion, h_n will increase at the $\sqrt{n}$ rate. The following theorem summarizes the asymptotic distributions for those cases. The result does not apply to the case that the number of households in any district is limited and the sample grows by surveying more districts. For the latter, m would likely be much smaller than r and $\{h_n\}$ would be a bounded sequence.

https://doi.org/10.1017/S0266466602182028 Published online by Cambridge University Press

THEOREM 4. *Suppose that Assumptions 1–5 hold and* $b_L = \sigma^2 \lim_{n\to\infty}(h_n/n)\,\mathrm{tr}(G_n)$ *exists.*

1. *If* $\lim_{n\to\infty}(h_n/\sqrt{n}) = 0$, *then* $h_n(\hat{\delta}_{LS} - \delta) \xrightarrow{P} \Gamma^{-1}(0, b_L)'$.
2. *If* $h_n = n^{1/2}$, *then* $\sqrt{n}(\hat{\delta}_{LS} - \delta) \xrightarrow{D} N(\Gamma^{-1}(0, b_L)', \sigma^2\Gamma^{-1})$.

Theorem 4 indicates that when h_n goes to infinity at a rate slower than $\sqrt{n}$, $\hat{\delta}_{LS}$ may converge to δ in probability at the slower rate h_n rather than the $\sqrt{n}$ rate. There is an asymptotic bias, and its limiting distribution is degenerate. When h_n is at the $\sqrt{n}$ rate, $\hat{\delta}_{LS}$ converges in probability to δ at the $\sqrt{n}$ rate; its variance is that of the optimal IVE and MLE, but the asymptotic bias persists.

In estimation, the existence of finite sample bias of the OLSE is expected as $E(Z_n' V_n)$ is, in general, not zero and the OLSE is also a nonlinear function of Y_n. Investigation of exact finite sampling properties of the OLSE is beyond the scope of this paper. However, possible small sample bias of the OLSE based on an asymptotic expansion can be easily derived. The asymptotic expansion of the OLSE may be useful to provide some guidance as to the use of OLSE in relation to the sample size, relevant regressors, and configuration of a spatial weights matrix. From the normal equation $Z_n'Z_n\hat{\delta}_{LS} = Z_n'Y_n$, $\hat{\delta}_{LS} - \delta = \Delta_{n1} + \Delta_{n2}$ where

$$\Delta_{n1} = \left(\frac{1}{n}E(Z_n')E(Z_n)\right)^{-1}\left(\frac{1}{n}E(Z_n')V_n + \left(0, \frac{1}{n}V_n'G_nV_n\right)'\right) \tag{4.1}$$

and

$$\Delta_{n2} = -\left(\frac{1}{n}E(Z_n')E(Z_n)\right)^{-1} \times \begin{pmatrix} 0 & \frac{1}{n}X_n'G_nV_n \\ \frac{1}{n}V_n'G_n'X_n & \frac{2}{n}(X_n\beta)'G_n'G_nV_n + \frac{1}{n}V_n'G_n'G_nV_n \end{pmatrix}(\hat{\delta}_{LS} - \delta).$$

From Theorems 1 and 4, $\hat{\delta}_{LS} - \delta$ has the stochastic order $\max\{O_P(1/\sqrt{n}), O_P(1/h_n)\}$. From the central limit theorem (CLT) in Lemma A.2 in Appendix A, $(1/n)X_n'G_nV_n$ and $(1/n)(X_n\beta)'G_n'G_nV_n$ are of order $O_P(1/\sqrt{n})$. Here $(1/n)V_n'G_n'G_nV_n = O_P(1/h_n)$ as $E((1/n)V_n'G_n'G_nV_n) = (\sigma^2/n)\mathrm{tr}(G_nG_n') = O(1/h_n)$ from Proposition 1. Therefore, $\Delta_{n2} = \max\{O_P(1/n), O_P(1/h_n^2)\}$. For the case $h_n \to \infty$ that we are considering, Δ_{n2} has a smaller stochastic order than that of Δ_{1n}, which is $\max\{O_P(1/\sqrt{n}), O_P(1/h_n)\}$. Thus, a stochastic expansion for the OLSE is

$$\hat{\delta}_{LS} - \delta = \Delta_{n1} + \max\left\{O_P\left(\frac{1}{n}\right), O_P\left(\frac{1}{h_n^2}\right)\right\}. \tag{4.2}$$

https://doi.org/10.1017/S0266466602182028 Published online by Cambridge University Press

The term $E(\Delta_{n1})$ is the finite sample bias of $\hat{\delta}_{LS}$ from the leading term of the asymptotic expansion (4.2). By the inverse of a partition matrix,

$$E(\Delta_{n1}) = [E(Z_n')E(Z_n)]^{-1}(0, E(V_n' G_n V_n))' = [E(Z_n')E(Z_n)]^{-1}(0, \sigma^2 \operatorname{tr}(G_n))'$$
$$= \sigma^2 \operatorname{tr}(G_n)(D_{12}', D_{22})',$$

where $D_{22} = \{(G_n X_n \beta)'[I_n - X_n(X_n'X_n)^{-1}X_n'](G_n X_n \beta)\}^{-1}$ and $D_{12} = -(X_n'X_n)^{-1}X_n'(G_n X_n \beta)D_{22}$. From the conformable matrix partition, $\sigma^2 \operatorname{tr}(G_n)D_{22}$ and $\sigma^2 \operatorname{tr}(G_n)D_{12}$ provide, respectively, the leading finite sample biases for the OLSE of λ and β. These biases are related to the configuration of the spatial weights matrix and relevant regressors of model (2.1). Consider the artificial regression of regressing $G_n X_n$ on X_n, i.e., $G_n X_n = X_n \gamma + \epsilon_n$. The OLSE of γ is $\hat{\gamma} = (X_n'X_n)^{-1}X_n' G_n X_n$, and the estimated residual matrix of the artificial regression is $\hat{\epsilon}_n = [I_n - X_n(X_n'X_n)^{-1}X_n']G_n X_n$. The biases of the leading term of the asymptotic expansion (4.2) can be rewritten as

$$\sigma^2 \operatorname{tr}(G_n)D_{22} = \sigma^2 \frac{\operatorname{tr}(G_n)}{\beta'\hat{\epsilon}_n'\hat{\epsilon}_n\beta} = \sigma^2\lambda \frac{\operatorname{tr}(W_{n,n}G_n)}{\beta'\hat{\epsilon}_n'\hat{\epsilon}_n\beta} \tag{4.3}$$

and

$$\sigma^2 \operatorname{tr}(G_n)D_{12} = -\hat{\gamma}\beta \cdot \sigma^2 \frac{\operatorname{tr}(G_n)}{\beta'\hat{\epsilon}_n'\hat{\epsilon}_n\beta} = -\hat{\gamma}\beta \cdot \sigma^2\lambda \frac{\operatorname{tr}(W_{n,n}G_n)}{\beta'\hat{\epsilon}_n'\hat{\epsilon}_n\beta}. \tag{4.4}$$

From (4.3) and (4.4), these finite sample biases of λ and β are proportional to σ^2 and approximately proportional to λ. The denominator in (4.3) and (4.4) could become small if $G_n X_n \beta$ could be accurately explained by X_n in an artificial regression. Such a scenario would correspond to multicolinearity of $G_n X_n \beta$ and X_n. If $G_n X_n \beta$ and X_n were highly colinear, the finite sample biases in (4.3) and (4.4) would be larger. In addition, the bias for β is also proportional to the regression coefficients, $\hat{\gamma}\beta$, of the artificial regression of $G_n X_n \beta$ on X_n. These biases are directly related to the value of $\operatorname{tr}(W_{n,n}^2)$ as $W_{n,n}^2$ is the leading term of $W_{n,n}G_n$ in its series expansion in $W_{n,n}$. Because $(1/n)\operatorname{tr}(W_{n,n}G_n) = O(1/h_n)$, these finite sample biases become smaller for larger h_n.

5. OLS AND GLS ESTIMATION FOR MODELS WITH SPATIAL AUTOREGRESSIVE DISTURBANCES

The first-order mixed regressive, spatial autoregressive model with first-order spatial autoregressive disturbances generalizes model (2.1) with its disturbances being spatially correlated such that

$$Y_n = X_n\beta + \lambda W_{n,n}Y_n + U_n, \qquad U_n = \rho W_{n,n}U_n + V_n, \tag{5.1}$$

where ρ is a scale parameter with $|\rho| < 1$ and the other variables are defined as in (2.1).[8] Let $R_n = I_n - \rho W_{n,n}$. The model implies that $U_n = R_n^{-1}V_n$ and $Y_n = S_n^{-1}X_n\beta + S_n^{-1}R_n^{-1}V_n$.

https://doi.org/10.1017/S0266466602182028 Published online by Cambridge University Press

It is of interest to investigate the asymptotic properties of the OLSE $\hat{\delta}_{LS}$ in (2.3) for this model. It follows from (5.1) that $\hat{\delta}_{LS} = \delta + (Z_n'Z_n)^{-1}Z_n'U_n$. Even though the disturbances u_{ni}, $i = 1,\ldots,n$, in U_n are spatially correlated, the consistency and asymptotic normality of $\hat{\delta}_{LS}$ can still hold. We assume that the following additional regularity conditions are satisfied for model (5.1).

Assumption 6. $\{R_n^{-1}\}$ are uniformly bounded in both row and column sums.

Assumption 7. The limit of $(1/n)E(Z_n')R_n^{-1}R_n'^{-1}E(Z_n)$ exists and is finite.

Assumption 6 is similar to Assumption 3. It applies to the spatial disturbance process instead of the mixed regression equation. The following theorem shows that even though the spatial correlation of U_n does influence the limiting variance of $\hat{\delta}_{LS}$, the correlation of $W_{n,n}Y_n$ and U_n may have no effect on the asymptotic distribution of the OLSE.

THEOREM 5. *Under the assumed regularity conditions,*

1. *if* $\lim_{n\to\infty} h_n = \infty$, $\hat{\delta}_{LS}$ *is a consistent estimator of* δ *for model (5.1);*
2. *if* $\lim_{n\to\infty}(h_n/\sqrt{n}) = \infty$, *then*

$$\sqrt{n}(\hat{\delta}_{LS} - \delta) = \left[\frac{1}{n}E(Z_n')E(Z_n)\right]^{-1}\frac{1}{\sqrt{n}}E(Z_n')U_n + o_P(1) \xrightarrow{D} N(0, \sigma^2\Omega_L), \tag{5.2}$$

where $\Omega_L = \lim_{n\to\infty}[(1/n)E(Z_n')E(Z_n)]^{-1}[(1/n)E(Z_n')R_n^{-1}R_n'^{-1}E(Z_n)] \times [(1/n)E(Z_n')E(Z_n)]^{-1}$.

From Theorem 5, we see that, for the case when $\lim_{n\to\infty}(h_n/\sqrt{n}) = \infty$, the correlation of $W_{n,n}Y_n$ and U_n does not have any effect on the limiting distribution of $\hat{\delta}_{LS}$ for model (5.1). It is of interest to note that the OLSE for the spatial model in (5.1) behaves differently than the OLSE for a corresponding time series model. It is well known that in a dynamic regression model with serially correlated disturbances, i.e., the model with $y_t = \lambda y_{t-1} + x_t\beta + u_t$ and $u_t = \rho u_{t-1} + v_t$, the OLSE's of β and λ are inconsistent when $\rho \neq 0$. This contrasts with the results in Theorem 5 for the spatial model. The difference occurs because of differences in the weights matrices. The preceding theorem highlights the important distinction of spatial models from time series models.

The (feasible) GLS method can provide more efficient estimates of β and λ than the OLS approach. By a Cochrane–Orcutt transformation, (5.1) implies that $R_nY_n = R_nZ_n\delta + V_n$. To implement the feasible GLS method, suppose that ρ of the spatial autoregressive disturbances has been consistently estimated by $\tilde{\rho}$. This can be done, e.g., by the method of moments with consistently estimated disturbances of U_n in Kelejian and Prucha (1999a) under some circumstances. Denote $\tilde{R}_n = I_n - \tilde{\rho}W_{n,n}$. The feasible GLS estimator (GLSE) is

$$\hat{\delta}_{FG} = (Z_n'\tilde{R}_n'\tilde{R}_nZ_n)^{-1}Z_n'\tilde{R}_n'\tilde{R}_nY_n. \tag{5.3}$$

https://doi.org/10.1017/S0266466602182028 Published online by Cambridge University Press

For the asymptotic analysis of $\hat{\delta}_{FG}$, we assume that the following additional regularity conditions hold.

Assumption 8. The initial estimate $\tilde{\rho}$ is a consistent estimator of ρ.

Assumption 9. $\lim_{n\to\infty}(1/n)E(Z_n)'R_n'R_nE(Z_n)$ exists and is nonsingular.

It is of interest to note that, because ρ appears linearly in the Cochrane–Orcutt transformed equation, all that will be needed of $\tilde{\rho}$ to establish the asymptotic distribution of $\hat{\delta}_{FG}$ is its consistency but not its rate of convergence.[9] The asymptotic properties of $\hat{\delta}_{FG}$ are in the following theorem.

THEOREM 6. *Under the assumed regularity conditions,*

1. *the feasible GLSE* $\hat{\delta}_{FG}$ *is a consistent estimator of* δ *for model (5.1) as* $\lim_{n\to\infty}h_n = \infty$;
2. *if* $\lim_{n\to\infty}(h_n/\sqrt{n}) = \infty$, *then* $\sqrt{n}(\hat{\delta}_{FG} - \delta) \xrightarrow{D} N(0, \sigma^2\Omega_G^{-1})$, *where* $\Omega_G = \lim_{n\to\infty}(1/n)E(Z_n')R_n'R_nE(Z_n)$.

Kelejian and Prucha (1998) suggest an IV method for the estimation of the Cochrane–Orcutt transformed equation. Their suggested IVE is consistent and asymptotically normal, but they have not derived the optimal IVE. By extending their approach, Lee (1999a) has derived the optimal IVE. By comparing the limiting variance of $\hat{\delta}_{FG}$ with those of the IVE in Kelejian and Prucha (1998) and Lee (1999a), one can conclude that $\hat{\delta}_{FG}$ has the same limiting distribution as the optimal IVE. The feasible GLSE is computationally simpler because the computation of the inverse of S_n can be bypassed completely.

Finally, σ^2 can be estimated by $\hat{\sigma}^2 = (1/n)\hat{V}_n'\hat{V}_n$ where $\hat{V}_n = \tilde{R}_n(Y_n - Z_n\hat{\delta}_{FG})$. The term $\hat{\sigma}^2$ is consistent because

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{n}(U_n - Z_n(\hat{\delta}_{FG} - \delta))'\tilde{R}_n'\tilde{R}_n(U_n - Z_n(\hat{\delta}_{FG} - \delta)) \\ &= \frac{1}{n}V_n'R_n'^{-1}\tilde{R}_n'\tilde{R}_nR_n^{-1}V_n - 2(\hat{\delta}_{FG} - \delta)'\frac{1}{n}Z_n'\tilde{R}_n'\tilde{R}_nR_n^{-1}V_n \\ &\quad + (\hat{\delta}_{FG} - \delta)'\frac{1}{n}Z_n'\tilde{R}_n'\tilde{R}_nZ_n(\hat{\delta}_{FG} - \delta) \\ &= \sigma^2 + o_P(1).\end{aligned}$$

The asymptotic variances of $\hat{\delta}_{LS}$ and $\hat{\delta}_{FG}$ can then be estimated by $\hat{\sigma}^2(Z_n'Z_n)^{-1}Z_n'\tilde{R}_n^{-1}\tilde{R}_n'^{-1}Z_n(Z_n'Z_n)^{-1}$ and $\hat{\sigma}^2(Z_n'\tilde{R}_n'^{-1}\tilde{R}_n^{-1}Z_n)^{-1}$, respectively.

6. ESTIMATION OF PURE SPATIAL AUTOREGRESSIVE PROCESSES

The preceding sections have focused on the OLS estimation of the mixed regressive, spatial autoregressive models. The importance of the presence of relevant exogenous variables is summarized in Assumption 5. One might be curious whether the OLS approach could be applicable to the estimation of purely spatial autoregressive processes. The answer to the question is negative even when

https://doi.org/10.1017/S0266466602182028 Published online by Cambridge University Press

h_n is divergent to infinity.[10] An example can illustrate the problem. A pure first-order spatial autoregressive process is

$$Y_n = \lambda W_{n,n} Y_n + V_n. \tag{6.1}$$

It differs from model (2.1) without the presence of regressors. As an example, consider the model design as in Case (1992), where "neighbors" refers to farmers living in the same district. The spatial weights matrix $W_{n,n} = \text{diag}(B^{(1)}_{m,m}, \ldots, B^{(r)}_{m,m})$ is a block diagonal symmetric matrix with $B^{(k)}_{m,m} = (1/(m-1))(l_m l_m' - I_m)$ for each $k = 1, \ldots, r$. With r districts and m farmers in each district, $n = mr$ and $h_n = m - 1$.

Consider the sampling when large sample size n corresponds to both m and r going to infinity. The OLSE of λ in (6.1) is $\hat{\lambda}_{LS} = (Y_n' W_{n,n}^2 Y_n)^{-1} (W_{n,n} Y_n)' Y_n$. In this case,

$$\begin{aligned} E(Y_n' W_{n,n}^2 Y_n) &= E(V_n' (W_{n,n} S_n^{-1})^2 V_n) = \sigma^2 \,\text{tr}[(W_{n,n} S_n^{-1})^2] \\ &= \sigma^2 r \cdot \text{tr}\{[B_{m,m}(I_m - \lambda B_{m,m})^{-1}]^2\} \\ &= r \frac{\sigma^2}{(1-\lambda)^2} \left(\frac{m}{m + \lambda - 1} \right)^2 \left[1 - 2 \frac{(1-\lambda)}{m} + \frac{(1-\lambda)^2}{m} \right] \end{aligned}$$

and

$$\begin{aligned} E(Y_n' W_{n,n} V_n) &= E(V_n' W_{n,n} S_n^{-1} V_n) = \sigma^2 \,\text{tr}(W_{n,n} S_n^{-1}) \\ &= r \frac{\sigma^2 \lambda}{(1-\lambda)} \left[1 - \frac{(1-\lambda)}{m} \right]^{-1}. \end{aligned}$$

Lemma A.1 in Appendix A shows that the variances of $(h_n/n) V_n' (W_{n,n} S_n^{-1})^2 V_n$ and $(h_n/n) V_n' W_{n,n} S_n^{-1} V_n$ have order $O(h_n/n)$. This sampling implies that $m \to \infty$ and $h_n/n = (m-1)/mr \to 0$. It follows by Chebyshev's inequality that

$$\text{plim}_{n\to\infty} \frac{h_n}{n} Y_n' W_{n,n}^2 Y_n = \frac{\sigma^2}{(1-\lambda)^2}$$

and

$$\text{plim}_{n\to\infty} \frac{h_n}{n} Y_n' W_{n,n} V_n = \frac{\sigma^2 \lambda}{(1-\lambda)}.$$

Therefore,

$$\hat{\lambda}_{LS} - \lambda = (Y_n' W_{n,n}^2 Y_n)^{-1} (W_{n,n} Y_n)' V_n \xrightarrow{p} (1-\lambda)\lambda.$$

This demonstrates the inconsistency of $\hat{\lambda}_{LS}$ if $\lambda \neq 0$. As $h_n/n = o(1)$, this example is for the case that h_n is divergent to infinity but at a rate slower than n.

For the situation when h_n diverges to infinity at the rate n, consider the example that $W_{n,n}$ has $w_{n,ij} = 1/(n-1)$ for all $i \neq j$. For this case,

$$V_n' W_{n,n} S_n^{-1} V_n = \frac{1}{(n + \lambda - 1)} V_n' \left(\frac{l_n l_n'}{1-\lambda} - I_n \right) V_n$$

https://doi.org/10.1017/S0266466602182028 Published online by Cambridge University Press

and

$$V_n'(W_{n,n}S_n^{-1})^2V_n = \frac{1}{(n+\lambda-1)^2}\,V_n'\left(\frac{n-2(1-\lambda)}{(1-\lambda)^2}\,l_nl_n' - I_n\right)V_n.$$

As v_is are i.i.d., the CLT implies that $(1/\sqrt{n})l_n'V_n \xrightarrow{D} N(0,\sigma^2)$ and, therefore, $(1/n)V_n'l_nl_n'V_n \xrightarrow{D} \sigma^2\xi$ where ξ is a χ^2 random variable with one degree of freedom. Hence,

$$\hat{\lambda}_{LS} - \lambda = (V_n'(W_{n,n}S_n^{-1})^2V_n)^{-1}(W_{n,n}S_n^{-1}V_n)'V_n \xrightarrow{p} \left(\frac{\xi}{(1-\lambda)^2}\right)^{-1}\left(\frac{\xi}{1-\lambda}-1\right)$$
$$= (1-\lambda)\,\frac{[\xi-(1-\lambda)]}{\xi},$$

and the OLSE is not consistent.

Finally, for completeness, we would like to point out briefly some of the recent findings on alternative estimators of the pure spatial autoregressive process. For the estimation of the process (6.1), proper methods are the ML method and the method of moments by Kelejian and Prucha (1999a). Kelejian and Prucha (1999a) have shown that their method of moments estimator of λ is consistent. In their setup, they also allow for spatial weights matrices that are not of the form (2.2). But for spatial weights matrices of the form (2.2), they only consider the case where $\{h_n\}$ is a bounded sequence.[11] Lee (1999b) considers situations where $\{h_n\}$ can be a bounded or divergent sequence. Both the ML and method of moments are considered. He shows that as long as h_n does not diverge or diverge at a rate slower than n, the MLE and the method of moments estimator can be consistent. However, the rate of convergence of these estimators is $\sqrt{n/h_n}$. Thus, if $\{h_n\}$ diverges, their rates of convergence will be slower than the conventional $\sqrt{n}$ rate. Both the MLE and the method of moments estimator will not, in general, be consistent if h_n diverges at the rate n.[12] For detailed analyses, please see those papers. By contrasting these results with those of the mixed regressive, spatial autoregressive model, one can see that the presence of relevant regressors plays the key role and the results are quite different.

7. CONCLUSIONS

In this paper, we consider the estimation of mixed regressive, spatial autoregressive models by the method of least squares. The OLSE can be consistent under some spatial scenarios. The situations emphasized are relevant for spatial economies in which units may have small spatial impacts on other units but each unit can be influenced aggregately by a significant portion of the total units. For those models, the OLSE can even be as relatively efficient as the MLE and optimal IVE. But the OLSE is computationally simpler than the MLE and IVE. Our results clarify the widely held beliefs that least squares estimators were inconsistent for mixed regressive, spatial autoregressive models. Although such beliefs are valid for models with sparse spatial weights matrices, they are not

https://doi.org/10.1017/S0266466602182028 Published online by Cambridge University Press

so for the cases under consideration. The OLSE is, however, inconsistent for the estimation of purely spatial autoregressive processes.

NOTES

1. This model has also been referred to as the spatial, first-order autoregressive model in the literature. We would prefer to follow the terminology that originated in Anselin (1988) to emphasize the presence of regressors in the model.

2. The other component is $(1/n)X_n'V_n$, which converges in probability to zero by some conventional law of large numbers as X_n are regressors. This component will not create problems.

3. If $\lambda = 0$, $G_n = W_{n,n}$ and $\text{tr}(G_n) = \text{tr}(W_{n,n}) = 0$ because the diagonal elements of $W_{n,n}$ are zeros.

4. The number of regions is finite but can be large enough for valid asymptotic approximation. As n tends to infinity, by regarding r as a random variable, it is possible to use the uniform law of large numbers (Andrews, 1987, 1992) to argue that $(1/n)\sum_{j=1}^{n} d_{ij}$ may converge to a positive limit function in probability uniformly in i.

5. This is so, even for a purely spatial autoregressive process without an intercept term. This point will be elaborated later.

6. We conjecture that the MLE might be asymptotically efficient. However, a rigorous proof is needed for justification. I appreciate having comments from Professor H.J. Bierens on this problem.

7. For complex μ's, the absolute value of the determinant $|I_n - \lambda W_{n,n}|$ is $\prod_{i=1}^{n}|1 - \lambda\mu_i|$ with $|1 - \lambda\mu_i| = [(1 - \lambda\text{Re}(\mu_i))^2 + (\lambda\text{Im}(\mu_i))^2]^{1/2}$, where $\text{Re}(\mu)$ and $\text{Im}(\mu)$ are the real and imaginary parts of μ, respectively. I am indebted to an anonymous referee for clarification on this point.

8. A more general specification for the disturbances is $U_n = \rho M_{n,n}U_n + V_n$, where $M_{n,n}$ can be a spatial weight matrix different from $W_{n,n}$. For our subsequent theorems to be valid, it is sufficient to assume that $M_{n,n}$ possesses similar properties as $W_{n,n}$ in Assumptions 2 and 3. In practice, it is logical to have $M_{n,n} = W_{n,n}$. We adopt the simplicity in our presentation.

9. If ρ appeared nonlinearly, an $\sqrt{n}$-consistent initial estimate would likely be needed.

10. For the case with h_n being a bounded sequence, the inconsistency of the OLSE is well recognized in the existing literature.

11. Kelejian and Prucha (1999a) do not consider h_n explicitly. However, if the spatial weights matrix had h_n diverged to infinity, their Assumption 5 would not be valid because it could be shown that $\Gamma_N = O(1/h_n)$ by Proposition A.3 in Lee (1999b).

12. The difficulty of estimating λ consistently in this situation does not mean that λ will be underidentified in (6.1). Under specific distributions for v_i's, such as the normality distribution, λ is identifiable under the criteria in Rothenberg (1971) as different numerical values of λ are associated with different probability laws.

REFERENCES

Andrews, D.W.K. (1987) Consistency in nonlinear econometric models: A generic uniform law of large numbers. *Econometrica* 55, 1465–1471.

Andrews, D.W.K. (1992) Generic uniform convergence. *Econometric Theory* 8, 241–257.

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

Anselin, L. (ed.) (1992) Space and Applied Econometrics [Special Issue]. *Regional Science and Urban Economics* 22.

Anselin, L. & A.K. Bera (1998) Spatial dependence in linear regression models with an introduction to spatial econometrics. In A. Ullah & D.E.A. Giles (eds.), *Handbook of Applied Economics Statistics*. New York: Marcel Dekker.

Anselin, L. & R. Florax (1995) *New Directions in Spatial Econometrics*. Berlin: Springer-Verlag.

Anselin, L. & S. Rey (1991) Properties of tests for spatial dependence in linear regression models. *Geographical Analysis* 23, 110–131.

https://doi.org/10.1017/S0266466602182028 Published online by Cambridge University Press

Anselin, L. & S. Rey (eds.) (1997) Spatial Econometrics [Special Issue]. *International Regional Science Review* 20.
Billingsley, P. (1995) *Probability and Measure*. 3rd ed. New York: Wiley.
Case, A.C. (1991) Spatial patterns in household demand. *Econometrica* 59, 953–965.
Case, A.C. (1992) Neighborhood influence and technological change. *Regional Science and Urban Economics* 22, 491–508.
Case, A.C., H.S. Rosen, & J.R. Hines (1993) Budget spillovers and fiscal policy interdependence: Evidence from the states. *Journal of Public Economics* 52, 285–307.
Cliff, A. & J. Ord (1973) *Spatial Autocorrelation*. London: Pion.
Cliff, A. & J. Ord (1981) *Spatial Processes, Models and Applications*. London: Pion.
Cressie, N. (1993) *Statistics for Spatial Data*. New York: Wiley.
Doreian, P. (1980) Linear models with spatially distributed data, spatial disturbances, or spatial effects. *Sociological Methods and Research* 9, 29–60.
Haining, R. (1990) *Spatial Data Analysis in the Social and Environmental Sciences*. New York: Cambridge University Press.
Horn, R. & C. Johnson (1985) *Matrix Analysis*. New York: Cambridge University Press.
Kelejian, H.H. & I.R. Prucha (1997) Estimation of spatial regression models with autoregressive errors by two-stage least squares procedures: A serious problem. *International Regional Science Review* 20 (1 and 2), 103–111.
Kelejian, H.H. & I.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.
Kelejian, H.H. & I.R. Prucha (1999a) A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economics Review* 40, 509–533.
Kelejian, H.H. & I.R. Prucha (1999b) On the Asymptotic Distribution of the Moran I Test Statistic with Applications. Manuscript, Department of Economics, University of Maryland.
Kelejian, H.H. & D. Robinson (1993) A suggested method of estimation for spatial interdependent models with autocorrelated errors, and an application to a county expenditure model. *Papers in Regional Science* 72, 297–312.
Lee, L.F. (1999a) Best Spatial Two-Stage Least Squares Estimators for a Spatial Autoregressive Model with Autoregressive Disturbances. Manuscript, Department of Economics, Hong Kong University of Science and Technology, Hong Kong.
Lee, L.F. (1999b) Asymptotic distributions of maximum likelihood estimators for spatial autoregressive models. Manuscript, Department of Economics, Hong Kong University of Science and Technology, Hong Kong.
Mead, R. (1967) A mathematical model for the estimation of interplant competition. *Biometrics* 23, 189–205.
Ord, J.K. (1975) Estimation methods for models of spatial interaction. *Journal of the American Statistical Association* 70, 120–297.
Pace, R.K. & R. Barry (1997) Sparse spatial autoregressions. *Statistics and Probability Letters* 33, 291–297.
Paelinck, J. & L. Klaassen (1979) *Spatial Econometrics*. Farnborough: Saxon House.
Rothenberg, T.J. (1971) Identification in parametric models. *Econometrica* 39, 577–591.
Whittle, P. (1954) On stationary processes in the plane. *Biometrika* 41, 434–449.

APPENDIX A: SOME USEFUL LEMMAS

In all the lemmas that follow, $V_n = (v_1, \cdots, v_n)'$ where v_i's are i.i.d. random variables with zero mean and finite variance σ^2. The first four moments of v are assumed to exist except Lemma A.2, where the existence of the first three absolute moments suffices. We

https://doi.org/10.1017/S0266466602182028 Published online by Cambridge University Press

denote $\mu_k = E(v_i^k)$ where $k = 3,4$. The term $W_{n,n}$ is a row-normalized spatial weights matrix with its (i,j)th element, $w_{n,ij} = d_{ij}/\sum_{l=1}^{n} d_{il}$, where $d_{ij} \geq 0$ for all i and j. The terms $W_{n,n}$ and S_n^{-1} satisfy Assumption 3, and R_n^{-1} satisfies Assumption 6.

LEMMA A.1. *Suppose that the column and row sums of* $\{A_n\}$ *are uniformly bounded in absolute value,* $\sum_{l=1}^{n} d_{il}$ *is bounded away from zero uniformly in* i *at the rate* h_n*, and* d_{il}*'s are uniformly bounded for all* i *and* l*. Then,* $E(V_n' W_{n,n} A_n V_n)^2 = \max\{O(n/h_n), O((n/h_n)^2)\}$ *and* $\operatorname{var}(V_n' W_{n,n} A_n V_n) = O(n/h_n)$.

Proof. Let $e_{i,n}$ denote the ith unit row vector and I_n be the identity matrix of dimension n. Because

$$E(V_n v_i^2 V_n') = \sigma^4 I_n + (\mu_4 - \sigma^4) e_{i,n}' e_{i,n} \quad \text{and}$$

$$E(V_n v_i v_j V_n') = \sigma^4 (e_{i,n}' e_{j,n} + e_{j,n}' e_{i,n}), \qquad i \neq j, \tag{A.1}$$

it follows that $E(V_n' W_{n,n} A_n V_n) = \sigma^2 \operatorname{tr}(W_{n,n} A_n)$ and

$$\begin{aligned}
E(V_n' W_{n,n} A_n V_n)^2 &= E\left(\sum_{i=1}^{n} v_i w_{i,n} A_n V_n\right)^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,n} A_n E(V_n v_i v_j V_n') A_n' w_{j,n}' \\
&= \sigma^4 \sum_{i=1}^{n}\sum_{j\neq i}^{n} (w_{i,n} A_n e_{i,n}' e_{j,n} A_n' w_{j,n}' + w_{i,n} A_n e_{j,n}' e_{i,n} A_n' w_{j,n}') \\
&\quad + \sigma^4 \sum_{i=1}^{n} w_{i,n} A_n A_n' w_{i,n}' + (\mu_4 - \sigma^4) \sum_{i=1}^{n} (w_{i,n} A_n e_{i,n}')^2 \\
&= \sigma^4 \sum_{i=1}^{n}\sum_{j=1}^{n} (w_{i,n} A_n e_{i,n}' e_{j,n} A_n' w_{j,n}' + w_{i,n} A_n e_{j,n}' e_{i,n} A_n' w_{j,n}') \\
&\quad + \sigma^4 \sum_{i=1}^{n} w_{i,n} A_n A_n' w_{i,n}' + (\mu_4 - 3\sigma^4) \sum_{i=1}^{n} (w_{i,n} A_n e_{i,n}')^2 \\
&= \sigma^4 [\operatorname{tr}^2(W_{n,n} A_n) + \operatorname{tr}((W_{n,n} A_n)^2) + \operatorname{tr}(W_{n,n} A_n A_n' W_{n,n}')] \\
&\quad + (\mu_4 - 3\sigma^4) \sum_{i=1}^{n} (w_{i,n} A_n e_{i,n}')^2,
\end{aligned}$$

where the final expression follows by rearranging terms. Because A_n is uniformly bounded in row and column sums, $A_n A_n'$ and $A_n W_{n,n} A_n$ are uniformly bounded in column sums. Proposition 1 implies that $\operatorname{tr}(W_{n,n} A_n) = \sum_{i=1}^{n} w_{i,n} A_n e_{i,n}'$, $\operatorname{tr}((W_{n,n} A_n)^2) = \sum_{i=1}^{n} w_{i,n} A_n W_{n,n} A_n e_{i,n}'$, and $\operatorname{tr}(W_{n,n} A_n A_n' W_{n,n}') = \sum_{i=1}^{n} w_{i,n} A_n A_n' w_{i,n}'$ are of order $O(n/h_n)$. On the other hand, $\sum_{i=1}^{n} (w_{i,n} A_n e_{i,n}')^2 = O(n/h_n^2)$, which is not larger than $O(n/h_n)$ regardless of whether $\{h_n\}$ is a bounded or divergent sequence. For the variance, its order follows because $\operatorname{var}(V_n' W_{n,n} A_n V_n) = E(V_n' W_{n,n} A_n V_n)^2 - \sigma^4 \operatorname{tr}^2(W_{n,n} A_n)$. ■

LEMMA A.2. *Suppose that* A_n *is a square matrix with its column sums being uniformly bounded and elements of the* $n \times k$ *matrix* C_n *are uniformly bounded. Then,* $(1/\sqrt{n}) C_n' A_n V_n = O(1)$. *Furthermore, if the limit of* $(1/n) C_n' A_n A_n' C_n$ *exists and it is positive definite, then* $(1/\sqrt{n}) C_n' A_n V_n \xrightarrow{D} N(0, \sigma^2 \lim_{n\to\infty} (1/n) C_n' A_n A_n' C_n)$.

https://doi.org/10.1017/S0266466602182028 Published online by Cambridge University Press

Proof. Let $a_{n,j}$ denote the jth column of A_n. It follows that $(1/\sqrt{n})C_n'A_nV_n = (1/\sqrt{n})\sum_{j=1}^n q_{nj}v_j$ where $q_{nj} = C_n'a_{n,j}$. The first result follows from Chebyshev's inequality because q_{nj} are uniformly bounded and $\mathrm{var}((1/\sqrt{n})C_n'A_nV_n) = (\sigma^2/n)\sum_{j=1}^n q_{nj}q_{nj}'$. The second result follows from the Liapounov double array CLT and the Cramér–Wold device (Billingsley, 1995, Theorem 27.3 and Theorem 29.4). To check the Liapounov condition, let α be a nonzero row vector of constants and $B_n = \mathrm{var}(\alpha C_n'A_nV_n) = \sigma^2\alpha C_n'A_nA_n'C_n\alpha'$. The assumptions imply that $\lim_{n\to\infty}(1/n)B_n^2 > 0$ and there exists a constant c such that $|\alpha q_{nj}| < c$, for all n and j. Hence, the Liapounov condition $\sum_{j=1}^n (1/B_n^3)E(|\alpha q_{nj}v_j|^3) \le c^3E|v^3|/((1/n)B_n^2)^{3/2}n^{1/2} \to 0$ holds. ■

LEMMA A.3. *Suppose that the column and row sums of* $\{A_n\}$ *are uniformly bounded in absolute value,* $\sum_{l=1}^n d_{il}$ *is bounded away from zero uniformly in i at the rate* h_n*, and* d_{il}*'s are uniformly bounded for all i and l. Then, for the first-order spatial autoregressive process* $U_n = \rho W_{n,n}U_n + V_n$*,* $E(U_n'W_{n,n}A_nU_n)^2$ *and* $E(U_n'W_{n,n}A_nV_n)^2$ *have the order* $\max\{O(n^2/h_n^2), \min[O(n^3/h_n^4), O(n^2/h_n)]\}$.

Proof. As $U_n = R_n^{-1}V_n$,

$$E(U_n'W_{n,n}A_nU_n)^2 = \sum_{i=1}^n\sum_{j=1}^n w_{i,n}B_nE(V_nu_{ni}u_{nj}V_n')B_n'w_{j,n}', \tag{A.2}$$

where $B_n = A_nR_n^{-1}$. Let $p_{i,n} = \rho w_{i,n}R_n^{-1}$ and $p_{n,il}$ be the lth element of $p_{i,n}$. As $u_{ni} = \rho w_{i,n}R_n^{-1}V_n + v_i = \sum_{l=1}^n p_{n,il}v_l + v_i$, by substituting the expressions in (A.1) and rearranging terms, one has

$$\begin{aligned}
E(V_nu_{ni}u_{nj}V_n') &= E\left[V_n\left(\sum_{l=1}^n p_{n,il}v_l + v_i\right)\left(\sum_{k=1}^n p_{n,jk}v_k + v_j\right)V_n'\right] \\
&= \sum_{l=1}^n\sum_{k=1}^n p_{n,il}p_{n,jk}E(V_nv_lv_kV_n') + \sum_{l=1}^n p_{n,jl}E(V_nv_lv_iV_n') \\
&\quad + \sum_{l=1}^n p_{n,il}E(V_nv_lv_jV_n') + E(V_nv_iv_jV_n') \\
&= \sigma^4\sum_{l=1}^n\sum_{k=1}^n p_{n,il}p_{n,jk}(e_{l,n}'e_{k,n} + e_{k,n}'e_{l,n}) \\
&\quad + \sum_{l=1}^n p_{n,il}p_{n,jl}[\sigma^4I_n + (\mu_4 - 3\sigma^4)e_{l,n}'e_{l,n}] \\
&\quad + \sigma^4\sum_{l=1}^n p_{n,jl}(e_{l,n}'e_{i,n} + e_{i,n}'e_{l,n}) + p_{n,ji}[\sigma^4I_n + (\mu_4 - 3\sigma^4)e_{i,n}'e_{i,n}] \\
&\quad + \sigma^4\sum_{l=1}^n p_{n,il}(e_{l,n}'e_{j,n} + e_{j,n}'e_{l,n}) + p_{n,ij}[\sigma^4I_n + (\mu_4 - 3\sigma^4)e_{j,n}'e_{j,n}] \\
&\quad + E(V_nv_iv_jV_n) \\
&= \sigma^4(p_{i,n}'p_{j,n} + p_{j,n}'p_{i,n} + p_{j,n}'e_{i,n} + e_{i,n}'p_{j,n} + p_{i,n}'e_{j,n} + e_{j,n}'p_{i,n}) \\
&\quad + \sigma^4(p_{i,n}p_{j,n}' + p_{n,ji} + p_{n,ij})I_n \\
&\quad + (\mu_4 - 3\sigma^4)\left(\sum_{l=1}^n p_{n,il}p_{n,jl}e_{l,n}'e_{l,n} + p_{n,ji}e_{i,n}'e_{i,n} + p_{n,ij}e_{j,n}'e_{j,n}\right) \\
&\quad + E(V_nv_iv_jV_n).
\end{aligned} \tag{A.3}$$

https://doi.org/10.1017/S0266466602182028 Published online by Cambridge University Press

Substituting both (A.1) and (A.3) into (A.2) and using the explicit expressions $p_{n,ij} = \rho w_{i,n} R_n^{-1} e'_{j,n}$ and $p_{i,n}$, it follows that

$$
\begin{aligned}
&E(U_n' W_{n,n} A_n U_n)^2 \\
&\quad = \sum_{i=1}^{n} \sum_{j=1}^{n} \{\sigma^4 \rho^2 w_{i,n} B_n R_n'^{-1} (w'_{i,n} w_{j,n} + w'_{j,n} w_{i,n}) R_n^{-1} B_n' w'_{j,n} \\
&\qquad + \sigma^4 \rho w_{i,n} B_n R_n'^{-1} (w'_{j,n} e_{i,n} + w'_{i,n} e_{j,n}) B_n' w'_{j,n} \\
&\qquad + \sigma^4 \rho w_{i,n} B_n (e'_{i,n} w_{j,n} + e'_{j,n} w_{i,n}) R_n^{-1} B_n' w'_{j,n} \\
&\qquad + \sigma^4 \rho (\rho w_{i,n} R_n^{-1} R_n'^{-1} w'_{j,n} + w_{j,n} R_n^{-1} e'_{i,n} + w_{i,n} R_n^{-1} e'_{j,n}) w_{i,n} B_n B_n' w'_{j,n} \\
&\qquad + (\mu_4 - 3\sigma^4) \\
&\qquad \times [w_{i,n} B_n P_{in} P_{jn} B_n' w'_{j,n} + \rho (w_{j,n} R_n^{-1} e'_{i,n}) w_{i,n} B_n e'_{i,n} e_{i,n} B_n' w'_{j,n} \\
&\qquad\quad + \rho (w_{i,n} R_n^{-1} e'_{j,n}) w_{i,n} B_n e'_{j,n} e_{j,n} B_n' w'_{j,n}] \\
&\qquad + \sigma^4 w_{i,n} B_n (e'_{i,n} e_{j,n} + e'_{j,n} e_{i,n}) B_n' w'_{j,n}\} \\
&\quad + \sum_{i=1}^{n} w_{i,n} B_n (\sigma^4 I_n + (\mu_4 - 3\sigma^4) e'_{i,n} e_{i,n}) B_n' w'_{i,n}, \qquad \textbf{(A.4)}
\end{aligned}
$$

where P_{in} is an $n \times n$ diagonal matrix with diagonal elements $p_{n,il}$ for $l = 1, \ldots, n$. The results in Proposition 1 can be directly applied to almost all terms of (A.4) except one. The new term is

$$
\sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,n} B_n P_{in} P_{jn} B_n' w'_{j,n} = \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{l=1}^{n} (w_{i,n} B_n e'_{l,n}) p_{n,il} p_{n,jl} (w_{j,n} B_n e'_{l,n}) = O\left(\frac{n^3}{h_n^4}\right),
$$

because $w_{i,n} B_n e'_{l,n}$ and $p_{n,il} = \rho w_{i,n} R_n^{-1} e'_{l,n}$ are of order $O(1/h_n)$ uniformly for all i and l from Proposition 1. The $\sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,n} B_n P_{in} P_{jn} B_n' w'_{j,n}$ also has the order $O(n^2/h_n)$ because $\{P_{in}\}$ are uniformly bounded in row and column sums uniformly in i. The remaining terms of (A.4) are either of orders $O(n^2/h_n^2)$ or $O(n^2/h_n^3)$, which are not larger than $O(n^2/h_n^2)$ regardless of whether $\{h_n\}$ is a bounded or divergent sequence. Similarly, the second result holds. ■

LEMMA A.4. *Suppose that the column and row sums of* $\{A_n\}$ *are uniformly bounded in absolute value,* $\sum_{l=1}^{n} d_{il}$ *are bounded away from zero uniformly in* i *at the rate* h_n*, and* d_{il} *are uniformly bounded for all* i *and* l*. Then,* $E((W_{n,n} A_n V_n)' W_{n,n} V_n)^2 = \max\{O(n^2/h_n^2), \min[O(n^3/h_n^4), O(n^2/h_n)]\}$.

Proof. As $E(V_n (w_{i,n} V_n)(w_{j,n} V_n) V_n') = \sum_{l=1}^{n} \sum_{k=1}^{n} w_{n,il} w_{n,jk} E(V_n v_l v_k V_n')$, by substituting the expressions in (A.1) and rearranging terms, one has

$$
\begin{aligned}
&E(V_n (w_{i,n} V_n)(w_{j,n} V_n) V_n') \\
&\quad = \sum_{l=1}^{n} w_{n,il} w_{n,jl} (\sigma^4 I_n + (\mu_4 - 3\sigma^4) e'_{l,n} e_{l,n}) \\
&\qquad + \sigma^4 \sum_{l=1}^{n} \sum_{k=1}^{n} w_{n,il} w_{n,jk} (e'_{l,n} e_{k,n} + e'_{k,n} e_{l,n}) \\
&\quad = \sigma^4 w_{i,n} w'_{j,n} I_n + (\mu_4 - 3\sigma^4) Q_{in} Q_{jn} + \sigma^4 (w'_{i,n} w_{j,n} + w'_{j,n} w_{i,n}),
\end{aligned}
$$

https://doi.org/10.1017/S0266466602182028 Published online by Cambridge University Press

where Q_{in} is a diagonal matrix with $w_{n,il}$ as its lth diagonal element. Therefore,

$$
\begin{aligned}
E((W_{n,n}A_nV_n)'W_{n,n}V_n)^2 &= E\left(\sum_{i=1}^{n} w_{i,n}A_nV_n\cdot w_{i,n}V_n\right)^2 \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,n}A_nE(V_n(w_{i,n}V_n)(w_{j,n}V_n)V_n')A_n'w_{j,n}' \\
&= \sigma^4\sum_{i=1}^{n}\sum_{j=1}^{n}(w_{i,n}w_{j,n}')w_{i,n}A_nA_n'w_{j,n}' \\
&\quad + (\mu_4-3\sigma^4)\sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,n}A_nQ_{in}Q_{jn}A_n'w_{j,n}' \\
&\quad + \sigma^4\sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,n}A_n(w_{i,n}'w_{j,n}+w_{j,n}'w_{i,n})A_n'w_{j,n}' \\
&= \max\left\{O\left(\frac{n^2}{h_n^2}\right), \min\left[O\left(\frac{n^3}{h_n^4}\right), O\left(\frac{n^2}{h_n}\right)\right]\right\},
\end{aligned}
$$

by Proposition 1 because $\sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,n}A_nQ_{in}Q_{jn}A_n'w_{j,n}' = O(n^2/h_n)$ but also

$$
\sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,n}A_nQ_{in}Q_{jn}A_n'w_{j,n}' = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n}(w_{i,n}A_ne_{l,n}')w_{n,il}w_{n,jl}(e_{l,n}A_n'w_{j,n}') = O\left(\frac{n^3}{h_n^4}\right),
$$

and the remaining terms are of order $O(n^2/h_n^3)$ or $O(n^2/h_n^2)$, which are not larger than $O(n^2/h_n^2)$. ■

APPENDIX B: PROOFS

Proof of Lemma 1. Given that the entries of $W_{n,n}$ are bounded by c/n for some constant c, the column sums and also the row sums of $\{W_{n,n}\}$ must be uniformly bounded in absolute value. Consider the entries of $W_{n,n}^m$ for $m \geq 2$. The (i,j)th entry of $W_{n,n}^m$ is $w_{i,n}W_{n,n}^{m-2}w_{n,j}$ where $w_{i,n}$ and $w_{n,j}$ are, respectively, the ith row and jth column of $W_{n,n}$. Because entries of the weights matrix are all nonnegative and are of order $O(1/n)$, $w_{i,n}W_{n,n}^{m-2}w_{n,j} \leq (c/n)w_{i,n}W_{n,n}^{m-2}l_n'$ for some constant c. Because the row sums of $W_{n,n}$ are unity, $w_{i,n}l_n' = 1$ and $W_{n,n}l_n' = l_n'$. Therefore, $w_{i,n}W_{n,n}^{m-2}l_n' = 1$ for all $m \geq 2$. It follows that all the entries of $W_{n,n}^m$ for $m = 1,2,\ldots,$ are bounded by c/n. The results for S_n^{-1} follow from the matrix expansion $S_n^{-1} = \sum_{i=0}^{\infty}\lambda^iW_{n,n}^i$ as all entries of $\sum_{i=0}^{\infty}\lambda^iW_{n,n}^i$ are of order $O(1/n)$ uniformly. ■

Proof of Lemma 2. It is obvious that $\{W_{n,n}\}$ is uniformly bounded in row sums. For column sums, let c_1 and c_2 be constants such that $c_1h_n \leq \sum_{j=1}^{n} d_{ij}$ and $\sum_{i=1}^{n} d_{ij} \leq c_2h_n$. It follows that

$$
\max_{1\leq j\leq n}\sum_{i=1}^{n}\frac{d_{ij}}{\sum_{l=1}^{n} d_{il}} \leq \max_{1\leq j\leq n}\frac{1}{c_1h_n}\sum_{i=1}^{n} d_{ij} \leq \frac{c_2h_n}{c_1h_n} = \frac{c_2}{c_1}.
$$

■

https://doi.org/10.1017/S0266466602182028 Published online by Cambridge University Press

Proof of Proposition 1. Let c_1, c_2, and c_3 be constants such that $d_{il} \leq c_1$ for all i and l, $0 < c_2 \leq (1/h_n)\sum_{l=1}^{n} d_{il}$ for all i and n, and $\max_{1\leq j\leq n}\sum_{k=1}^{n}|a_{kj}| \leq c_3$ for all n. The result holds because

$$|w_{i,n}A_n b_n'| = \left|\left(\frac{d_{i1}}{\sum_{l=1}^{n} d_{il}},\ldots,\frac{d_{in}}{\sum_{l=1}^{n} d_{il}}\right)A_n b_n'\right| = \left|\left(\frac{\sum_{k=1}^{n} d_{ik}a_{k1}}{\sum_{l=1}^{n} d_{il}},\cdots,\frac{\sum_{k=1}^{n} d_{ik}a_{kn}}{\sum_{l=1}^{n} d_{il}}\right)b_n'\right|$$

$$\leq \frac{c_1 c_3}{c_2 h_n}\sum_{i=1}^{n}|b_{ni}| = \frac{c_1 c_3}{c_2 h_n}O(k_n). \quad \blacksquare$$

Proof of Theorem 1. First, it will be shown that $(1/n)Z_n'Z_n = (1/n)E(Z_n')E(Z_n) + o_P(1)$. Model (2.1) implies that $(1/n)X_n'W_{n,n}Y_n = (1/n)X_n'W_{n,n}E(Y_n) + L_n$, where $L_n = (1/n)X_n'W_{n,n}S_n^{-1}V_n$, and

$$\frac{1}{n}(W_{n,n}Y_n)'(W_{n,n}Y_n) = \frac{1}{n}(W_{n,n}E(Y_n))'(W_{n,n}E(Y_n)) + 2B_{1n} + B_{2n},$$

where $B_{1n} = (1/n)(W_{n,n}E(Y_n))'W_{n,n}S_n^{-1}V_n$ and $B_{2n} = (1/n)\mathrm{tr}(W_{n,n}S_n^{-1}V_nV_n'S_n'^{-1}W_{n,n}')$. Lemma A.2 implies that L_n and B_{1n} are $o_P(1)$. The term B_{2n} is nonnegative and $E(B_{2n}) = \sigma^2(1/n)\mathrm{tr}(W_{n,n}S_n^{-1}S_n'^{-1}W_{n,n}') = O(1/h_n)$ by Proposition 1, which implies $B_{2n} = o_P(1)$ by the Markov inequality as $\lim_{n\to\infty}h_n = \infty$. Therefore, $(1/n)Z_n'Z_n = (1/n)E(Z_n')E(Z_n) + o_P(1)$. For the consistency of $\hat{\delta}_{LS}$, it remains to show that $(1/n)Z_n'V_n = o_P(1)$. By Lemma A.2, $(1/n)X_n'V_n = o_P(1)$ and $(1/n)(W_{n,n}Y_n)'V_n = (1/n)(W_{n,n}E(Y_n))'V_n + (1/\sqrt{n})C_n = (1/\sqrt{n})C_n + o_P(1)$, where $C_n = (1/\sqrt{n})\times V_n'W_{n,n}S_n^{-1}V_n$. By Lemma A.1, $E(C_n^2) = (1/n)E(V_n'W_{n,n}S_n^{-1}V_n)^2 = \max\{O(1/h_n), O(n/h_n^2)\}$, which implies that $(1/\sqrt{n})C_n = o_P(1)$. Hence $(1/n)Z_n'V_n = o_P(1)$ and the consistency of $\hat{\delta}_{LS}$ follows from the Slutsky lemma.

For asymptotic distribution of $\hat{\delta}_{LS}$, it remains for us to consider the limiting distribution of $(1/\sqrt{n})Z_n'V_n$. Under the assumption that $h_n/\sqrt{n}$ tends to infinity, $C_n = o_P(1)$ by Chebyshev's inequality. Therefore, $(1/\sqrt{n})(W_{n,n}Y_n)'V_n = (1/\sqrt{n})(W_{n,n}E(Y_n))'V_n + C_n = (1/\sqrt{n})(W_{n,n}E(Y_n))'V_n + o_P(1)$ and $(1/\sqrt{n})Z_n'V_n = (1/\sqrt{n})E(Z_n')V_n + o_P(1)$. These results imply that $\sqrt{n}(\hat{\delta}_{LS} - \delta) = [(1/n)E(Z_n')E(Z_n) + o_P(1)]^{-1}((1/\sqrt{n})E(Z_n')V_n + o_P(1))$. The limiting distribution of $\hat{\delta}_{LS}$ follows from Lemma A.2 and the Slutsky lemma. $\blacksquare$

Proof of Theorem 2. The result follows because $\Delta^* = \Gamma^{-1}$. $\blacksquare$

Proof of Theorem 3. From Proposition 1, $\mathrm{tr}(G_n) = \sum_{i=1}^{n} w_{i,n}S_n^{-1}e_{i,n}'$, $\mathrm{tr}(G_n^2) = \sum_{i=1}^{n} w_{i,n}S_n^{-1}G_n e_{i,n}'$, and $\mathrm{tr}(G_n'G_n) = \sum_{i=1}^{n} w_{i,n}S_n^{-1}S_n'^{-1}w_{i,n}'$, are of order $O(n/h_n)$. Hence, $\mathrm{tr}(G_n)/n$, $\mathrm{tr}(G_n^2)/n$ and $\mathrm{tr}(G_n'G_n)/n$ are $O(1/h_n)$, which vanish as $h_n \to \infty$. $\blacksquare$

Proof of Theorem 4. The OLSE implies that $h_n(\hat{\delta}_{LS} - \delta) = ((1/n)Z_n'Z_n)^{-1}(h_n/n)Z_n'V_n$. As the proof of Theorem 1 has shown that, as $h_n \to \infty$, $(1/n)\sum_{i=1}^{n} Z_n'Z_n = (1/n)E(Z_n')E(Z_n) + o_P(1)$, it suffices for us to consider the limiting distribution of $(h_n/n)Z_n'V_n$. Let $Q_n = V_n'G_nV_n - \sigma^2\mathrm{tr}(G_n)$. One has

$$\frac{h_n}{n}Z_n'V_n = \frac{h_n}{n}E(Z_n')V_n + \left(0, \frac{h_n}{n}V_n'G_nV_n\right)' = \frac{h_n}{n}E(Z_n')V_n + \left(0, \frac{h_n}{n}Q_n\right)'$$

$$+ \left(0, \sigma^2\frac{h_n}{n}\mathrm{tr}(G_n)\right)'.$$

https://doi.org/10.1017/S0266466602182028 Published online by Cambridge University Press

From Proposition 1, $(h_n/n)\mathrm{tr}(G_n) = O(1)$. On the other hand, $E(Q_n) = 0$ and $\mathrm{var}((h_n/n)Q_n) = (h_n/n)^2\mathrm{var}(Q_n) = O(h_n/n)$ by Lemma A.1. As long as (h_n/n) goes to zero, $(h_n/n)Q_n = o_P(1)$ by Chebyshev's inequality and, hence,

$$\frac{h_n}{n} Z_n' V_n = \frac{h_n}{n} E(Z_n')V_n + \left(0, \sigma^2 \frac{h_n}{n} \mathrm{tr}(G_n)\right)' + o_P(1).$$

When $\lim_{n\to\infty}(h_n/\sqrt{n}) = 0$, $(h_n/n)X_n'V_n = (h_n/\sqrt{n})\cdot(1/\sqrt{n})X_n'V_n = o_P(1)$ and, similarly, $(h_n/n)(W_{n,n}E(Y_n))'V_n = o_P(1)$. Thus, for case 1, $(h_n/n)E(Z_n')V_n = o_P(1)$ and $h_n(\hat{\delta}_{LS} - \delta) \xrightarrow{p} \Gamma^{-1}(0, b_L)'$.

For the case that $h_n = \sqrt{n}$, $(h_n/n)E(Z_n')V_n = (1/\sqrt{n})E(Z_n')V_n$ and the result follows from Lemma A.2. ■

Proof of Theorem 5. First of all, it can be shown that $(1/n)Z_n'Z_n = (1/n)E(Z_n')E(Z_n) + o_P(1)$. From (5.1), $(1/n)X_n'(W_{n,n}Y_n) = (1/n)X_n'(W_{n,n}E(Y_n)) + L_n$, where $L_n = (1/n)X_n'W_{n,n}S_n^{-1}U_n = (1/n)X_n'W_{n,n}S_n^{-1}R_n^{-1}V_n$, and

$$\frac{1}{n}(W_{n,n}Y_n)'(W_{n,n}Y_n) = \frac{1}{n}(W_{n,n}E(Y_n))'(W_{n,n}E(Y_n)) + 2B_{1n} + B_{2n},$$

where $B_{1n} = (1/n)(W_{n,n}E(Y_n))'(W_{n,n}S_n^{-1}U_n)$ and $B_{2n} = (1/n)\mathrm{tr}(W_{n,n}S_n^{-1}U_n \times U_n'S_n'^{-1}W_{n,n}')$. Lemma A.2 implies that $L_n = o_P(1)$ and $B_{1n} = (1/n)E(Y_n') \times W_{n,n}'W_{n,n}S_n^{-1}R_n^{-1}V_n = o_P(1)$. As B_{2n} is nonnegative and $E(B_{2n}) = (\sigma^2/n)\mathrm{tr}(W_{n,n}'S_n^{-1} \times R_n^{-1}R_n'^{-1}S_n'^{-1}W_{n,n}') = O(1/h_n)$, $B_{2n} = o_P(1)$ as $\lim_{n\to\infty}h_n = \infty$. Therefore, $\mathrm{plim}_{n\to\infty} \times (1/n)Z_n'Z_n = \lim_{n\to\infty}(1/n)E(Z_n')E(Z_n)$.

Consider the term $Z_n'U_n = E(Z_n')U_n + (0, W_{n,n}S_n^{-1}U_n)'U_n$. Here $(1/n)E(Z_n')U_n = (1/n)E(Z_n')R_n^{-1}V_n = o_P(1)$ by Lemma A.2. Lemma A.3 implies that $E((1/n)U_n'W_{n,n}S_n^{-1}U_n)^2 = O(1/h_n)$. Thus, as $\lim_{n\to\infty}h_n = \infty$, $(1/n)U_n'W_{n,n}S_n^{-1}U_n = o_P(1)$. Consequently, $(1/n)Z_n'U_n = o_P(1)$ and $\hat{\delta}_{LS}$ is a consistent estimate of δ. For asymptotic distribution, consider $(1/\sqrt{n})Z_n'U_n = (1/\sqrt{n})E(Z_n')U_n + (0, (1/\sqrt{n})U_n'W_{n,n}S_n^{-1}U_n)'$. As $E((1/\sqrt{n})U_n'W_{n,n}S_n^{-1}U_n)^2 = \max\{O(n/h_n^2), O(n^2/h_n^4)\}$ by Lemma A.3, $(1/\sqrt{n})U_n'W_{n,n}S_n^{-1}U_n = o_P(1)$ when $\lim_{n\to\infty}(n/h_n^2) = 0$. These imply that

$$\begin{aligned}\sqrt{n}(\hat{\delta}_{LS} - \delta) &= \left[\frac{1}{n}Z_n'Z_n\right]^{-1}\frac{1}{\sqrt{n}}Z_n'U_n \\ &= \left[\frac{1}{n}E(Z_n')E(Z_n) + o_P(1)\right]^{-1}\frac{1}{\sqrt{n}}E(Z_n')R_n^{-1}V_n + o_P(1).\end{aligned}$$

The final result follows from Lemma A.2. ■

Proof of Theorem 6. First, we want to show that $(1/n)(\tilde{R}_nZ_n)'(\tilde{R}_nZ_n) = (1/n)(R_nE(Z_n))'(R_nE(Z_n)) + o_P(1)$. By expanding relevant terms at ρ,

$$\begin{aligned}\frac{1}{n}(\tilde{R}_nZ_n)'(\tilde{R}_nZ_n) = {} & \frac{1}{n}(R_nZ_n)'(R_nZ_n) - (\tilde{\rho} - \rho)\frac{1}{n}(W_{n,n}Z_n)'(R_nZ_n) \\ & - (\tilde{\rho} - \rho)\frac{1}{n}(R_nZ_n)'(W_{n,n}Z_n) \\ & + (\tilde{\rho} - \rho)^2\frac{1}{n}(W_{n,n}Z_n)'(W_{n,n}Z_n).\end{aligned}$$

https://doi.org/10.1017/S0266466602182028 Published online by Cambridge University Press

Therefore, it is sufficient to show that $(1/n)(R_nZ_n)'(R_nZ_n) = (1/n)(R_nE(Z_n))' \times (R_nE(Z_n)) + o_P(1)$ and the terms $(1/n)(W_{n,n}Z_n)'(R_nZ_n)$ and $(1/n)(R_nZ_n)'(W_{n,n}Z_n)$ are $O_P(1)$. Let $T_n = S_n^{-1}R_n^{-1}$. It follows that

$$\frac{1}{n}(R_nX_n)'(R_nW_{n,n}Y_n) = \frac{1}{n}(R_nX_n)'(R_nW_{n,n}E(Y_n)) + L_n,$$

where $L_n = (1/n)(R_nX_n)'R_nW_{n,n}T_nV_n$, and

$$\frac{1}{n}(R_nW_{n,n}Y_n)'(R_nW_{n,n}Y_n) = \frac{1}{n}(R_nW_{n,n}E(Y_n))'(R_nW_{n,n}E(Y_n)) + 2C_{1n} + C_{2n},$$

where $C_{1n} = (1/n)(R_nW_{n,n}E(Y_n))'(R_nW_{n,n}T_nV_n)$ and $C_{2n} = (1/n)\mathrm{tr}(R_nW_{n,n}T_nV_n \times V_n'T_n'W_{n,n}'R_n')$. Lemma A.2 implies that L_n and C_{1n} are $o_P(1)$. By Proposition 1, $E(C_{2n}) = (\sigma^2/n)\mathrm{tr}(W_{n,n}T_nT_n'W_{n,n}'R_n'R_n) = O(1/h_n)$, which implies that $C_{2n} = o_P(1)$ by the Markov inequality as $\lim_{n\to\infty} h_n = \infty$. Hence, $(1/n)(R_nZ_n)'(R_nZ_n) = (1/n) \times (R_nE(Z_n))'(R_nE(Z_n)) + o_P(1)$. With similar analyses, $(1/n)(W_{n,n}Z_n)'(R_nZ_n) = (1/n)(W_{n,n}E(Z_n))'(R_nE(Z_n)) + o_P(1) = O_P(1)$ and $(1/n)(W_{n,n}Z_n)'(W_{n,n}Z_n) = (1/n)(W_{n,n}E(Z_n))'(W_{n,n}E(Z_n)) + o_P(1) = O_P(1)$. Because $\tilde{\rho}$ is a consistent estimate of ρ, $\tilde{\rho} - \rho = o_P(1)$. Hence, we conclude that, as $\lim_{n\to\infty} h_n = \infty$, $(1/n)(\tilde{R}_nZ_n)'(\tilde{R}_nZ_n) = (1/n)(R_nZ_n)'(R_nZ_n) + o_P(1)$. As $\tilde{R}_n = R_n - (\tilde{\rho} - \rho)W_{n,n}$, $\tilde{R}_nU_n = V_n + (\rho - \tilde{\rho}) \times W_{n,n}U_n$. These results imply that

$$\begin{aligned}\hat{\delta}_{FG} - \delta &= \left[\frac{1}{n}(\tilde{R}_nZ_n)'(\tilde{R}_nZ_n)\right]^{-1}\frac{1}{n}(\tilde{R}_nZ_n)'(\tilde{R}_nU_n)\\ &= \left[\frac{1}{n}(R_nE(Z_n))'(R_nE(Z_n)) + o_P(1)\right]^{-1}\\ &\quad\times\frac{1}{n}(R_nZ_n + (\rho - \tilde{\rho})W_{n,n}Z_n)'(V_n + (\rho - \tilde{\rho})W_{n,n}U_n).\end{aligned}$$

The asymptotic properties of $\hat{\delta}_{FG}$ will depend on the stochastic orders of $(R_nZ_n)'V_n$, $(W_{n,n}Z_n)'V_n$, $(R_nZ_n)'W_{n,n}U_n$, and $(W_{n,n}Z_n)'W_{n,n}U_n$. Explicitly, these terms can be decomposed:

$$\begin{aligned}(R_nZ_n)'V_n &= (R_nE(Z_n))'V_n + [0,(I_n - \rho W_{n,n})W_{n,n}T_nV_n]'V_n,\\ (W_{n,n}Z_n)'V_n &= (W_{n,n}E(Z_n))'V_n + (0,W_{n,n}^2T_nV_n)'V_n,\\ (R_nZ_n)'W_{n,n}U_n &= (R_nE(Z_n))'W_{n,n}U_n + [0,(I_n - \rho W_{n,n})W_{n,n}T_nV_n]'W_{n,n}U_n;\\ (W_{n,n}Z_n)'W_{n,n}U_n &= (W_{n,n}E(Z_n))'W_{n,n}U_n + (0,W_{n,n}G_nU_n)'W_{n,n}U_n.\end{aligned}$$

By Lemma A.1, $E[(W_{n,n}A_nV_n)'V_n]^2$, where $A_n = T_n$ or $W_{n,n}T_n$, has the order $\max\{O(n/h_n), O(n/h_n)^2\}$. If $\rho \neq 0$, $((I_n - \rho W_{n,n})W_{n,n}T_nV_n)'W_{n,n}U_n = (1/\rho)((I_n - \rho W_{n,n})W_{n,n}T_nV_n)'(U_n - V_n)$ and $(W_{n,n}G_nU_n)'W_{n,n}U_n = (1/\rho)(W_{n,n}G_nU_n)'(U_n - V_n)$. If $\rho = 0$, $U_n = V_n$ and, then,

$$((I_n - \rho W_{n,n})W_{n,n}T_nV_n)'W_{n,n}U_n = ((I_n - \rho W_{n,n})W_{n,n}T_nV_n)'W_{n,n}V_n$$

and $(W_{n,n}G_nU_n)'W_{n,n}U_n = (W_{n,n}G_nV_n)'W_{n,n}V_n$. These expressions allow the applications of Lemmas A.1, A.3, and A.4. By Lemma A.1, $E[(W_{n,n}G_nU_n)'V_n]^2 =$

https://doi.org/10.1017/S0266466602182028 Published online by Cambridge University Press

$E[(W_{n,n}G_nR_n^{-1}V_n)'V_n]^2 = \max\{O(n/h_n), O(n/h_n)^2\}$. Lemmas A.3 and A.4 imply that $E[(W_{n,n}A_nV_n)'U_n]^2$, $E(W_{n,n}A_nU_n)'U_n]^2$, and $E[(W_{n,n}A_nV_n)'W_{n,n}V_n]^2$, where $A_n = T_n, W_{n,n}T_n$, or G_n, are all of order $\max\{O(n/h_n)^2, \min[O(n^3/h_n^4), O(n^2/h_n)]\}$. The stochastic orders of these terms are relevant for asymptotic analysis.

For consistency, as $h_n \to \infty$, the preceding stochastic orders imply that $(1/n)((I_n - \rho W_{n,n})W_{n,n}T_nV_n)'V_n$, $(1/n)(W_{n,n}^2T_nV_n)'V_n$, $(1/n)(I_n - \rho W_{n,n})W_{n,n}T_nV_n)'W_{n,n}U_n$, and $(1/n)(W_{n,n}G_nU_n)'W_{n,n}U_n$ are of order $o_P(1)$. Lemma A.2 implies that $(1/n)(R_nE(Z_n))'V_n$ and $(1/n)(W_{n,n}E(Z_n))'V_n$ are $o_P(1)$. Similarly, because $U_n = R_n^{-1}V_n$, $(1/n)(R_nE(Z_n))'W_{n,n}U_n$ and $(1/n)(W_{n,n}E(Z_n))'W_{n,n}U_n$ are also $o_P(1)$. Therefore, $(1/n)(R_nZ_n + (\rho - \tilde{\rho})W_{n,n}Z_n)'(V_n + (\rho - \tilde{\rho})W_{n,n}U_n) = o_P(1)$ and, hence, $\hat{\delta}_{FG}$ is consistent.

For asymptotic distribution, as $(n/h_n^2) \to 0$, the stochastic orders of relevant terms imply that $(1/\sqrt{n})((I_n - \rho W_{n,n})W_{n,n}T_nV_n)'V_n$, $(1/\sqrt{n})(W_{n,n}^2T_nV_n)'V_n$, $(1/\sqrt{n})(I_n - \rho W_{n,n})W_{n,n}T_nV_n)'W_{n,n}U_n$ and $(1/\sqrt{n})(W_{n,n}G_nU_n)'W_{n,n}U_n$ are of order $o_P(1)$. Therefore,

$$\frac{1}{\sqrt{n}}(R_nZ_n + (\rho - \tilde{\rho})W_{n,n}Z_n)'(V_n + (\rho - \tilde{\rho})W_{n,n}U_n)$$

$$= \frac{1}{\sqrt{n}}(R_nE(Z_n) + (\rho - \tilde{\rho})W_{n,n}E(Z_n))'(V_n + (\rho - \tilde{\rho})W_{n,n}U_n) + o_P(1).$$

Lemma A.2 implies that $(1/\sqrt{n})(R_nE(Z_n))'V_n$ and $(1/\sqrt{n})(M_{n,n}E(Z_n))'V_n$ are stochastically bounded. Because $U_n = R_n^{-1}V_n$, $(1/\sqrt{n})(R_{n,n}E(Z_n))'W_{n,n}U_n$ and $(1/\sqrt{n}) \times (W_{n,n}Z_n)'W_{n,n}U_n$ are also stochastically bounded. As $\tilde{\rho}$ is a consistent estimate of ρ, it follows that

$$\frac{1}{\sqrt{n}}(R_nZ_n + (\rho - \tilde{\rho})W_{n,n}Z_n)'(V_n + (\rho - \tilde{\rho})W_{n,n}U_n) = \frac{1}{\sqrt{n}}(R_nE(Z_n))'V_n + o_P(1).$$

Finally, the asymptotic distribution of $\hat{\delta}_{FG}$ follows from Lemma A.2 and the Slutsky lemma. ■

https://doi.org/10.1017/S0266466602182028 Published online by Cambridge University Press