INTERNATIONAL ECONOMIC REVIEW
Vol. 40, No. 2, May 1999

A GENERALIZED MOMENTS ESTIMATOR FOR THE AUTOREGRESSIVE PARAMETER IN A SPATIAL MODEL*

BY HARRY H. KELEJIAN AND INGMAR R. PRUCHA[1]

University of Maryland, U.S.A.

This paper is concerned with the estimation of the autoregressive parameter in a widely considered spatial autocorrelation model. The typical estimator for this parameter considered in the literature is the (quasi) maximum likelihood estimator corresponding to a normal density. However, as discussed in this paper, the (quasi) maximum likelihood estimator may not be computationally feasible in many cases involving moderate- or large-sized samples. In this paper we suggest a generalized moments estimator that is computationally simple irrespective of the sample size. We provide results concerning the large and small sample properties of this estimator.

1. INTRODUCTION

There exists a large body of literature that considers autocorrelation of the disturbances across cross-sectional units for panel data, that is, data that are observed both across cross-sectional units and over time. However, the estimation of models that permit for autocorrelation of the disturbances across cross-sectional units for cases in which the data are only observed in one time period has, until recently, only received relatively little attention in the theoretical econometrics literature. For example, in most econometric textbooks there is no discussion relating to spatial models when only a single cross section of data is available.[2] This is unfortunate because issues relating to geographic proximity, transportation, spillover effects, etc., suggest that such models are important. Indeed, in recent years there have been a number of theoretical and applied econometric studies involving spatial issues, which include contributions by Case (1991), Conley (1996), Delong and Summers (1991), Dubin (1988), Kelejian and Robinson (1993), Moulton (1990), Quah (1992) and Topa (1996).[3]

*Manuscript received February 1995; revised December 1997.

[1] We would like to thank Michael Binder, Benedikt Pötscher, You-Qiang Wang, an anonymous referee, and the editors for helpful comments and Dennis Robinson for providing some of the weighting matrices.

[2] Of course, if panel data are available one can consider, for example, a seemingly unrelated regression model, or an error component model to permit for cross sectional correlation, and estimate the cross-sectional correlations via the time dimension of the panel if the time dimension is sufficiently large.

[3] There is an extensive literature relating to spatial models in the regional science and geography literature; see, for example, Anselin (1988), Bennett and Hordijk (1986), Cliff and Ord (1973, 1981), and Cressie (1993) and the references cited therein. For critical comments, see Kelejian and Robinson (1995).

One of the most widely referenced models of spatial autocorrelation is one that was put forth by Cliff and Ord (1973, 1981). This model is a variant of the model considered by Whittle (1954) and is sometimes referred to as a *spatial autoregressive* (*SAR*) *model* (see, for example Anselin 1988). In the SAR model, the disturbance term corresponding to a cross-sectional unit is, as discussed in more detail below, modeled as a weighted average of disturbances corresponding to other cross-sectional units, plus an innovation. This weighted average involves a scalar parameter, say ρ, and a set of weights that describe the spatial interactions. The innovations are typically assumed to be i.i.d. $N(0, \sigma^2)$. In a regression framework, the parameters of interest would then be ρ, σ^2 and the vector of regression coefficients. Typically, the spatial weights do not involve unknown parameters.[4]

Regression models containing spatially correlated disturbance terms based on the SAR model are typically estimated by the (quasi) maximum likelihood (ML) estimator, where the likelihood function corresponds to the normal distribution. We use the term *(quasi) ML estimator* rather than the term *ML estimator* to cover specifications where the actual distribution is permitted to differ from the normal distribution, as is the case in our analysis below. Given appropriate conditions, these (quasi) ML estimators should be consistent and asymptotically normally distributed. However, to the best of our knowledge, formal results establishing these properties under a specific set of low-level assumptions do not seem to be available for the SAR model considered here. We note, however, that Mardia and Marshall (1984) give a general result concerning the consistency and asymptotic normality of the ML estimator for regression models with general disturbance covariances, provided that the disturbances are normally distributed. Clearly, their theorem will cover many Gaussian spatial processes. However, in a formal sense their theorem is not applicable to the typical SAR model, even in the case where the disturbances are normally distributed. The reason for this is that Mardia and Marshall assume that the elements of the disturbance covariance matrix do not depend on the sample size. As will be seen below, this assumption is not satisfied for the typical SAR model.[5]

A practical difficulty with the (quasi) ML method in SAR models is that the estimation of ρ entails significant computational complexities. As our discussion will make clear, these complexities can be overwhelming if the spatial weights are not symmetric, which is typically the case in practice, even if the sample size is only moderate, or if the sample size is large, which is also the case in various applications

[4] See, for example, Anselin (1988, 1990) and the references cited therein. For an empirical study involving a parameterized weighting matrix, see Dubin (1988).

[5] Of course, the general literature on (quasi) ML estimation contains various sets of sufficient conditions under which (quasi) ML estimators are consistent and asymptotically normally distributed; see, for example, Gallant and White (1988), Heijmans and Magnus (1986, 1987), and Pötscher and Prucha (1991a, 1991b) for recent contributions in the econometrics literature as well as for other references. One approach to formally establish the asymptotic properties of the (quasi) ML estimators under a specific set of low-level assumptions for the SAR model considered here would be to formally establish that those assumptions are covered by one of the sets of sufficient conditions given in the general literature on (quasi) ML estimation. We note, however, that such a demonstration may be involved. Also, if N cross sections are observed not only for one but for T periods, spatial autocorrelation can be modeled in a general fashion via a seemingly unrelated regression model, and standard large sample theory can be applied to the case in which N is fixed and $T \to \infty$.

(e.g., there are more than 3000 counties in the United States). These practical difficulties are troublesome since, as Cliff and Ord (1981, p. 153) suggest, thus far the only available alternative to the (quasi) ML estimator of ρ in the SAR model is a moments estimator, which was suggested by Ord (1975). This estimator, however, is generally not seriously considered because of its inefficiency (see, for example, Ord 1975, p. 122).[6]

The purpose of this paper is twofold. First, on a theoretical level, we suggest an estimator for the parameter ρ in the SAR model based on a "generalized" moments approach. This estimator is, relative to the (quasi) ML estimator, computationally simple. We then provide a formal proof for the consistency of the estimator under an explicit set of conditions. We note that these conditions do not involve the assumption of normality. Second, we give Monte Carlo results relating to, among other things, the small sample distribution of our suggested estimator and the (quasi) ML estimator. These results suggest that under a variety of distributions, including the normal distribution, our estimator of ρ is "virtually as efficient" as the (quasi) ML estimator, defined as the maximizer of the likelihood function corresponding to the normal distribution.

In the context of a regression model, we also demonstrate that, under typical assumptions, ρ is a nuisance parameter in the sense that the feasible generalized least squares (feasible GLS) estimator based on a consistent estimator of ρ is asymptotically equivalent to the GLS estimator. Therefore, the importance of our results concerning the estimation of ρ also relate to the computational simplicity of feasible GLS estimators. As a by-product, we also establish the limiting distribution of those estimators. We note that this requires the use of a central limit theorem for triangular arrays.

Recently, in an interesting dissertation, Conley (1996) has considered a class of generalized method of moments estimators within a spatial setting. Rather than to assume a specific model for the generation of the data, he maintains that the data are stationary and spatially mixing. Clearly, avoiding specific modeling assumptions is appealing with regard to issues of potential misspecification. On the other hand, Conley's stationarity assumption may be restrictive in many applied settings. Also, this assumption is in general not satisfied by the class of spatial ARMA processes as defined, for example, in Anselin and Florax (1995), including the SAR model considered here, because of the nature of the spatial weighting matrices used in modeling those processes.[7] Additionally, the derivation of asymptotic results for M-estimators, and in particular, generalized method of moments estimators, typically involves a demonstration that the objective function of the estimator converges uniformly over the parameter space to its asymptotic counterpart. Provided that the functions forming the objective function are "first moment continuous," that a

[6] Also, this estimator is specified as the solution of a single quadratic equation and hence, in general, is not well defined unless a further selection mechanism between the two possible roots is specified.

[7] As a further technical detail, let (z_i) denote the data generating process, and let θ denote the vector of unknown parameters. Then Conley considers moments of the form $Eg(z_i, \theta) = 0$, $i = 1, \ldots, N$, where g is some vector valued function. In contrast, the moments utilized in this paper are of the form $Eg_{i,N}(z_1, \ldots, z_N, \theta) = 0$, $i = 1, \ldots, N$.

"local" law of large numbers holds, and given compactness of the parameter space, the desired uniform convergence follows immediately from Wald's (1949) approximation technique (see, for example, Pötscher and Prucha 1989, pp. 680–681). Conley maintains "first moment continuity" as an assumption toward establishing uniform convergence. However, in particular applications, a verification of this high-level assumption may be "involved." In this paper we deduce the needed uniform convergence from a set of lower-level assumptions. We note further that Conley's dissertation also provides a treatment of covariance matrix estimators in a spatial setting.

The SAR model is specified and interpreted in Section 2. This section also contains a discussion relating to (quasi) maximum likelihood estimation. Our estimator and a variation of it are defined and discussed in Section 3. Results showing that ρ is a nuisance parameter in a regression framework are given in Section 4. The Monte Carlo study is described and results relating to our suggested estimators, as well as to the (quasi) maximum likelihood estimator, are given in Section 5. Section 6 contains suggestions for further work. All proofs are relegated to the Appendix.

2. THE SPATIAL AUTOREGRESSIVE MODEL

In the SAR model an $N \times 1$ disturbance vector u is generated as follows:

$$u = \rho M u + \epsilon \tag{1}$$

where M is an $N \times N$ matrix of known constants, ρ is a scalar parameter, which is typically referred to as the *spatial autoregressive parameter*, and ϵ is an $N \times 1$ vector of innovations. For reasons that will become evident, M is often referred to as a *spatial weighting matrix*. For reasons of generality, we permit the elements of M and ϵ to depend on N, that is, to form triangular arrays. However, for simplicity of notation, we do not indicate this possible dependence on N explicitly in the following.

It proves helpful to introduce the following notational conventions: In general, we denote the ith element of a vector v as v_i and the (i, j)th element of a matrix A as a_{ij}. Correspondingly, we denote the ith row and jth column of A as $a_{i.}$ and $a_{.j}$. Given this notation, the typical assumptions of the SAR model are as follows[8]:

ASSUMPTION 1. *The innovations $\epsilon_1, \ldots, \epsilon_N$ are independently and identically distributed (for all N) with zero mean and variance σ^2, where $0 < \sigma^2 < b$ with $b < \infty$. Additionally, the innovations are assumed to possess finite fourth moments.*

ASSUMPTION 2. (a) *All diagonal elements of M are zero.* (b) $|\rho| < 1$. (c) *The matrix $I - \rho M$ is nonsingular for all $|\rho| < 1$.*

Given these assumptions, it follows from Eq. (1) that $u = (I - \rho M)^{-1}\epsilon$. Thus $E(u) = 0$ and $E(uu') = \Omega(\rho)$, where

$$\Omega(\rho) = \sigma^2 (I - \rho M)^{-1} (I - \rho M')^{-1} \tag{2}$$

[8] Generalizations and variations on these assumptions have been considered (see, for example, Anselin 1988 and Cliff and Ord 1973, 1981).

We note that, in general, the elements of $(I-\rho M)^{-1}$ will depend on the sample size N. As a consequence, in general, the elements of u also will depend on N and thus form a triangular array, even if the elements of ϵ do not depend on N. It also follows that, in general, the elements of $\Omega(\rho)$ will depend on N.[9]

The specification in Eq. (1) implies that $u_i = \rho \sum_{j=1}^{N} m_{ij} u_j + \epsilon_i$, $i = 1, \ldots, N$. In a cross-sectional setting, the nonzero weights m_{ij} are often specified to be those which correspond to units which relate to the ith unit in a meaningful way. Such units are often said to be *neighbors* of unit i. As one example, if the cross-sectional units are geographic regions, one might take $m_{ij} \neq 0$ if the ith and jth regions are contiguous and $m_{ij} = 0$ otherwise. In this setting, each disturbance consists of a weighted sum of disturbances in related regions and a term that is i.i.d. over the regions. Clearly, Assumption 2(a) is a normalization of the model, Assumption 2(b) is a stability condition for certain specifications of M, and Assumption 2(c) ensures that the disturbance vector u is uniquely defined in terms of the innovation vector ϵ.[10] One implication of a model such as Eq. (1) is that, unlike for most time series models, m_{ij} need not be zero for $j > i$. Thus one distinguishing feature of a spatial model is that the ith disturbance term may be directly related to both "future" and "past" disturbances. Also, in a spatial model there is typically no natural order for arranging the sample.

Assuming for the moment that u is observable and normally distributed, the log likelihood for the model in Eq. (1) is, using evident notation, given by

$$\text{(3)} \qquad \ln(\mathscr{L}) = -\frac{N}{2}\left[\ln(\sigma^2) + \ln(2\pi)\right] - \frac{1}{2\sigma^2} u'(I - \rho M')(I - \rho M)u + \ln\|I - \rho M\|$$

As remarked earlier, the normality of u is not one of our maintained assumptions, and hence we refer to the maximizers of Eq. (3) as (quasi) ML estimators. In the following we denote these (quasi) ML estimators for ρ and σ^2 as $\hat{\rho}_{QML}$ and $\hat{\sigma}^2_{QML}$, respectively. As is evident from Eq. (3), the computation of the (quasi) ML estimators involves the repeated evaluation of the determinant of the $N \times N$ matrix $I - \rho M$. To minimize the computational burden, Ord (1975) suggested that the troublesome term in Eq. (3) be expressed as $\ln\|I - \rho M\| = \sum_{i=1}^{N} \ln(|1 - \rho\lambda_i|)$, where λ_i denotes the ith eigenvalue of M. The advantage of this approach is that (since M is a known matrix) the eigenvalues of M only have to be computed once at the outset of the numerical optimization procedure employed in finding the (quasi) ML estimates and not repeatedly at each of the necessary numerical iterations. However, this still leaves the researcher with the task of finding the eigenvalues of the $N \times N$ matrix M. Unless M has a particular structure, this task is typically "challenging," especially if N is large—recall, for example, that there are over 3000 counties in the

[9] As remarked in the Introduction, this violates one of the assumptions maintained by Mardia and Marshall's (1984) theorem regrading the consistency and asymptotic normality of ML estimators for Gaussian processes.

[10] Kelejian and Robinson (1995) give results that suggest that Assumption 2(c) is satisfied for many specifications of M considered in the literature.

United States. In fact, in many cases it will be practically impossible to compute those eigenvalues accurately based on computing technology typically available to empirical researchers. As an illustration, in some of the Monte Carlo experiments reported below we use "idealized" symmetric M matrices in which each spatial unit has the same number of neighbors, say J. Clearly, for those matrices, all eigenvalues are real. However, when we employed a standard subroutine for computing the eigenvalues of a general matrix from the IMSL program library the routine reported eigenvalues with imaginary parts that differed substantially from zero even for the moderate sample size $N = 400$, when the number of neighbors J was 6 or larger.[11] In fact, some of the reported imaginary parts differed from zero by more than .5 in absolute value. Only when we employed a subroutine that utilized the symmetric nature of those M matrices were we able to compute the eigenvalues accurately. Since, in practice, spatial weighting matrices are typically not symmetric, this suggests that an accurate computation of the (quasi) ML estimator may not be feasible in many cases even for moderate sample sizes.[12] Given these computational problems, it is clearly important to have an alternative to the (quasi) ML estimator, which
is computationally feasible for general weighting matrices M, and large sample sizes N.

3. DEFINITION AND CONSISTENCY OF A GENERALIZED MOMENTS ESTIMATOR OF ρ

Suppose u defined in Eq. (1) represents the disturbance vector in a model, and based on that model, $\tilde{u}$ is a predictor of u. For notational convenience, let $\bar{u} = Mu$ and $\bar{\bar{u}} = MMu$, and correspondingly, $\bar{\tilde{u}} = M\tilde{u}$, and $\bar{\bar{\tilde{u}}} = MM\tilde{u}$. Similarly, let $\bar{\epsilon} = M\epsilon$ and note that under Assumptions 1 and 2:

$$(4) \qquad E\left[\frac{1}{N}\epsilon'\epsilon\right] = \sigma^2 \qquad E\left[\frac{1}{N}\bar{\epsilon}'\bar{\epsilon}\right] = \sigma^2 N^{-1} Tr(M'M) \qquad E\left[\frac{1}{N}\bar{\epsilon}'\epsilon\right] = 0$$

Our generalized moments estimator for ρ is based on these three moments. Specifically, noting from Eq. (1) that $\epsilon = u - \rho\bar{u}$ and so $\bar{\epsilon} = \bar{u} - \rho\bar{\bar{u}}$, consider the following three-equation system implied by Eqs. (1) and (4):

$$(5) \qquad \Gamma_N[\rho, \rho^2, \sigma^2]' - \gamma_N = 0$$

[11] The IMSL subroutine employed was DEVLGR, which itself is based on subroutines from the EISPACK program library. It seems that other packages such as MATLAB also employ routines from EISPACK.

[12] We also experimented with MATLAB 4.2 for Windows, using a PC with a Pentium 133-MHz processor and 32 MB of memory, to calculate the eigenvalues for our "idealized" M matrices. In those experiments we encountered "out of memory" errors for M matrices with $N \geq 2000$ and $J = 10$, even when using a routine for sparse symmetric matrices. In terms of computational time, it took, for example, 22 minutes to compute the eigenvalues in the case $N = 1500$ and $J = 10$, again using a routine for sparse symmetric matrices. The subsequent computation of the (quasi) ML estimator based on those eigenvalues and using TSP 4.2 only took seconds. The computation of our generalized moments estimator, which does not require the computation of eigenvalues, also took only seconds.

where

$$\Gamma_N = \begin{bmatrix} \frac{2}{N}E(u'\bar{u}) & \frac{-1}{N}E(\bar{u}'\bar{u}) & 1 \\ \frac{2}{N}E(\bar{\bar{u}}'\bar{u}) & \frac{-1}{N}E(\bar{\bar{u}}'\bar{\bar{u}}) & \frac{1}{N}Tr(M'M) \\ \frac{1}{N}E(u'\bar{\bar{u}} + \bar{u}'\bar{u}) & \frac{-1}{N}E(\bar{u}'\bar{\bar{u}}) & 0 \end{bmatrix} \qquad \gamma_N = \begin{bmatrix} \frac{1}{N}E(u'u) \\ \frac{1}{N}E(\bar{u}'\bar{u}) \\ \frac{1}{N}E(u'\bar{u}) \end{bmatrix}$$

Now consider the following analogue to Eq. (5) in terms of sample moments based on $\tilde{u}$:

$$G_N[\rho, \rho^2, \sigma^2]' - g_N = \nu_N(\rho, \sigma^2) \tag{6}$$

where

$$G_N = \begin{bmatrix} \frac{2}{N}\tilde{u}'\tilde{\bar{u}} & \frac{-1}{N}\tilde{\bar{u}}'\tilde{\bar{u}} & 1 \\ \frac{2}{N}\tilde{\bar{\bar{u}}}'\tilde{\bar{u}} & \frac{-1}{N}\tilde{\bar{\bar{u}}}'\tilde{\bar{\bar{u}}} & \frac{1}{N}Tr(M'M) \\ \frac{1}{N}\left(\tilde{u}'\tilde{\bar{\bar{u}}} + \tilde{\bar{u}}'\tilde{\bar{u}}\right) & \frac{-1}{N}\tilde{\bar{u}}'\tilde{\bar{\bar{u}}} & 0 \end{bmatrix} \qquad g_N = \begin{bmatrix} \frac{1}{N}\tilde{u}'\tilde{u} \\ \frac{1}{N}\tilde{\bar{u}}'\tilde{\bar{u}} \\ \frac{1}{N}\tilde{u}'\tilde{\bar{u}} \end{bmatrix}$$

and where the 3×1 vector $\nu_N(\rho, \sigma^2)$ can be viewed as a vector of residuals. We now define our generalized moments estimator for ρ and σ^2 as the nonlinear least squares estimator, say $\hat{\rho}_{NLS,N}$ and $\hat{\sigma}^2_{NLS,N}$, corresponding to Eq. (6). More specifically,

(7)

$$\left(\hat{\rho}_{NLS,N}, \hat{\sigma}^2_{NLS,N}\right) = \operatorname{argmin}\left\{\nu_N\left(\underline{\rho}, \underline{\sigma}^2\right)'\nu_N\left(\underline{\rho}, \underline{\sigma}^2\right) : \underline{\rho} \in [-a, a], \underline{\sigma}^2 \in [0, b]\right\}$$

where $a \geq 1$.

Remark 1. Note that Eq. (7) implies that $|\hat{\rho}_{NLS,N}| \leq a$ with $a \geq 1$. Since $|\rho| < 1$, if the bound a is sufficiently large, $\hat{\rho}_{NLS,N}$ is essentially the unconstrained nonlinear least squares estimator of ρ. The existence and measurability of $\hat{\rho}_{NLS,N}$ and $\hat{\sigma}^2_{NLS,N}$ is ensured by, for example, Lemma 2 in Jennrich (1969).

In the following, let $P(\rho) = (I - \rho M)^{-1}$. We now specify three additional assumptions.

Assumption 3. *(a) The sums* $\Sigma_{i=1}^N |m_{ij}|$ *and* $\Sigma_{j=1}^N |m_{ij}|$ *are bounded by, say,* $c_m < \infty$ *for all* $1 \leq i, j \leq N$, $N \geq 1$. *(b) The sums* $\Sigma_{i=1}^N |p_{ij}(\rho)|$ *and* $\Sigma_{j=1}^N |p_{ij}(\rho)|$ *are bounded by, say,* $c_p < \infty$ *for all* $1 \leq i, j \leq N$, $N \geq 1$, $|\rho| < 1$, *where* c_p *may depend on* ρ.

ASSUMPTION 4. *Let $\tilde{u}_i$ denote the ith element of $\tilde{u}$, where again we suppress the dependence of $\tilde{u}$ and its elements on N for notational convenience. We then assume that there exist (finite dimensional) random vectors d_{iN} and Δ_N such that $|\tilde{u}_i - u_i| \leq \|d_{iN}\| \, \|\Delta_N\|$ with $N^{-1}\sum_{i=1}^N \|d_{iN}\|^{2+\delta} = O_p(1)$ for some $\delta > 0$ and $N^{1/2}\|\Delta_N\| = O_p(1)$.*[13]

ASSUMPTION 5. *The smallest eigenvalue of $\Gamma_N'\Gamma_N$ is bounded away from zero, that is, $\lambda_{\min}(\Gamma_N'\Gamma_N) \geq \lambda_* > 0$, where λ_* may depend on ρ and σ^2.*

REMARK 2. (a) In practice, spatial models are often formulated in such a way that each cross-sectional unit has a limited number of "neighbors" regardless of the sample size (see, for example, Case 1991 and Kelejian and Robinson 1995). In such cases, the weighting matrix M is sparse for large N, and so Assumption 3(a) would be satisfied. As a point of information, we note that in many of these cases the elements of M are taken to be nonnegative and row normalized in that $\sum_j m_{ij} = 1$. In still other cases, the weighting matrix does not contain zeros, but its elements are assumed to decline rapidly in certain directions because they are defined in terms of variables such as distance (see for example, Dubin 1988 and De Long and Summers 1991). Again, under further reasonable (but idealized) conditions, Assumption 3(a) would be expected to hold.

(b) Recall from Eq. (2) that $\Omega = \sigma^2 PP'$. Assumption 3(b) then implies that $N^{-1}\sum_{i=1}^N \sum_{j=1}^N |\omega_{ij}|$ is bounded, thus limiting the degree of correlation.[14] In a time series context this condition ensures that the process possesses a fading memory. We also note that Assumption 3(b) is closely related to Condition A5 in Mandy and Martins-Filho (1994) in their study of large sample properties of feasible GLS estimators.

REMARK 3. Assumption 4 should be satisfied for most cases in which $\tilde{u}$ is based on $N^{1/2}$-consistent estimators of the regression coefficients. For example, using evident notation, consider the nonlinear regression model $y_i = f(x_i, \beta) + u_i$. Let $\tilde{\beta}_N$ denote the nonlinear least squares estimator, and let $\tilde{u}_i = y_i - f(x_i, \tilde{\beta}_N)$. Assuming that f is continuously differentiable and applying the mean value theorem, it is readily seen that $|\tilde{u}_i - u_i| \leq \|d_{iN}\| \, \|\Delta_N\|$ with $d_{iN} = \sup_{\underline{\beta}} |\partial f(x_i, \underline{\beta})/\partial \underline{\beta}|$ and $\Delta_N = \tilde{\beta}_N - \beta$. Under typical assumptions maintained for the nonlinear regression model, d_{iN} and Δ_N will satisfy the conditions postulated in Assumption 3 (see, for example, Pötscher and Prucha 1986).

REMARK 4. It will become evident that Assumption 5 is an identifiability condition.

[13] For definiteness, let A be some vector or matrix; then $\|A\| = [Tr(A'A)]^{1/2}$. We note that this norm is submultiplicative, that is, $\|AB\| \leq \|A\| \, \|B\|$. We also define $|A|$ as the vector or matrix of absolute values.

[14] Observe that $N^{-1}\sum_{i=1}^N \sum_{j=1}^N |\omega_{ij}| \leq \sigma^2 N^{-1}\sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^N |p_{ik}| \, |p_{jk}| = \sigma^2 N^{-1} \sum_{k=1}^N \sum_{i=1}^N |p_{ik}| \sum_{j=1}^N |p_{jk}| \leq \sigma^2 c_p^2 < \infty$.

Our basic result is Theorem 1, whose proof is given in the Appendix.

THEOREM 1. *Let* $\hat{\rho}_{NLS,N}$ *and* $\hat{\sigma}^2_{NLS,N}$ *be the nonlinear least squares estimators defined by Eq.* (7). *Then, given Assumptions* 1 *to* 5,

$$\left(\hat{\rho}_{NLS,N}, \hat{\sigma}^2_{NLS,N}\right) \xrightarrow{p} \left(\rho, \sigma^2\right) \text{ as } N \to \infty$$

An obvious variation on $\hat{\rho}_{NLS,N}$ in Theorem 1 is based on an overparameterization of Eq. (6). Specifically, let $\varphi = \rho^2$, $\alpha = (\rho, \varphi, \sigma^2)$, and let $\hat{\alpha}_{OLS,N} = (\hat{\rho}_{OLS,N}, \hat{\varphi}_{OLS,N}, \hat{\sigma}^2_{OLS,N})$ be the ordinary least squares estimator of α based on Eq. (6). Then, it is evident from the proof of Theorem 1 that, under the same conditions, $(\hat{\rho}_{OLS,N}, \hat{\varphi}_{OLS,N}, \hat{\sigma}^2_{OLS,N}) \xrightarrow{p} (\rho, \varphi, \sigma^2)$ as $N \to \infty$.

4. AN APPLICATION TO THE GENERALIZED LEAST SQUARES MODEL

As discussed, the vector u defined in Eq. (1) often will represent the vector of disturbances of some econometric model. In such cases, ρ often will be a nuisance parameter in the sense that the asymptotic distribution of some estimator of the model parameters of interest will be the same if ρ is known or if ρ is replaced by a consistent estimator. In many of these cases it will be possible to estimate the disturbances $N^{1/2}$-consistently in a first step. The force of Theorem 1 is that based on those estimated disturbances, a simple and consistent estimator of ρ is available.

In the following we illustrate this point within the context of a linear regression model with spatially autoregressive disturbances. In particular, consider the following model:

$$y = X\beta + u \tag{8}$$

where y is the $N \times 1$ vector of observations on the dependent variable, X is the $N \times K$ matrix of observations on the explanatory variables, β is the $K \times 1$ vector of unknown model parameters, and u is the vector of disturbances assumed to be generated by Eq. (1). As discussed in Section 2, in general, the elements of u and hence those of y will depend on N. For reasons of generality, we also permit the elements of X to depend on N, but again, we do not indicate this possible dependence on N explicitly. We maintain the following typical assumptions for the regressor matrix X and the variance covariance matrix $\Omega(\delta)$ of the disturbance vector u.

ASSUMPTION 6. *The elements of X are nonstochastic and bounded in absolute value by* c_x, $0 < c_x < \infty$. *Also, X has full column rank, and the matrix* $Q_x = \lim_{N\to\infty} N^{-1}X'X$ *is finite and nonsingular. Furthermore, the matrices* $\underline{Q}_x(\rho) = \lim_{N\to\infty} N^{-1}X'\Omega(\rho)^{-1}X$ *and* $\overline{Q}_x(\rho) = \lim_{N\to\infty} N^{-1}X'\Omega(\rho)X$ *are finite and nonsingular for all* $|\rho| < 1$.

The true GLS estimator for β is defined as $\tilde{\beta}_N^G = [X'\Omega(\rho)^{-1}X]^{-1}X'\Omega(\rho)^{-1}y$, and the feasible GLS estimator for β corresponding to some estimator of ρ, say $\tilde{\rho}_N$, is defined as $\tilde{\beta}_N^{FG} = [X'\Omega(\tilde{\rho}_N)^{-1}X]^{-1}X'\Omega(\tilde{\rho}_N)^{-1}y$ (where A^{-1} denotes the

Moore-Penrose generalized inverse of a matrix A, if that matrix is singular). The following theorem first establishes the asymptotic distribution of $\tilde{\beta}_N^G$ and then shows that $\tilde{\beta}_N^{FG}$ has the same asymptotic distribution as $\tilde{\beta}_N^G$ if $\tilde{\rho}_N$ is a consistent estimator for ρ. All proofs are relegated to the Appendix.

THEOREM 2. *Given that Assumptions* 1 *to* 3 *and* 6 *hold*:

(a) *The true GLS estimator* $\tilde{\beta}_N^G$ *is a consistent estimator for* β, *and*

$$N^{1/2}\left[\tilde{\beta}_N^G - \beta\right] \xrightarrow{D} N\left[0, \sigma^2 \underline{Q}_x(\rho)^{-1}\right]$$

(b) *Let* $\tilde{\rho}_N$ *be a consistent estimator for* ρ. *Then the true GLS estimator* $\tilde{\beta}_N^G$ *and the feasible GLS estimator* $\tilde{\beta}_N^{FG}$ *have the same asymptotic distribution.*

(c) *Suppose further that* $\tilde{\sigma}_N^2$ *is a consistent estimator for* σ^2. *Then* $\tilde{\sigma}_N^2[N^{-1}X'\Omega(\tilde{\rho}_N)^{-1}X]^{-1}$ *is a consistent estimator for* $\sigma^2\underline{Q}_x(\rho)^{-1}$.

As remarked in the Introduction, for the spatial model considered here, a rigorous proof of the asymptotic distribution of the GLS estimator $\tilde{\beta}_N^G$ requires the use of a central limit theorem for triangular arrays (even if the elements of M and X do not depend on N). Such a central limit theorem is given in the Appendix.

Theorem 2 assumes the existence of a consistent estimator of ρ and σ^2. We demonstrate in the Appendix that under Assumptions 1 to 3 and 5 and 6 the ordinary least squares (OLS) estimator $\tilde{\beta}_N = [X'X]^{-1}X'y$ is $N^{1/2}$-consistent. Given this, the corresponding residuals $\tilde{u}_i = y_i - x_{i.}\tilde{\beta}_N$ satisfy Assumption 4 with $d_{iN} = |x_{i.}|$ and $\Delta_N = \tilde{\beta}_N - \beta$. Thus, via the suggested generalized moments estimator and Theorem 1, these residuals can be used to obtain consistent estimators of ρ and σ^2. According to Theorem 2, these estimators can then be used in formulating a feasible GLS estimator (and an estimator for its asymptotic variance covariance matrix) with the feasible and true GLS estimator being asymptotically equivalent.

5. A MONTE CARLO MODEL STUDY

It is of interest to analyze the small sample properties of the generalized moments estimators $\hat{\rho}_{NLS}$ and $\hat{\rho}_{OLS}$ and compare them with those of the (quasi) maximum likelihood estimator $\hat{\rho}_{QML}$ defined as the maximizer of the normal log likelihood function (3).[15] For this purpose, we have conducted a two-part Monte Carlo study. The first part of the Monte Carlo study is based on "idealized" weighting matrices M that differ in size and in the number of neighbors. For these idealized weighting matrices, the number of neighbors per unit is taken to be the same in each of the respective matrices. For future reference, we note that, for a given sample size, the number of neighbors per unit can be viewed as a measure of the sparseness of that matrix. In using these idealized weighting matrices, we can readily explore the effects of sample size and number of neighbors on the small sample properties of our considered estimators. Of course, the use of idealized weighting matrices raises

[15] We note that $\hat{\rho}_{QML}$ (and $\hat{\sigma}_{QML}^2$) denote the (joint) maximizers of the normal log likelihood function (3), even if the actual distribution is not normal.

the concern that results corresponding to those matrices may not be representative of results corresponding to "real world" matrices. The second part of the Monte Carlo study is hence based on real-world weighting matrices.

For both parts of the Monte Carlo study we consider three distributions of ϵ and seven selections of ρ. As discussed in more detail below, we consider a total of 36 cases for each distribution of ϵ. The results for each case are based on 500 Monte Carlo replications. To summarize the results of the respective Monte Carlo experiments, we estimate response functions. It turns out that the estimated response functions based on idealized and real-world weighting matrices are not "significantly" different.[16] The estimates for the response functions reported below hence will be based on both sets of weighting matrices. These response functions also can be used to interpolate results for other cases.

We now describe the design of the Monte Carlo experiments in more detail. Note first from Eq. (1) that σ is a scale factor for u, as well as for $\bar{u}$ and $\bar{\bar{u}}$, in that their standard deviations are proportional to σ. Because of this, the estimators for ρ defined earlier do not depend on σ^2. Hence, without loss of generality, we took $\sigma^2 = 1$ in generating the data for all the experiments considered; however, in all the experiments, σ^2 was viewed as an unknown parameter concerning estimation.

The first distribution for ϵ explored in the experiments is the normal. More specifically, we assume that the ϵ_i are i.i.d. $N(0,1)$. This case is viewed as a base case for the small sample comparisons, since in this case $\hat{\rho}_{QML}$ is actually the maximum likelihood estimator. The second distribution considered is a normalized version of the log normal. More specifically, we assume in this case that $\epsilon_i = [\exp(\xi_i) - \exp(.5)]/[\exp(2) - \exp(1)]^{.5}$, where the ξ_i are i.i.d. $N(0,1)$. The normalization implies that the ϵ_i are i.i.d. $(0,1)$. This distribution was considered because it is not symmetric. The third distribution considered is a normalized version of a mixture of normals in which one normally distributed random variable is contaminated by another that has a larger variance. More specifically, we assume here that $\epsilon_i = [\lambda_i \xi_i + (1-\lambda_i)\zeta_i]/(5.95)^{.5}$, where the λ_i are i.i.d. Bernoulli variables with $\text{Prob}(\lambda_i = 1) = .95$, the ξ_i are i.i.d. $N(0,1)$, and the ζ_i are i.i.d. $N(0,100)$. Also, the processes (λ_i), (ξ_i) and (ζ_i) are assumed to be jointly independent. Again, the normalization implies that ϵ_i is i.i.d. $(0,1)$. This case was considered because the implied distribution has thicker tails than the normal.[17] In particular, for the specification considered $E\epsilon_i^4/(E\epsilon_i^2)^2 \doteq 14.15$.

As mentioned, the first part of the Monte Carlo study is based on idealized weighting matrices M. For each of the three distributions of ϵ we consider 15 cases that relate to seven selections of ρ, three selections of the weighting matrix M, and three selections of the sample size N. We note that the total number of combinations of these selections of ρ, M, and N would lead to $7 \times 3 \times 3 = 63$ cases for each distribution of ϵ. To keep the Monte Carlo study manageable, we consider only 15 of those cases per distribution of ϵ but summarize the results of the Monte Carlo experiments in terms of response functions. The three specifications of the weighting

[16] For our "test of significance," we employed the Chow test in a classical fashion. While in this context this testing procedure is not a formal one, it should be illustrative.

[17] We note that mixtures of normals are frequently used to model the effects of outliers.

matrices M differ in terms of sparseness and therefore in terms of the extent of implied autocorrelation concerning the disturbance terms u_i defined by Eq. (1). In the first specification, which we henceforth refer to as "1 ahead and 1 behind," M was selected such that each element of u is directly related to the one immediately after and immediately before it. In doing this, we specified a "circular" world so that, for example, u_N is directly related to u_1 and to u_{N-1} and, similarly, u_1 to u_2 and u_N. Furthermore, we specified M such that all nonzero elements of M are equal and that the respective rows sum to unity. That is, in this case, each row of M has two nonzero elements that are equal to $\frac{1}{2}$. Correspondingly, the next two specifications of M are "3 ahead and 3 behind" and "5 ahead and 5 behind," again in a circular world. The nonzero elements of M in these two cases are, respectively, taken as $\frac{1}{6}$ and $\frac{1}{10}$.[18] Let J denote the average number of neighbors for each unit. We can then characterize the preceding matrices with $J = 2, 6, 10$, respectively.

The second part of the Monte Carlo experiment is based on three real-world weighting matrices M. In particular, these matrices represent the spatial weighting matrices for 58, 100, and 254 counties in the states of California, North Carolina, and Texas. For these matrices, two counties are defined as neighbors if they are in the same state and if a 50-mile circle centered at the population center of one county includes the population center of the other county. Neighbors are indicated by nonzero elements in the M matrix. These nonzero elements are specified to be equal in each row and to sum to unity in each row. Again, we characterize these matrices by their average number of neighbors, that is, with $J = 3.8, 10.9, 6.6$, respectively. Given seven selections of ρ, the three real-world weighting matrices lead to $7 \times 3 = 21$ additional cases per distribution.

Table 1 gives results on two characteristics of the distributions of $\hat{\rho}_{NLS}$, $\hat{\rho}_{OLS}$, and $\hat{\rho}_{QML}$ for each of the $15 + 21 = 36$ cases (defined in terms of N, J, and ρ) for each of the three disturbance distributions considered. These characteristics are closely related to the standard measures of bias and root mean squared error (RMSE) but, unlike these measures, are assured to exist. Our measure of bias is defined as the difference between the median and the true parameter value. Our measure corresponding to the RMSE is defined as $[bias^2 + (IQ/1.35)^2]^{1/2}$, where IQ is the interquantile range. That is, $IQ = c_1 - c_2$, where c_1 is the .75 quantile and c_2 is the .25 quantile. If the distribution is normal, $IQ/1.35$ is (apart from rounding errors) equal to the standard deviation. In the following we will refer to our measures simply as bias and RMSE. The results in Table 1 are Monte Carlo estimates of these measures based on quantiles computed from the empirical distributions corresponding 500 Monte Carlo replications. Before discussing response functions for the RMSEs, we note some points.

The average absolute biases are generally similar for $\hat{\rho}_{QML}$ and $\hat{\rho}_{NLS}$ but higher for $\hat{\rho}_{OLS}$ for all three cases of considered distributions. The biases, while typically negative, are relatively small in absolute value. The RMSEs for $\hat{\rho}_{QML}$ and $\hat{\rho}_{NLS}$ are also generally very close in magnitude and considerably lower than those relating to

[18] We emphasize that the estimators for ρ considered in this paper do not depend on the particular ordering of the data. Thus any M matrix obtained from a rearrangement of the data would yield the same results.

TABLE 1A

BIASES AND RMSE OF ESTIMATORS FOR ρ, NORMAL ERROR DISTRIBUTION

			Bias			RMSE		
N	J	ρ	$\hat{\rho}_{QML}$	$\hat{\rho}_{NLS}$	$\hat{\rho}_{OLS}$	$\hat{\rho}_{QML}$	$\hat{\rho}_{NLS}$	$\hat{\rho}_{OLS}$
49.	2.0	−0.90	.0034	−.0012	−.0205	.0364	.0439	.2111
49.	2.0	0.90	−.0051	−.0031	.0222	.0351	.0436	.2155
49.	6.0	−0.50	−.0008	−.0345	−.0155	.2879	.2952	.2973
49.	6.0	0.50	−.0193	−.0184	−.0078	.1599	.1649	.4807
49.	10.0	0.00	−.0294	−.0416	−.0210	.3139	.3410	.7245
58.	3.8	−0.90	.0205	−.0068	−.0429	.1110	.1190	.4705
58.	3.8	−0.50	.0175	−.0126	.0217	.1615	.1715	.2964
58.	3.8	−0.25	.0113	−.0127	.0246	.1663	.1836	.2173
58.	3.8	0.00	−.0023	−.0153	−.0076	.1691	.1814	.2991
58.	3.8	0.25	−.0080	−.0135	.0423	.1610	.1632	.4441
58.	3.8	0.50	−.0139	−.0132	.1085	.1332	.1324	.6179
58.	3.8	0.90	−.0070	−.0049	.1209	.0434	.0475	.5320
100.	2.0	0.00	.0019	.0013	.0029	.0903	.0907	.0931
100.	6.0	−0.50	.0006	−.0106	−.0128	.1925	.2043	.2000
100.	6.0	0.50	−.0092	−.0064	.0141	.1076	.1121	.3553
100.	10.0	−0.25	−.0068	−.0234	.0067	.2292	.2431	.4278
100.	10.0	0.25	−.0145	−.0167	.0448	.1734	.1751	.5031
100.	10.9	−0.90	−.0143	−.0275	−.0477	.2871	.3195	.3025
100.	10.9	−0.50	−.0165	−.0249	−.0554	.2762	.2958	.3431
100.	10.9	−0.25	−.0148	−.0234	−.0300	.2573	.2721	.3937
100.	10.9	0.00	−.0147	−.0198	−.0227	.2390	.2432	.4449
100.	10.9	0.25	−.0197	−.0150	−.0105	.2024	.2058	.4635
100.	10.9	0.50	−.0123	−.0128	−.0007	.1588	.1618	.4610
100.	10.9	0.90	−.0061	−.0042	.0140	.0510	.0622	.3492
254.	6.6	−0.90	.0126	−.0039	−.0158	.0617	.0960	.1543
254.	6.6	−0.50	−.0020	−.0059	.0026	.1068	.1206	.1249
254.	6.6	−0.25	−.0012	−.0085	−.0042	.1116	.1171	.1221
254.	6.6	0.00	.0017	−.0080	.0016	.1093	.1097	.1525
254.	6.6	0.25	−.0041	−.0058	.0137	.0953	.0972	.1795
254.	6.6	0.50	−.0074	−.0061	.0195	.0763	.0795	.2025
254.	6.6	0.90	−.0042	−.0029	.0139	.0267	.0303	.1642
400.	2.0	−0.25	−.0010	−.0017	−.0035	.0464	.0463	.0628
400.	2.0	0.25	−.0026	−.0021	−.0019	.0449	.0461	.0606
400.	6.0	0.00	−.0093	−.0114	−.0022	.0811	.0833	.1408
400.	10.0	−0.90	−.0189	−.0122	−.0251	.1379	.1557	.1466
400.	10.0	0.90	−.0017	−.0018	−.0086	.0201	.0213	.1435
Column averages of absolute values			.0093	.0121	.0231	.1378	.1466	.2999

$\hat{\rho}_{OLS}$. This suggests that the generalized moments estimator $\hat{\rho}_{NLS}$ and the (quasi) maximum likelihood estimator $\hat{\rho}_{QML}$ possess very similar small sample properties, under both normality and nonnormality. We conjecture that a reason for this is that $\hat{\rho}_{QML}$ and $\hat{\rho}_{NLS}$ are both, in essence, defined in terms of second-order moments. Given the similarity of the small sample properties of $\hat{\rho}_{QML}$ and $\hat{\rho}_{NLS}$, a major advantage of the generalized moments estimator $\hat{\rho}_{NLS}$ as compared with the (quasi) maximum likelihood estimator $\hat{\rho}_{QML}$ seems to be that $\hat{\rho}_{NLS}$ remains readily computable even for large sample sizes N and general spatial weighting matrices M, as was discussed in some detail at the end of Section 2. We note, however, that the overparameterization underlying the definition of $\hat{\rho}_{OLS}$ is costly in terms of small

TABLE 1B

BIASES AND RMSE OF ESTIMATORS FOR ρ, LOG-NORMAL ERROR DISTRIBUTION

			Bias			RMSE		
N	J	ρ	$\hat{\rho}_{QML}$	$\hat{\rho}_{NLS}$	$\hat{\rho}_{OLS}$	$\hat{\rho}_{QML}$	$\hat{\rho}_{NLS}$	$\hat{\rho}_{OLS}$
49.	2.0	−0.90	.0036	−.0030	−.0087	.0337	.0357	.2005
49.	2.0	0.90	−.0013	−.0038	.0042	.0299	.0358	.2013
49.	6.0	−0.50	.0018	−.0008	−.0183	.2348	.2706	.2827
49.	6.0	0.50	−.0041	−.0010	.0451	.1426	.1456	.4277
49.	10.0	0.00	−.0211	−.0138	.0158	.2789	.3165	.5880
58.	3.8	−0.90	.0521	−.0317	−.2430	.1642	.1485	.7087
58.	3.8	−0.50	.0277	−.0556	.0199	.1948	.1936	.3784
58.	3.8	−0.25	−.0031	−.0557	.0675	.1832	.1942	.3569
58.	3.8	0.00	−.0147	−.0586	.0676	.1532	.1884	.4968
58.	3.8	0.25	−.0225	−.0479	.1685	.1478	.1755	.7617
58.	3.8	0.50	−.0223	−.0326	.2603	.1362	.1428	1.0143
58.	3.8	0.90	−.0106	−.0111	.1848	.0504	.0540	.8143
100.	2.0	0.00	−.0016	−.0014	−.0007	.0870	.0851	.0851
100.	6.0	−0.50	.0058	−.0040	−.0093	.1653	.1816	.1816
100.	6.0	0.50	−.0007	−.0007	.0117	.0938	.0953	.3031
100.	10.0	−0.25	−.0085	−.0143	.0155	.2268	.2405	.3735
100.	10.0	0.25	−.0116	−.0110	.0434	.1718	.1745	.4255
100.	10.9	−0.90	−.0038	−.0177	−.0306	.2547	.2920	.2772
100.	10.9	−0.50	−.0002	−.0169	−.0269	.2423	.2706	.3407
100.	10.9	−0.25	−.0018	−.0163	−.0218	.2306	.2535	.3927
100.	10.9	0.00	−.0091	−.0123	−.0102	.2029	.2272	.4627
100.	10.9	0.25	−.0134	−.0094	−.0029	.1789	.1891	.5049
100.	10.9	0.50	−.0135	−.0083	.0057	.1412	.1451	.5032
100.	10.9	0.90	−.0091	−.0055	.0192	.0516	.0558	.3820
254.	6.6	−0.90	.0395	−.0155	−.0215	.0851	.0983	.1734
254.	6.6	−0.50	.0033	−.0158	.0105	.1038	.1109	.1321
254.	6.6	−0.25	−.0044	−.0132	.0280	.1032	.1079	.1415
254.	6.6	0.00	−.0052	−.0076	.0458	.0949	.1002	.2061
254.	6.6	0.25	−.0056	−.0060	.0646	.0862	.0889	.2697
254.	6.6	0.50	−.0067	−.0041	.0854	.0727	.0708	.2978
254.	6.6	0.90	−.0058	−.0019	.0661	.0296	.0286	.2211
400.	2.0	−0.25	−.0012	−.0014	.0013	.0442	.0442	.0573
400.	2.0	0.25	−.0009	−.0015	−.0012	.0434	.0444	.0600
400.	6.0	0.00	.0010	.0007	−.0053	.0827	.0874	.1352
400.	10.0	−0.90	−.0066	−.0057	−.0077	.1417	.1538	.1452
400.	10.0	0.90	−.0013	−.0003	.0027	.0194	.0217	.1366
Column averages of absolute values			.0096	.0141	.0456	.1307	.1408	.3456

sample efficiency. For example, on average, the RMSE corresponding to $\hat{\rho}_{OLS}$ is more than twice as large as those of $\hat{\rho}_{NLS}$ and $\hat{\rho}_{QML}$. For this reason, we will henceforth focus attention only on $\hat{\rho}_{NLS}$ and $\hat{\rho}_{QML}$.

Observations concerning the response of the RMSEs to the sample size N, the average number of neighbors J of the weighting matrix M, and the value of ρ are not readily apparent from Table 1. For this reason, we describe the general results in the table via response functions. In doing this, we estimate separate response functions for $\hat{\rho}_{NLS}$ and $\hat{\rho}_{QML}$ for each of the three distributions considered. These six functions have the same form but different parameters. These response functions describe the results in Table 1 and should be useful for inferring corresponding results for experiments that have "similar" sets of parameter values.

TABLE 1C
BIASES AND RMSE OF ESTIMATORS FOR ρ, CONTAMINATED ERROR DISTRIBUTION

			Bias			RMSE		
N	J	ρ	$\hat{\rho}_{QML}$	$\hat{\rho}_{NLS}$	$\hat{\rho}_{OLS}$	$\hat{\rho}_{QML}$	$\hat{\rho}_{NLS}$	$\hat{\rho}_{OLS}$
49.	2.0	−0.90	.0021	.0005	.0094	.0257	.0268	0.1630
49.	2.0	0.90	.0005	.0016	.0031	.0252	.0273	0.1667
49.	6.0	−0.50	.0148	.0035	.0049	.1947	.2219	0.2329
49.	6.0	0.50	−.0020	.0012	.0120	.1183	.1206	0.3708
49.	10.0	0.00	.0023	−.0042	.0605	.2289	.2505	0.5745
58.	3.8	−0.90	.0563	−.0215	−.2058	.1679	.1218	0.7515
58.	3.8	−0.50	.0377	−.0395	−.0112	.1707	.1637	0.3739
58.	3.8	−0.25	.0224	−.0388	.0520	.1370	.1677	0.3178
58.	3.8	0.00	−.0035	−.0319	.0892	.1203	.1687	0.5490
58.	3.8	0.25	−.0117	−.0261	.2298	.1345	.1596	0.9347
58.	3.8	0.50	−.0110	−.0171	.3210	.1271	.1345	1.2607
58.	3.8	0.90	−.0040	−.0058	.2363	.0508	.0524	0.7936
100.	2.0	0.00	−.0034	−.0020	−.0023	.0582	.0582	0.0621
100.	6.0	−0.50	.0051	.0081	.0006	.1493	.1514	0.1615
100.	6.0	0.50	.0035	.0043	.0060	.0832	.0851	0.2636
100.	10.0	−0.25	.0122	.0016	.0210	.1984	.1975	0.3424
100.	10.0	0.25	.0013	.0017	.0479	.1384	.1403	0.3728
100.	10.9	−0.90	.0358	.0084	.0078	.2505	.2614	0.2890
100.	10.9	−0.50	.0301	.0056	.0095	.2313	.2390	0.3676
100.	10.9	−0.25	.0232	.0040	.0277	.2158	.2193	0.4303
100.	10.9	0.00	.0173	.0032	.0463	.1871	.1952	0.4862
100.	10.9	0.25	.0104	.0035	.0547	.1635	.1650	0.5121
100.	10.9	0.50	−.0009	.0023	.0654	.1264	.1264	0.5069
100.	10.9	0.90	−.0018	−.0004	.0713	.0455	.0484	0.3719
254.	6.6	−0.90	.0458	−.0132	−.0256	.0872	.0961	0.1961
254.	6.6	−0.50	.0086	−.0102	.0148	.1012	.1069	0.1289
254.	6.6	−0.25	.0007	−.0075	.0338	.0891	.1046	0.1685
254.	6.6	0.00	.0006	−.0075	.0598	.0846	.1037	0.2428
254.	6.6	0.25	−.0035	−.0046	.0981	.0838	.0921	0.3248
254.	6.6	0.50	−.0062	−.0041	.1165	.0737	.0760	0.3618
254.	6.6	0.90	−.0055	−.0022	.0887	.0290	.0315	0.2571
400.	2.0	−0.25	.0000	−.0002	−.0019	.0369	.0360	0.0490
400.	2.0	0.25	−.0002	.0001	.0024	.0363	.0361	0.0535
400.	6.0	0.00	−.0021	−.0031	−.0008	.0791	.0775	0.1252
400.	10.0	−0.90	.0004	−.0049	.0026	.1356	.1419	0.1341
400.	10.0	0.90	−.0011	−.0013	.0033	.0177	.0195	0.1232
Column averages of absolute values			.0108	.0082	.0568	.1168	.1229	.3561

Let $s = 1, \ldots, 36$ denote the sth case considered in Table 1 corresponding to a particular distribution. Using evident notation, we then specify the response functions for the RMSE of $\hat{\rho} = \hat{\rho}_{QML}$ or $\hat{\rho} = \hat{\rho}_{NLS}$ for a particular distribution as follows:

$$(9) \qquad RMSE(\hat{\rho}_s \mid N_s, J_s, \rho_s) = N_s^{-1/2} \exp\left[a_1 + a_2(1/J_s) + a_3 \rho_s + a_4(\rho_s/J_s) + a_5 \rho_s^2\right]$$

where $a_1, \ldots, a_5$ are parameters to be estimated using the data from Table 1 on the corresponding 36 cases. We estimate $a_1, \ldots, a_5$ by least squares (taking logs on both sides).

A few points concerning the response function in Eq. (9) should be noted. First, rather than being empirically determined, the exponent of the sample size is taken as $-\frac{1}{2}$ because of evident large sample considerations. Second, the function in Eq. (9) is relatively simple but yet nonnegative and able to accommodate certain patterns that might be suggested from time series considerations. For example, for an AR(1) model (with autocorrelation coefficient ρ), the variance in the asymptotic distribution of the (quasi) maximum likelihood estimator for ρ is, under typical assumptions, proportional to $1-\rho^2$.[19] It should be noted that this variance is symmetric about, and maximized at, zero; in addition, it approaches its minimum value as ρ approaches the "critical" points ± 1. Although the spatial models considered in our Monte Carlo study are not identical to an AR(1) model, one might nevertheless expect the relationship between the RMSE and the parameter ρ to peak at some point and then decline as ρ approaches "critical" points at which $I-\rho M$ is singular. For all the spatial weighting matrices considered in our experiments, the smallest positive critical point is 1.0; however, the largest negative critical point is equal to -1 only for the case in which $J=2$; for all other cases considered, the largest negative critical point is less than -1. In allowing for an interaction term between ρ and $1/J$ in Eq. (9), our response function permits *a priori* that the RMSEs might peak at a value of ρ that varies with J. There is also another avenue by which J might affect the RMSEs. Specifically, recall that $u_i = \rho \bar{u}_i + \epsilon_i$. The weighting matrices considered in the experiments are such that $\bar{u}_i$ is a straight average of the disturbances that correspond to the "neighbors" of the ith region. Because of this, the variance of $\bar{u}_i$ (relative to that of u_i) should be inversely related to J, the average number of neighbors. *Ceteris paribus*, one might expect large values of J to be associated with large RMSEs because estimation efficiency is typically an increasing function of regressor variances. Finally, other forms of the response functions were considered but were dominated by the form in Eq. (9).

The estimation results for the six response functions are given in Table A in the Appendix. Overall, the results in that table suggest that the response functions fit the data well. The R^2 values and the t-ratios are all quite high, suggesting both a tight fit and that each term considered is important. For all cases considered, the estimated value of a_5 is negative, and so each function peaks at a given value of ρ and then declines. The estimates of the coefficients are such that this "maximizing" value of ρ declines as J increases. For all cases considered, if $J > 2$ the value of ρ at which each function peaks is negative but greater than $-.25$. For all cases in which $J=2$, this "maximizing" value of ρ is very close to zero, namely, between $-.03$ and .04. The estimated coefficients are also such that increases in J are, again in all cases, associated with increases in the RMSEs. These results are consistent with prior notions. Graphs of the estimated response functions for the case of a normal error distribution are given in Figures 1 and 2 for $\hat{\rho}_{QML}$ and $\hat{\rho}_{NLS}$. Of course, in this case the (quasi) ML estimator $\hat{\rho}_{QML}$ is the ML estimator. The graphs for the case of a log-normal and contaminated error distribution are similar but are not given here to conserve space.

[19] See, for example, Johnston (1984, p. 329).

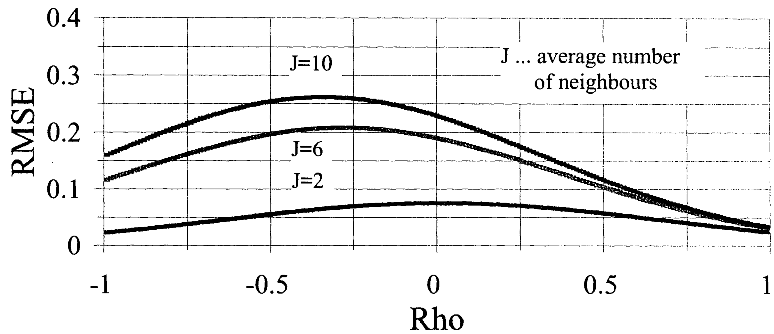

FIGURE 1

RMSE OF QML ESTIMATOR (NORMAL DISTRIBUTION, $N = 100$)

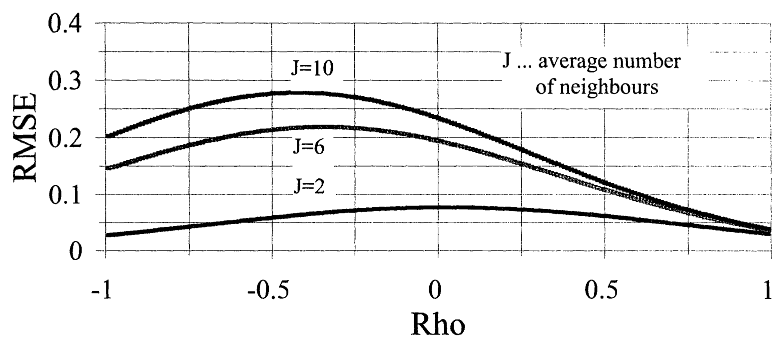

FIGURE 2

RMSE OF NLS ESTIMATOR (NORMAL DISTRIBUTION, $N = 100$)

The Monte Carlo results reported here correspond to the case in which the disturbances u_i are observable. We also performed corresponding experiments involving estimated disturbances, but we do not report here the details of those experiments because of space limitations. Those experiments suggest that the statements based on Table 1 and Table A in the Appendix concerning the relative efficiency of the three estimators carry over qualitatively to cases in which ρ is estimated from estimated disturbances.

6. SUGGESTIONS FOR FURTHER WORK

The autocorrelation model considered in this paper is sometimes referred to as a *spatial autoregressive model of order one* in that only one "spatial lag" of the disturbance term, represented by ρMu in Eq. (1), is being considered. Higher-order spatial models involving more than one spatial lag of the disturbance term (for example, using evident notation, $\rho_1 M_1 u + \cdots + \rho_p M_p u$) as well as of the innovation term (for example, $\epsilon + \rho_{p+1} M_{p+1}\epsilon + \cdots + \rho_{p+q} M_{p+q}\epsilon$) also have been considered in the literature. It should be of interest to extend the generalized moments approach suggested in this paper to those models and to determine corresponding large sample properties.

APPENDIX

In proving Theorem 1, we have to consider the following moments:

$$
\text{(A.1)}\qquad
\begin{aligned}
\vartheta_{1,N} &= N^{-1}u'u = N^{-1}\epsilon'(C_{1,N})\epsilon, & C_{1,N} &= P'P\\
\vartheta_{2,N} &= N^{-1}u'\bar{u} = N^{-1}\epsilon'(C_{2,N})\epsilon, & C_{2,N} &= P'MP\\
\vartheta_{3,N} &= N^{-1}\bar{u}'\bar{u} = N^{-1}\epsilon'(C_{3,N})\epsilon, & C_{3,N} &= P'M'MP\\
\vartheta_{4,N} &= N^{-1}\bar{\bar{u}}'\bar{u} = N^{-1}\epsilon'(C_{4,N})\epsilon, & C_{4,N} &= P'(M')^2MP\\
\vartheta_{5,N} &= N^{-1}\bar{\bar{u}}'\bar{\bar{u}} = N^{-1}\epsilon'(C_{5,N})\epsilon, & C_{5,N} &= P'(M')^2M^2P\\
\vartheta_{6,N} &= N^{-1}u'\bar{\bar{u}} = N^{-1}\epsilon'(C_{6,N})\epsilon, & C_{6,N} &= P'M^2P\\
\vartheta_{7,N} &= N^{-1}\bar{\epsilon}'\bar{\epsilon} = N^{-1}\epsilon'(C_{7,N})\epsilon, & C_{7,N} &= M'M\\
\vartheta_{8,N} &= N^{-1}\bar{\epsilon}'\epsilon = N^{-1}\epsilon'(C_{8,N})\epsilon, & C_{8,N} &= M'
\end{aligned}
$$

The corresponding moments based on $\tilde{u}$, $\bar{\tilde{u}}$ and $\bar{\bar{\tilde{u}}}$ in place of, respectively, u, $\bar{u}$ and $\bar{\bar{u}}$ will be denoted by $\tilde{\vartheta}_{h,N}$, $h = 1,\ldots,6$. In the following we will suppress the subscript N for the matrices $C_{h,N}$ and their elements, $h = 1,\ldots,8$. To prove Theorem 1, we need several lemmata.

LEMMA 1. *Under Assumption* 3 *the elements of the matrices* C_h *defined in Eq.* (*A*.1) *have the following properties,* $h = 1,\ldots,8$: $\sum_{i=1}^{N}|c_{h,ij}| \le c$, $\sum_{j=1}^{N}|c_{h,ij}| \le c$ *for all* $N \ge 1$ *and* $1 \le i, j \le N$ *for some* $0 < c < \infty$. *Furthermore,* $N^{-2}\sum_{i=1}^{N}\sum_{j=1}^{N}(c_{h,ij} + c_{h,ji})^2 = o(1)$.

PROOF. The first claim follows because by Assumption 3 the row and column sums of the absolute values of the elements of the matrices P and M are bounded, and this property is preserved under matrix multiplication.[20] Next, observe that the row and column sums of the absolute values of the elements of the matrices $C_h + C_h'$ and $[C_h + C_h'][C_h + C_h']$ are then bounded by $2c$ and $4c^2$, respectively. The second claim of the lemma now follows because $N^{-2}\sum_{i=1}^{N}\sum_{j=1}^{N}(c_{h,ij} + c_{h,ji})^2 = N^{-2}Tr\{[C_h + C_h'][C_h + C_h']\} \le 4c^2/N \to 0$ as $N \to \infty$. □

LEMMA 2. *Under Assumptions* 1 *to* 3, *the moments* $\vartheta_{h,N}$ *have the following properties,* $h = 1,\ldots,8$: $E\vartheta_{h,N} = O(1)$ *and* $\mathrm{var}(\vartheta_{h,N}) = o(1)$, *and hence* $\vartheta_{h,N} - E\vartheta_{h,N} \xrightarrow{p} 0$ *as* $N \to \infty$, *and* $\vartheta_{h,N} = O_p(1)$.

[20] To see this, consider matrices $A_N = (a_{ij,N})$, $B_N = (b_{ij,N})$, and $D_N = (d_{ij,N}) = A_N B_N$. Suppose $\sum_{i=1}^{N}|a_{ij,N}| \le c_a$, $\sum_{j=1}^{N}|a_{ij,N}| \le c_a$, $\sum_{i=1}^{N}|b_{ij,N}| \le c_b$, $\sum_{j=1}^{N}|b_{ij,N}| \le c_b$. Then $\sum_{i=1}^{N}|d_{ij,N}| \le \sum_{i=1}^{N}\sum_{k=1}^{N}|a_{ik,N}||b_{kj,N}| = \sum_{k=1}^{N}|b_{kj,N}|\sum_{i=1}^{N}|a_{ik,N}| \le c_a c_b$. Similarly, $\sum_{j=1}^{N}|d_{ij,N}| \le c_a c_b$.

PROOF. By Lemma 1, all elements $c_{h,ji}$ are bounded in absolute value. Hence $E\vartheta_{h,N} = \sigma^2 N^{-1}\sum_{i=1}^N c_{h,ii} = O(1)$. Observe further that $\mathrm{var}(\vartheta_{h,N}) = N^{-2}[(\mu_4 - \sigma^4)\sum_{i=1}^N c_{h,ii}^2 + \sigma^4 \sum_{i=1}^{N-1}\sum_{j=i+1}^N (c_{h,ij} + c_{h,ji})^2]$ with $\mu_4 = E\epsilon_i^4$, since $\mathrm{cov}(\epsilon_i \epsilon_j, \epsilon_r \epsilon_s) = 0$ unless $i = r$ and $j = s$, or $i = s$ and $j = r$. Clearly, a sufficient condition for $\mathrm{var}(\vartheta_{h,N}) = o(1)$ is that $N^{-2}\sum_{i=1}^N \sum_{j=1}^N (c_{h,ij} + c_{h,ji})^2 = o(1)$, which holds in light of Lemma 1. The last two claims follow from Chebychev's inequality and, for example, Corollary 5.1.1.2 in Fuller (1976, p. 186), respectively. □

LEMMA 3. *Consider random variables $v_{i,N}$, $w_{i,N}$, $\tilde{v}_{i,N}$, and $\tilde{w}_{i,N}$ and assume that*

$$(\text{A.2}) \qquad |\tilde{v}_{i,N} - v_{i,N}| \leq D_{iN}^v \tau_N^v \qquad |\tilde{w}_{i,N} - w_{i,N}| \leq D_{iN}^w \tau_N^w$$

where D_{iN}^v, D_{iN}^w, τ_N^v, and τ_N^w are, respectively, nonnegative random variables with $N^{-1}\sum_{i=1}^N (D_{iN}^v)^2 = O_p(1)$, $N^{-1}\sum_{i=1}^N (D_{iN}^w)^2 = O_p(1)$, $\tau_N^v = o_p(1)$, $\tau_N^w = o_p(1)$. Suppose furthermore that $N^{-1}\sum_{i=1}^N v_{i,N}^2 = O_p(1)$ and $N^{-1}\sum_{i=1}^N w_{i,N}^2 = O_p(1)$. Then $N^{-1}\sum_{i=1}^N \tilde{v}_{i,N}\tilde{w}_{i,N} - N^{-1}\sum_{i=1}^N v_{i,N} w_{i,N} \xrightarrow{p} 0$ as $N \to \infty$.

PROOF. Observe that

$$\left| N^{-1}\sum_{i=1}^N \tilde{v}_{i,N}\tilde{w}_{i,N} - N^{-1}\sum_{i=1}^N v_{i,N} w_{i,N} \right|$$

$$\leq N^{-1}\sum_{i=1}^N |\tilde{v}_{i,N} - v_{i,N}|\,|w_{i,N}| + N^{-1}\sum_{i=1}^N |\tilde{w}_{i,N} - w_{i,N}|\,|v_{i,N}|$$

$$+ N^{-1}\sum_{i=1}^N |\tilde{v}_{i,N} - v_{i,N}|\,|\tilde{w}_{i,N} - w_{i,N}|$$

$$\leq \left[N^{-1}\sum_{i=1}^N (D_{iN}^v)^2\right]^{1/2}\left[N^{-1}\sum_{i=1}^N w_{i,N}^2\right]^{1/2}\tau_N^v$$

$$+ \left[N^{-1}\sum_{i=1}^N (D_{iN}^w)^2\right]^{1/2}\left[N^{-1}\sum_{i=1}^N v_{i,N}^2\right]^{1/2}\tau_N^w$$

$$+ \left[N^{-1}\sum_{i=1}^N (D_{iN}^v)^2\right]^{1/2}\left[N^{-1}\sum_{i=1}^N (D_{iN}^w)^2\right]^{1/2}\tau_N^v \tau_N^w$$

The last inequality follows from Eq. (A.2) and Hölder's inequality. Since $\tau_N^v = o_p(1)$ and $\tau_N^w = o_p(1)$, the claim in the lemma follows by observing that all other terms are bounded in probability. □

LEMMA 4. *Under Assumptions* 1 *to* 4, $\tilde{\vartheta}_{h,N} - \vartheta_{h,N} \xrightarrow{p} 0$ *as* $N \to \infty$ *for* $h = 1, \ldots, 6$.

PROOF. To prove the lemma, it suffices to show, in light of Lemma 3, that u_i, $\bar{u}_i$, $\bar{\bar{u}}_i$, and the $\tilde{u}_i$, $\tilde{\bar{u}}_i$, $\tilde{\bar{\bar{u}}}_i$ satisfy the properties maintained for $v_{i,N}$ and $\tilde{v}_{i,N}$ in that lemma. First, observe that by Lemma 2, $N^{-1}\sum_{i=1}^N u_i^2 = O_p(1)$, $N^{-1}\sum_{i=1}^N \bar{u}_i^2 = O_p(1)$, and $N^{-1}\Sigma_{i=1}^{N}\bar{\bar{u}}_i^2 = O_p(1)$. Next, observe that by Assumption 4 we have $|\tilde{u}_i - u_i| \leq \|d_{iN}\|\,\|\Delta_N\|$ with $N^{-1}\Sigma_{i=1}^{N}\|d_{iN}\|^{2+\delta} = O_p(1)$ for some $\delta > 0$ and $N^{1/2}\|\Delta_N\| = O_p(1)$. Since $N^{-1}\Sigma_{i=1}^{N}\|d_{iN}\|^2 \leq [N^{-1}\Sigma_{i=1}^{N}\|d_{iN}\|^{2+\delta}]^{2/(2+\delta)}$ by Lyapunov's inequality, u_i and $\tilde{u}_i$ clearly satisfy the properties maintained for $v_{i,N}$ and $\tilde{v}_{i,N}$ in Lemma 3. Next, observe that

$$\sum_{j=1}^{N}|m_{ij}|^p = c_m^{p-1}\sum_{j=1}^{N}|m_{ij}|\big[|m_{ij}|/c_m\big]^{p-1} \leq c_m^{p-1}\sum_{j=1}^{N}|m_{ij}| \leq c_m^p$$

and that

$$\tilde{\bar{u}}_i = \sum_{j=1}^{N} m_{ij}\tilde{u}_j = \bar{u}_i + \sum_{j=1}^{N} m_{ij}(\tilde{u}_j - u_j)$$

and

$$\tilde{\bar{\bar{u}}}_i = \sum_{j=1}^{N} m_{ij}\tilde{\bar{u}}_j = \bar{\bar{u}}_i + \sum_{j=1}^{N} m_{ij}\sum_{s=1}^{N} m_{js}(\tilde{u}_s - u_s)$$

Hence, using the triangle and Hölder's inequalities with $q = 2 + \delta$ and $(1/q) + (1/p) = 1$,

$$|\tilde{\bar{u}}_i - \bar{u}_i| \leq \sum_{j=1}^{N}|m_{ij}|\,\|d_{jN}\|\,\|\Delta_N\|$$

$$\leq \left[\sum_{j=1}^{N}|m_{ij}|^p\right]^{1/p}\left[\sum_{j=1}^{N}\|d_{jN}\|^q\right]^{1/q}\|\Delta_N\| \leq \bar{D}_N\bar{\tau}_N$$

$$|\tilde{\bar{\bar{u}}}_i - \bar{\bar{u}}_i| \leq \sum_{j=1}^{N}|m_{ij}|\sum_{s=1}^{N}|m_{js}|\,\|d_{sN}\|\,\|\Delta_N\|$$

$$\leq \sum_{j=1}^{N}|m_{ij}|\left[\sum_{s=1}^{N}|m_{js}|^p\right]^{1/p}\left[\sum_{s=1}^{N}\|d_{sN}\|^q\right]^{1/q}\|\Delta_N\| \leq \bar{\bar{D}}_N\bar{\tau}_N$$

with $\bar{D}_N = c_m[N^{-1}\Sigma_{j=1}^{N}\|d_{jN}\|^q]^{1/q}$, $\bar{\bar{D}}_N = c_m^2[N^{-1}\Sigma_{j=1}^{N}\|d_{jN}\|^q]^{1/q}$ and $\bar{\tau}_N = N^{1/q}\|\Delta_N\|$. By Assumption 4, $\bar{D}_N = O_p(1)$, $\bar{\bar{D}}_N = O_p(1)$ and $\bar{\tau}_N = o_p(1)$. Hence $\bar{u}_i$ and $\tilde{\bar{u}}_i$ as well as $\bar{\bar{u}}_i$ and $\tilde{\bar{\bar{u}}}_i$ also satisfy the properties maintained for $v_{i,N}$ and $\tilde{v}_{i,N}$ in Lemma 3, and thus the claims of Lemma 4 follow from Lemma 3. □

PROOF OF THEOREM 1. The existence and measurability of $\hat{\rho}_{NLS,N}$ and $\hat{\sigma}_{NLS,N}^2$ are ensured by, for example, Lemma 2 in Jennrich (1969). The objective function of the nonlinear least squares estimator and its corresponding nonstochastic counterpart are given by, respectively,

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

$$\bar{R}_N(\underline{\theta}) = \big[\Gamma_N(\underline{\rho},\underline{\rho}^2,\underline{\sigma}^2)' - \gamma_N\big]'\big[\Gamma_N(\underline{\rho},\underline{\rho}^2,\underline{\sigma}^2)' - \gamma_N\big]$$

where $\underline{\theta} = (\underline{\rho}, \underline{\sigma}^2)'$. To prove the consistency of $(\hat{\rho}_{NLS,N}, \hat{\sigma}^2_{NLS,N})$, we show that the conditions of, for example, Lemma 3.1 in Pötscher and Prucha (1991a) are satisfied for the problem at hand. We first show that $\theta = (\rho, \sigma^2)'$ is identifiably unique [where $\theta = (\rho, \sigma^2)'$ denotes the vector of true parameters]. Observe that because of Eq. (5),

$$\begin{aligned}
&\overline{R}_N(\underline{\theta}) - \overline{R}_N(\theta) \\
&\quad = \left[\underline{\rho} - \rho, \underline{\rho}^2 - \rho^2, \underline{\sigma}^2 - \sigma^2\right]\Gamma_N'\Gamma_N\left[\underline{\rho} - \rho, \underline{\rho}^2 - \rho^2, \underline{\sigma}^2 - \sigma^2\right]' \\
&\quad \geq \lambda_{\min}(\Gamma_N'\Gamma_N)\left[\underline{\rho} - \rho, \underline{\rho}^2 - \rho^2, \underline{\sigma}^2 - \sigma^2\right]\left[\underline{\rho} - \rho, \underline{\rho}^2 - \rho^2, \underline{\sigma}^2 - \sigma^2\right]' \\
&\quad \geq \lambda_*\left[\underline{\rho} - \rho, \underline{\sigma}^2 - \sigma^2\right]\left[\underline{\rho} - \rho, \underline{\sigma}^2 - \sigma^2\right]' = \lambda_*\|\underline{\theta} - \theta\|^2
\end{aligned}$$

Hence for every $\epsilon > 0$ and any N, we have

$$\inf_{\{\underline{\theta}:\|\underline{\theta} - \theta\| \geq \epsilon\}} \left[\overline{R}_N(\underline{\theta}) - \overline{R}_N(\theta)\right] \geq \inf_{\{\underline{\theta}:\|\underline{\theta} - \theta\| \geq \epsilon\}} \lambda_*\|\underline{\theta} - \theta\|^2 = \lambda_*\epsilon^2 > 0$$

which proves that θ is identifiably unique. Next, let $F_N = [G_N, -g_N]$ and $\Phi_N = [\Gamma_N, -\gamma_N]$; then for $\rho \in [a, a]$ and $\sigma^2 \in [0, b]$.

$$\begin{aligned}
|R_N(\underline{\theta}) - \overline{R}_N(\underline{\theta})| &= |\left[\underline{\rho}, \underline{\rho}^2, \underline{\sigma}^2, 1\right][F_N'F_N - \Phi_N'\Phi_N]\left[\underline{\rho}, \underline{\rho}^2, \underline{\sigma}^2, 1\right]'| \\
&\leq \|F_N'F_N - \Phi_N'\Phi_N\|\|\underline{\rho}, \underline{\rho}^2, \underline{\sigma}^2, 1\|^2 \\
&\leq \|F_N'F_N - \Phi_N'\Phi_N\|[1 + a^2 + a^4 + b^2]
\end{aligned}$$

Since Lemmata 2 and 4 imply that $F_N - \Phi_N \xrightarrow{p} 0$ and that the elements of F_N and Φ_N are, respectively, $O_p(1)$ and $O(1)$, it follows that $R_N(\underline{\theta}) - \overline{R}_N(\underline{\theta})$ converge to zero uniformly over the (extended) parameter space, that is,

$$\sup_{\underline{\rho} \in [-a, a], \underline{\sigma}^2 \in [0, b]} |R_N(\underline{\theta}) - \overline{R}_N(\underline{\theta})| \leq \|F_N'F_N - \Phi_N'\Phi_N\|[1 + a^2 + a^4 + b^2] \xrightarrow{p} 0$$

as $N \to \infty$. The consistency of $(\hat{\rho}_{NLS,N}, \hat{\sigma}^2_{NLS,N})$ now follows directly from Lemma 3.1 in Pötscher and Prucha (1991a). □

The proof of Theorem 2 requires a central limit theorem (CLT) for triangular arrays. The CLT below follows readily from a corollary to the Lindeberg-Feller CLT for triangular arrays using the Cramer-Wold device. That corollary is, for example, given in Billingsley (1979, p. 319, Problem 27.6).[21]

[21] More precisely, we use a slight generalization of that corollary where the assumption that the random variables of interest are constructed from a sequence of i.i.d. random variables is replaced by the assumption that they are constructed from a triangular array of identically distributed and (within each sample) independent random variables.

THEOREM A. *Let $\{v_{iN}, 1 \le i \le N, N \ge 1\}$ be a triangular array of random variables that are identically distributed, and for each N (jointly) independent with $Ev_{iN} = 0$ and $Ev_{iN}^2 = \sigma^2$, $0 < \sigma^2 < \infty$. Let $\{z_{ij,N}, 1 \le i \le N, N \ge 1\}$, $j = 1, \dots, K$, be triangular arrays of real numbers that are bounded in absolute value, that is, $c_z = \sup_N \sup_{i \le N, j \le K} |z_{ij,N}| < \infty$. Further, let $\{V_N : n \ge 1\}$ and $\{Z_N : n \ge 1\}$ with $V_N = (v_{iN})_{i=1,\dots,N}$ and $Z_N = (z_{ij,N})_{i=1,\dots,N; j=1,\dots,K}$ denote corresponding sequences of $N \times 1$ random vectors and $N \times K$ real matrices, respectively, and let $\lim_{N \to \infty} N^{-1} Z_N' Z_N = Q$ be finite and positive definite. Then $N^{-1/2} Z_N' V_N \xrightarrow{D} N(0, \sigma^2 Q)$.*

PROOF OF THEOREM 2. To prove part (a) of the theorem, observe that $N^{1/2}[\tilde{\beta}_N^G - \beta] = [N^{-1} Z'Z]^{-1} N^{-1/2} Z'\epsilon$, where $Z = (I - \rho M)X$. Note that, in general, the elements of Z will depend on the sample size N. However, for notational simplicity we will not denote this dependence explicitly in the following. Observe further that under the maintained assumptions the elements of Z are bounded in absolute value by $(1 + c_m)c_x$ and that $\lim_{N \to \infty} N^{-1} Z'Z = \underline{Q}_x(\rho)$ is finite and nonsingular. Recall that the innovations $\epsilon_1, \dots, \epsilon_N$ are identically distributed and for each N (jointly) independent with $Ev_i N = 0 \dots$ with mean zero and variance σ^2. Hence it follows from Theorem A that $N^{-1/2} Z'\epsilon \xrightarrow{D} N[0, \sigma^2 \underline{Q}_x(\rho)]$ and consequently $N^{1/2}[\tilde{\beta}_N^G - \beta] \xrightarrow{D} N[0, \sigma^2 \underline{Q}_x(\rho)^{-1}]$. Of course, this also implies that $\tilde{\beta}_N^G$ is consistent.

We prove part (b) of the theorem by showing that $N^{1/2}[\tilde{\beta}_N^G - \tilde{\beta}_N^{FG}] \xrightarrow{p} 0$ as $N \to \infty$. To prove this, it suffices to show that

$$\text{(A.3)} \qquad N^{-1} X' \left[\Omega(\tilde{\rho}_N)^{-1} - \Omega(\rho)^{-1} \right] X \xrightarrow{p} 0$$

and

$$\text{(A.4)} \qquad N^{-1/2} X' \left[\Omega(\tilde{\rho}_N)^{-1} - \Omega(\rho)^{-1} \right] u \xrightarrow{p} 0$$

Clearly, $\Omega(\tilde{\rho}_N)^{-1} - \Omega(\rho)^{-1} = (\rho - \tilde{\rho}_N)(M + M') + (\rho^2 - \tilde{\rho}_N^2) M'M$. Hence

$$\text{(A.5)} \qquad N^{-1} X' \left[\Omega(\tilde{\rho}_N)^{-1} - \Omega(\rho)^{-1} \right] X = (\rho - \tilde{\rho}_N) N^{-1} X'(M + M')X + \left(\rho^2 - \tilde{\rho}_N^2\right) N^{-1} X'M'MX$$

and

$$\text{(A.6)} \qquad N^{-1/2} X' \left[\Omega(\tilde{\rho}_N)^{-1} - \Omega(\rho)^{-1} \right] u = (\rho - \tilde{\rho}_N) N^{-1/2} X'(M + M')u + \left(\rho^2 - \tilde{\rho}_N^2\right) N^{-1/2} X'M'Mu$$

Under the maintained assumptions, the elements of $N^{-1} X'(M + M')X$ and $N^{-1} X'M'MX$ are bounded in absolute value by $2c_x^2 c_m$ and $c_x^2 c_m^2$, respectively (see footnote 20). Condition (A.3) then follows from (A.5), since $\tilde{\rho}_N$ is assumed to be consistent.

Next, consider the terms $N^{-1/2} X'M'u$, $N^{-1/2} X'Mu$, and $N^{-1/2} X'M'Mu$. Clearly, the expected value of each element of these vectors is zero. The variance-covariance matrices of these vectors are given by, respectively,

$$N^{-1} X' \Phi_s X \qquad s = 1, 2, 3,$$

with $\Phi_1 = M'PP'M$, $\Phi_2 = MPP'M'$, $\Phi_3 = M'MPP'M'M$. Since the row and column sums of the absolute values of the matrices P and M are bounded, it follows that the row and column sums of the matrices Φ_s are also bounded by some finite constants, say, c_s $(s = 1,2,3)$; see footnote 20. Since the elements of X are bounded in absolute value by c_x, it then follows that the elements of the variance covariance matrices $N^{-1}X'\Phi_s X$ are bounded in absolute value by $c_x^2 c_s < \infty$ $(s = 1,2,3)$. It then follows from, for example, Corollary 5.1.1.2 in Fuller (1976), that the elements of $N^{-1/2}X'M'u$, $N^{-1/2}X'Mu$, and $N^{-1/2}X'M'Mu$ are $O_p(1)$. Condition (A.4) is now seen to hold from (A.6) because $\tilde{\rho}_N$ is assumed to be consistent.

Part (c) of the theorem follows immediately from (A.3) and Assumption 6 and the fact that $\tilde{\sigma}^2$ is a consistent estimator for σ^2. □

Next, we prove that under Assumptions 1 to 3 and 5 and 6 the OLS estimator $\tilde{\beta}_N$ is $N^{1/2}$-consistent, as was claimed in the discussion after Theorem 2. Observe that $N^{1/2}[\tilde{\beta}_N - \beta] = [N^{-1}X'X]^{-1}N^{-1/2}Z'\epsilon$, with Z defined here as $Z = (I - \rho M')^{-1}X$. Note again that, in general, the elements of Z will depend on the sample size N. By assumption $\lim_{N\to\infty} N^{-1}Z'Z = \lim_{N\to\infty} N^{-1}X'\Omega(\rho)X = \overline{Q}_x(\rho)$ is finite and nonsingular, and the innovations $\epsilon_1, \ldots, \epsilon_N$ are identically distributed, and for each N (jointly) independent with mean zero and variance σ^2. Hence it follows from Theorem A that $N^{-1/2}Z'\epsilon \xrightarrow{D} N[0, \sigma^2\overline{Q}_x(\rho)]$. Observing that $Q_x = \lim_{N\to\infty} N^{-1}X'X$ is finite and nonsingular, it follows that $N^{1/2}[\tilde{\beta}_N - \beta] \xrightarrow{D} N[0, Q_x^{-1}\overline{Q}_x(\rho)Q_x^{-1}]$. □

The following table contains the estimation results for the response functions for $\hat{\rho}_{QML}$ and $\hat{\rho}_{NLS}$ discussed in Section 5.

TABLE A

THE RESPONSE FUNCTIONS FOR $\hat{\rho}_{QML}$ AND $\hat{\rho}_{NLS}$

Estimator	Parameter Estimates $\hat{a}_1$	$\hat{a}_2$	$\hat{a}_3$	$\hat{a}_4$	$\hat{a}_5$	R^{2*}
	Normal Error Distribution					
$\hat{\rho}_{QML}$	1.11015	−2.77568	−0.98016	2.03430	−1.14638	.91
	(16.88)	(11.49)	(9.46)	(4.92)	(11.53)	
$\hat{\rho}_{NLS}$	1.12976	−2.77091	−1.04065	2.17021	−0.97829	.93
	(20.10)	(13.42)	(11.75)	(6.14)	(11.51)	
	Log-Normal Error Distribution					
$\hat{\rho}_{QML}$	1.01024	−2.70593	−0.98919	1.85930	−0.97475	.92
	(15.25)	(11.12)	(9.48)	(4.46)	(9.74)	
$\hat{\rho}_{NLS}$	1.08550	−2.77321	−1.06518	2.19157	−0.96662	.94
	(18.99)	(13.21)	(11.83)	(6.10)	(11.19)	
	Contaminated Error Distribution					
$\hat{\rho}_{QML}$	0.93495	−3.07788	−1.04396	2.10490	−0.85946	.90
	(12.40)	(11.11)	(8.78)	(4.44)	(7.54)	
$\hat{\rho}_{NLS}$	1.01993	−3.13841	−1.04142	2.24336	−0.92679	.93
	(16.09)	(13.47)	(10.43)	(5.63)	(9.67)	

*The R^2 statistic is the square of the correlation coefficient between the RMSE and its response function prediction based on the values in Table 1. The numbers in parentheses are t-ratios (in absolute value).

REFERENCES

ANSELIN, L., *Spatial Econometrics: Methods and Models* (Boston: Kluwer Academic Publishers, 1988).

———, "Some Robust Approaches to Testing and Estimation in Spatial Econometrics," *Regional Science and Urban Economics* 20 (1990), 141–163.

——— AND R.J.G.M. FLORAX, "Small Sample Properties of Tests for Spatial Dependence in Regression Models: Some Further Results," in L. Anselin and R.J.G.M. Florax (eds.), *New Directions in Spatial Econometrics* (New York: Springer-Verlag, 1995), 21–74.

BENNETT, R. AND L. HORDIJK, "Regional Econometric and Dynamic Models," in P. Nijkamp (ed.), *Handbook of Regional and Urban Economics*, Vol. 1 (Amsterdam: North Holland, 1986), 407–441.

BILLINGSLEY, P., *Probability and Measure* (New York: Wiley, 1979).

CASE, A., "Spatial Patterns in Household Demand," *Econometrica* 59 (1991), 953–966.

CLIFF, A. AND J. ORD, *Spatial Autocorrelation* (London: Pion, 1973).

——— AND ———, *Spatial Process: Models and Applications* (London: Pion, 1981).

CONLEY, T., "Econometric Modeling of Cross Sectional Dependence," Ph.D. Dissertation, University of Chicago, 1996.

CRESSIE, N.A.C., *Statistics for Spatial Data* (New York: Wiley, 1993).

DELONG, J.B. AND L. SUMMERS, "Equipment Investment and Economic Growth," *Quarterly Journal of Economics* 106 (1991), 445–502.

DUBIN, R., "Estimation of Regression Coefficients in the Presence of Spatially Autocorrelation Error Terms," *Review of Economics and Statistics* 70 (1988), 466–474.

FULLER, W., *Introduction to Statistical Time Series* (New York: Wiley, 1976).

GALLANT, R. AND H. WHITE, *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models* (Oxford: Oxford University Press, 1988).

HEIJMANS, R. AND J. MAGNUS, "Consistent Maximum Likelihood Estimation with Dependent Observations," *Journal of Econometrics* 32 (1986), 253–285.

——— AND ———, "Corrigenda," *Journal of Econometrics* 35 (1987), 395.

JENNRICH, R., "Asymptotic Properties of Non-Linear Least Squares Estimators," *Annals of Mathematical Statistics* 40 (1969), 633–643.

JOHNSTON, J., *Econometric Methods* (New York: McGraw-Hill, 1984).

KELEJIAN, H. AND D. ROBINSON, "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 (1993), 297–312.

——— AND ———, "Spatial Correlation: A Suggested Alternative to the Autoregressive Model," in L. Anselin and R.J.G.M. Florax, (eds.), *New Directions in Spatial Econometrics* (New York: Springer-Verlag, 1995), 75–95.

MANDY, D. AND C. MARTINS-FILHO, "A Unified Approach to Asymptotic Equivalence of Aitken and Feasible Aitken Instrumental Variable Estimators," *International Economic Review* 35 (1994), 957–980.

MARDIA, K.V. AND R.J. MARSHALL, "Maximum Likelihood Estimation of Models for Residual Covariance in Spatial Regression," *Biometrika* 71 (1984), 135–146.

MOULTON, B.R., "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units," *Review of Economics and Statistics* 72 (1990), 334–338.

ORD, J., "Estimation Methods for Models of Spatial Interaction," *Journal of the American Statistical Association* 70 (1975), 120–126.

PÖTSCHER, B. AND I. PRUCHA, "A Class of Partially Adaptive One-Step *M*-Estimators for the Nonlinear Regression Model with Dependent Observations," *Journal of Econometrics* 32 (1986), 219–251.

——— AND ———, "A Uniform Law of Large Numbers for Dependent and Heterogeneous Data Processes," *Econometrica* 57 (1989), 675–683.

——— AND ———, "Basic Structure of Asymptotic Theory in Dynamic Nonlinear Models: I. Consistency and Approximation Concepts," *Econometric Reviews* 10 (1991a), 125–216.

——— AND ———, "Basic Structure of Asymptotic Theory in Dynamic Nonlinear Models: II. Asymptotic Normality," *Econometric Reviews* 10 (1991b), 253–325.

QUAH, D., "International Patterns of Growth: I. Persistence in Cross-Country Disparities," Working Paper, Department of Economics, London School of Economics, 1992.
TOPA, G., "Social Interactions, Local Spillovers, and Unemployment," Ph.D. Dissertation, University of Chicago, 1996.
WALD, A., "A Note on the Consistency of the Maximum Likelihood Estimate," *Annals of Mathematical Statistics* 20 (1949), 595–600.
WHITTLE, P., "On Stationary Processes in the Plane," *Biometrica* 41 (1954), 434–449.