

On the inconsistency of the OLS estimator for spatial autoregressive models

Théophile Azomahou* and Agénor Lahatte

Université Louis Pasteur, Strasbourg I

Bureau d'Économie Théorique et Appliquée (BETA-Theme)

Abstract

We investigate the asymptotic bias of the ordinary least squares estimator for spatial autoregressive models. We show that this estimator is biased as well as inconsistent for the parameters regardless of the distribution of the disturbance. Illustrative examples are provided.

Key words: Spatial processes; consistency of OLS estimator

AMS Classification: 62F12; 62M30

*Corresponding author. Théophile Azomahou; *Université Louis Pasteur*, Strasbourg, 61, avenue de la Forêt Noire, F-67085 Strasbourg Cedex; France, phone: (33) 390 414 071; fax: (33) 390 414 050; e-mail: azomahou@cournot.u-strasbg.fr

1 Introduction

Many studies deal with the spatial modeling of phenomena in a regression context. Recent examples of empirical work that explicitly incorporate spatial dependence concern the forecasting of cigarette demand using panel data, Baltagi and Li (1999), the study of responses of real wages to local and aggregate unemployment rates over time, Ziliak et al. (1999) and the estimation of a hedonic model for residential sales transactions, Bell and Bockstael (2000). From a theoretical side, Kelejian and Prucha (1998, 1999) and Conley (1999) developed new estimators based on Generalized Method of Moments (GMM) to deal respectively with generalized spatial models and lattice models.¹ Readers are also referred to Ripley (1981, 1988), Cressie (1991) and Sajjan (2000) for special topics on spatial processes.

Two structures of spatial dependence mostly encountered both in the statistics and econometrics literature are sources of statistical inference problems: spatial dependencies across observations for the response variable (dependent variable) and spatial autocorrelation in the error terms. Indeed, contrary to time-series models which are associated with uni-directional time flow, there is no natural order for arranging spatial data. Spatial data are supplemented with a multi-directional motion. This particular characteristic of spatial processes precludes a simple transposition of time-series methodologies. As outlined by Anselin (1988, p.59), the asymptotic properties of the ordinary least squares estimator for the model with spatial residual autocorrelation are more in line with the times-series analogue. Indeed, parameter

¹By general spatial model we mean a cross-sectional regression model for spatial data containing both the spatial lag of the response variable as additional regressor and a disturbance term that is spatially autoregressive.

estimates will still be only unbiased but inefficient due to the nondiagonal structure of the variance-covariance matrix of the error terms. In light of this, Kelejian and Prucha (1997) have shown that, in a spatial autoregressive errors specification, the attempt to use a two-stage least squares procedure based on the Cochrane-Orcutt (1949) transformation leads to inconsistent estimates. Kelejian and Prucha (1997) demonstrated that the response function of the spatial autoregressive model associated with the Cochrane-Orcutt transformation violated Amemiya's (1985, p.246) rank condition, raising identification issues for the spatial parameter.

In a regression context and from a methodological viewpoint, when a spatial lag of the response variable is used as additional regressor, the spatial dynamics involved induces the problem of endogeneity. Indeed, contrary to time-series models where the lagged term of the dependent variable is uncorrelated with the error if there is no serial residual correlation, in the spatial context, this correlation occurs regardless of the properties of the disturbance. As a result, the ordinary least squares estimator will be inconsistent. This inconsistency is mentioned in papers presenting alternative estimation procedures, see e.g., Anselin and Bera (1998) for a review of various available estimators. However, to the best of our knowledge, the details of the theoretical reasons of this inconsistency have not yet been published.

The purpose of this paper is to investigate the asymptotic properties of the ordinary least squares estimator for autoregressive spatial models. We show that this estimator is biased as well as inconsistent regardless of the distribution of the disturbance. We provide a detailed proof that the bias of the ordinary least squares estimator for autoregressive spatial models is

$O_p(1)$ and not zero.

The paper is organized as follows. Section 2 presents the model and some examples illustrating the issue. Section 3 is devoted to theoretical results. Section 4 concludes the study.

2 Model and examples

Without loss of generality, let us consider a linear regression model generated from a spatial stochastic process of the form

$$y_n = \theta W_n y_n + \varepsilon_n, \quad |\theta| < 1, \quad (1)$$

where y_n is a $n \times 1$ vector of observations for the response variable (dependent variable), $W_n y_n$ is a spatial regressor with W_n denoting a $n \times n$ spatial weighting matrix; that is to say, W_n expresses for each (row) observation those (columns) locations that belong to its neighborhood set as non-zero elements. Clearly, W_n is the matrix of a graph, Berge (1983). This definition will be illustrated in the examples below. θ is a spatial parameter (to be estimated) and ε_n is a $n \times 1$ vector of error terms. The specification (1) is widely used in empirical analysis and usually referred as to pure first order spatial autoregressive model by analogy with time-series.² Indeed, if we assume that $E(\varepsilon_n) = 0$, then the conditional expectation of y_n is $E(y_n | \cdot) = \theta W_n y_n$. The dependent variable can then be viewed, in part, to depend on a weighted average of the values it takes on neighboring cross-sectional units. For reasons of generality, we permit the elements of the matrices and vectors to depend on the number of observations n , that is to

²The model (1) may be viewed as the spatial analogue of time-series AR(1) processes, that is $y_t = \theta y_{t-1} + \varepsilon_t$.

form a triangular array. Thus, in the notations an index n denotes the size of the sample. This notation allows us to fix ideas on elements of the model which depend on n .

Despite its simplicity, this model captures all the effects of the presence of a spatially lagged response variable, and can therefore be used for our purpose. To ease notations, let $z_n := W_n y_n$. If we assume the errors to be i.i.d., the OLS estimator for θ denoted $\hat{\theta}$ is

$$\hat{\theta}_n = (z_n' z_n)^{-1} z_n' y_n. \quad (2)$$

Substituting the expression for y_n in (1) the population parameter from (2) leads to

$$\hat{\theta}_n = \theta_0 + (z_n' z_n)^{-1} z_n' \varepsilon_n. \quad (3)$$

Based on relation (3), the convergence of $\hat{\theta}$ towards the true value of θ denoted θ_0 requires the following conditions:³

$$\text{plim}_{n \rightarrow \infty} n^{-1} (z_n' z_n) = M_z, \quad (4)$$

with M_z a non-zero scalar and

$$\text{plim}_{n \rightarrow \infty} n^{-1} (z_n' \varepsilon_n) = 0. \quad (5)$$

In the spatial framework, condition (4) can be satisfied with suitable restrictions on the value of θ and on the structure of the spatial weighting matrix. But except in the trivial case where $\theta = 0$, condition (5) does not hold. The aim is to show that the limit in probability of the bias $E(\hat{\theta}_n) - \theta_0$ is not zero unless $\theta = 0$. To show this, we make use of the following assumptions.

³It should be noted that the class of distributions allowed for ε_n precludes the Dirac mass.

Assumption 1 *The $\varepsilon_{i,n}$ are i.i.d. with mean zero and variance $\sigma_\varepsilon^2 < \infty$. Moreover, there exists a finite non negative constant c_ε such that for all i , $1 \leq i \leq n$, $\forall n \geq 1$, the s -th absolute moments are such that $E|\varepsilon_{i,n}|^s \leq c_\varepsilon < \infty, \forall s > 0$.*

Assumption 2 *All diagonal elements $\omega_{ii,n}$ of the spatial weighting matrix W_n are zero and the off-diagonal elements $\omega_{ij,n}$ are such that $\sum_{i=1}^n |\omega_{ij,n}| \leq \kappa_\omega$ and $\sum_{j=1}^n |\omega_{ij,n}| \leq \kappa_\omega$ for all $j = 1, \dots, n; \forall n \geq 1$, where κ_ω is a finite constant.*

Assumption 3 *The matrix $(I - \theta W_n)$ is nonsingular and the following sums: $\sum_{i=1}^n |m_{ij,n}|$ and $\sum_{j=1}^n |m_{ij,n}|$ are bounded by, say $\kappa_m < \infty$ for all $j = 1, \dots, n; n \geq 1$, with $m_{ij,n}$ being an element of $M_n = (I - \theta W_n)^{-1}$, for all $|\theta| < 1$.*

Using the Markov inequality, it is easy to show that the second part of Assumption 1 implies that $n^{-1} \sum_{i=1}^n |\varepsilon_{i,n}|^s = O_p(1), \forall s > 0$. Assumption 2 allows us to restrict the extent of the spatial dependence. Under some conditions related to the eigenvalues of W_n and the restriction that $|\theta| < 1$, the invertibility of M_n is ensured as stated in Assumption 3.⁴ We give some examples to illustrate the issue.

Consider a spatial system constituted by four units $S = \{s_1, s_2, s_3, s_4\}$ located on a straight line and arranged in the increasing order of the indices. We can define the neighborhood sets V_{s_i} , with V_{s_i} indicating the neighbors of s_i , by $V_{s_1} = \{s_2\}$, $V_{s_2} = \{s_1, s_3\}$, $V_{s_3} = \{s_2, s_4\}$ and $V_{s_4} = \{s_3\}$.

⁴If all eigenvalues of W_n are less than or equal to one in absolute value, $|\theta| < 1$ implies that all eigenvalues of θW_n are strictly below one in absolute value, which ensures that $M_n = \sum_{i=0}^{\infty} \theta^i W_n^i$, Horn and Johnson (1985).

In the first example, we use as spatial weighting matrix the binary matrix of dimension 4×4 denoted $W_{1,n}$ of the graph associated with S . This yields a first order contiguity matrix

$$W_{1,n} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \quad (6)$$

Let ε_n be a vector of error terms with suitable dimension. The assumptions on ε_n and $W_{1,n}$ yield

$$E \left[\varepsilon_n' W_{1,n} (I - \theta W_{1,n})^{-1} \varepsilon_n \right] = \frac{1}{D} 2\sigma_\varepsilon^2 \theta (3 - 2\theta^2), \quad (7)$$

with $D = 1 - 3\theta^2 + \theta^4$. Given the maintained assumptions of the model, the second term of relation (7) is zero if and only if $\theta = 0$ (this term is zero for $\theta_1 = 0$ and $\theta_2 = \pm\sqrt{3/2}$, but the latter violates the assumption $|\theta| < 1$). For the second example, consider a row-standardized version of $W_{1,n}$ denoted $W_{2,n}$. We get

$$W_{2,n} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \quad (8)$$

Then,

$$E \left[\varepsilon_n' (I - \theta W_{2,n}')^{-1} W_{2,n}' \varepsilon_n \right] = \frac{1}{D} 2\sigma_\varepsilon^2 \theta (5/4 - \theta^2), \quad (9)$$

with $D = 1 - (5/4)\theta^2 + (1/4)\theta^4$. In this case also, we have $\theta_1 = 0$ and $\theta_2 = \pm\sqrt{5/4}$, and only the first solution is acceptable.

From these two examples, we note that the expectations are null only for $\theta = 0$, which suggests that condition (5) does not hold unless $\theta = 0$. Note that the above examples can be extended to the case of $(S)_{i=1}^n$ spatial units on a straight line and it can be shown easily that the same conclusion applies. We now provide the theoretical proof.

3 Bias properties

Theorem 1 *Given Assumptions 1–3, the bias $E(\hat{\theta}_n) - \theta_0$ is $O_p(1)$.*

Proof.

The proof is given in two steps. We show respectively that $\text{plim}_{n \rightarrow \infty} n^{-1}(z'_n z_n)$ and $\text{plim}_{n \rightarrow \infty} n^{-1}(z'_n \varepsilon_n)$ are $O_p(1)$.

$$(i) \text{plim}_{n \rightarrow \infty} n^{-1}(z'_n z_n) = O_p(1).$$

Note that the $z_{i,n}$'s are not independent. Then using the law of large numbers for dependent observations (see, e.g., White (1984, p.42)), $\text{plim}_{n \rightarrow \infty} n^{-1}(z'_n z_n) \xrightarrow{\text{p.s.}} E(z'_n z_n)$. Also, $n^{-1} \sum_{i=1}^n z_{i,n}^2 \xrightarrow{\text{p.s.}} E(z_n^2)$ and $E(n^{-1} \sum_{i=1}^n z_{i,n}^2) = E(z_n^2)$. Then,

$$\begin{aligned} E|n^{-1}(z'_n z_n)| &= E|n^{-1} \varepsilon'_n M'_n W'_n W_n M_n \varepsilon_n| \\ &= E \left| n^{-1} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n \sum_{k=1}^n \sum_{p=1}^n \varepsilon_{i,n} m_{ki,n} \omega_{lk,n} \omega_{lp,n} m_{pj,n} \varepsilon_{j,n} \right| \\ &\leq n^{-1} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n \sum_{k=1}^n \sum_{p=1}^n |m_{ki,n}| |\omega_{lk,n}| |\omega_{lp,n}| |m_{pj,n}| \\ &\qquad\qquad\qquad E|\varepsilon_{i,n}| |\varepsilon_{j,n}| =: \xi_1. \end{aligned}$$

From Assumption 1, $E|\varepsilon_{i,n}| |\varepsilon_{j,n}| \leq c_\varepsilon$ for $i = j$ and $E|\varepsilon_{i,n}| |\varepsilon_{j,n}| \leq c_\varepsilon^2$ for

$i \neq j$. Assuming that $1 \leq c_\varepsilon$ and by the Cauchy-Schwarz inequality we have

$$\xi_1 \leq n^{-1} c_\varepsilon^2 \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n \sum_{k=1}^n \sum_{p=1}^n |m_{ki,n}| |\omega_{lk,n}| |\omega_{lp,n}| |m_{pj,n}| =: \xi_2.$$

Observing that only $|m_{pj,n}|$ depends on j and using Assumption 3, it follows that

$$\begin{aligned} \xi_2 &= n^{-1} c_\varepsilon^2 \sum_{i=1}^n \sum_{l=1}^n \sum_{k=1}^n \sum_{p=1}^n |m_{ki,n}| |\omega_{lk,n}| |\omega_{lp,n}| \sum_{j=1}^n |m_{pj,n}| \\ &\leq n^{-1} c_\varepsilon^2 \kappa_m \sum_{i=1}^n \sum_{l=1}^n \sum_{k=1}^n \sum_{p=1}^n |m_{ki,n}| |\omega_{lk,n}| |\omega_{lp,n}| =: \xi_3, \end{aligned}$$

From ξ_3 , only $|\omega_{lp,n}|$ depends on p and using Assumption 2 we can write

$$\begin{aligned} \xi_3 &= n^{-1} c_\varepsilon^2 \kappa_m \sum_{i=1}^n \sum_{l=1}^n \sum_{k=1}^n |m_{ki,n}| |\omega_{lk,n}| \sum_{p=1}^n |\omega_{lp,n}| \\ &\leq n^{-1} c_\varepsilon^2 \kappa_m \kappa_\omega \sum_{i=1}^n \sum_{l=1}^n \sum_{k=1}^n |m_{ki,n}| |\omega_{lk,n}| =: \xi_4. \end{aligned}$$

It then follows that

$$\begin{aligned} \xi_4 &= n^{-1} c_\varepsilon^2 \kappa_m \kappa_\omega \sum_{i=1}^n \sum_{k=1}^n |m_{ki,n}| \sum_{l=1}^n |\omega_{lk,n}| \\ &\leq n^{-1} c_\varepsilon^2 \kappa_m \kappa_\omega^2 \sum_{i=1}^n \sum_{k=1}^n |m_{ki,n}| =: \xi_5 \end{aligned}$$

Finally, we have

$$\begin{aligned} \xi_5 &\leq n^{-1} \sum_{i=1}^n c_\varepsilon^2 \kappa_m^2 \kappa_\omega^2 \\ &= c_\varepsilon^2 \kappa_m^2 \kappa_\omega^2 < \infty. \end{aligned}$$

We may conclude that $E|n^{-1}(z'_n z_n)| = E|n^{-1} \varepsilon'_n M'_n W'_n W_n M_n \varepsilon_n|$ is bounded and thus is $O_p(1)$.

$$(ii) \text{ plim}_{n \rightarrow \infty} n^{-1}(z'_n \varepsilon_n) = O_p(1).$$

Applying an analogue step to the previous we obtain

$$\begin{aligned}
E|n^{-1}(z'\varepsilon_n)| &= E|n^{-1}\varepsilon_n' M_n' W_n' \varepsilon_n| \\
&= E\left|n^{-1}\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n\varepsilon_{i,n}m_{ki,n}\omega_{jk,n}\varepsilon_{j,n}\right| \\
&\leq n^{-1}\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n|m_{ki,n}|\omega_{jk,n}|E|\varepsilon_{i,n}||\varepsilon_{j,n}| \\
&\leq c_\varepsilon^2\kappa_m\kappa_\omega < \infty,
\end{aligned}$$

and it follows that $E|n^{-1}(z'\varepsilon_n)|$ is $O_p(1)$. Thus, the bias is $O_p(1)$, as the product of $O_p(1)$ terms in the same probability space is $O_p(1)$. \square

Theorem 2 *Given Assumptions 1–3*

$$\text{plim}_{n\rightarrow\infty} n^{-1}(z'\varepsilon_n) \neq 0, \quad \forall \theta \neq 0. \quad (10)$$

Proof.

Note that the expression $\text{plim}_{n\rightarrow\infty} n^{-1}(z'\varepsilon_n)$ can be rewritten as

$$\text{plim}_{n\rightarrow\infty} n^{-1}\varepsilon_n'(I - \theta W_n')^{-1}W_n'\varepsilon_n. \quad (11)$$

Let $A_n = B_n'$ with $B_n = W_n M_n$. Define $X_n = (X_{n1}, X_{n2})$ with $X_{n1} = n^{-1/2}\varepsilon_n$ and $X_{n2} = n^{-1/2}B_n'\varepsilon_n$. Considering $S_n = [I, B_n]$ where I denotes an identity matrix of dimension n , we have $X_n = n^{-1/2}S_n'\varepsilon_n$. Using a corollary of the Lindeberg-Feller central limit theorem for triangular arrays (see, e.g., Billingsley (1979, p.319)), the asymptotic distribution of X_n is

$$X_n \xrightarrow[n\rightarrow\infty]{d} N(0, \sigma^2\Gamma_n).$$

with $\Gamma_n = \lim_{n\rightarrow\infty} n^{-1}S_n'S_n$. Applying the continuous mapping theorem (see, e.g., Serfling (1980, p.24)), it then follows that

$$X_{n1}'X_{n2} \xrightarrow[n\rightarrow\infty]{d} X_1'X_2,$$

with $X = (X_1, X_2)' \approx^d N(0, \sigma^2 \Gamma_n)$. Given the maintained assumptions of the model, it is straightforward to verify that each block of the matrix $S_n' S_n$ is nonsingular. As a result, X_1 and X_2 are not perfectly correlated and thus $P(|X_1' X_2| > \delta) > 0, \forall \delta > 0$. Then $\text{plim}_{n \rightarrow \infty} X_{n1}' X_{n2} \neq 0$. \square

4 Conclusion

In this paper we provide the theoretical proof of the inconsistency of ordinary least squares estimator for cross-sectional autoregressive spatial models. We show that the bias of this estimator is $O_p(1)$ for all $|\theta| < 1, \theta \neq 0$. The contribution of this study consists in providing arguments for not using the OLS estimator (contrary to time-series) in regression models for spatial data even if this is highly tentative.

Acknowledgements

We are grateful to François Laisney, for his detailed constructive criticisms in developing this work. We also thank Gaël Giraud for helpful comments. The usual disclaimer applies.

References

- AMEMIYA, T. (1985): *Advanced Econometrics*. Cambridge: Harvard University Press.
- ANSELIN, L. (1988): *Spatial Econometrics Methods and Models*. Dordrecht: Kluwer.

- ANSELIN, L., AND A. BERA (1998): "Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics," in *Handbook of Applied Economic Statistics*, A. Truchmuche (eds.), pp. 237–289.
- BALTAGI, B. H., AND D. LI (1999): "Prediction in the Panel Data Model with Spatial Correlation," *Texas A&M University, Department of Economics*.
- BELL, K. P., AND N. E. BOCKSTAEEL (2000): "Applying the Generalized-Moments Estimation Approach to Spatial Problems Involving Microlevel Data," *The Review of Economics and Statistics*, (82–1), 72–82.
- BERGE, C. (1983): *Graphes*. Paris: Gauthier-Villars.
- BILLINGSLEY, P. (1979): *Probability and Measure*. New York: Wiley.
- COCHRANE, D., AND G. ORCUTT (1949): "Application of Least Squares Regressions to Relationships Containing Autocorrelated Error Terms," *Journal of the American Statistical Association*, (44), 32–61.
- CONLEY, T. (1999): "GMM Estimation with Cross Sectional Dependence," *Journal of Econometrics*, (92), 1–45.
- CRESSIE, N. (1991): *Statistics for Spatial Data*. New York: Wiley-Interscience.
- HORN, R., AND C. JOHNSON (1985): *Matrix Analysis*. Cambridge: Cambridge University Press.
- KELEJIAN, H., AND I. PRUCHA (1997): "Estimation of Spatial Regression Models with Autoregressive Errors by Two-Stage Least Squares Proce-

- dures: A Serious Problem,” *International Regional Science Review*, (20-1), 103–111.
- (1998): “A Generalized Spatial two-stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbance,” *Journal of Real Estate Finance and Economics*, (17-1), 99–121.
- (1999): “A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model,” *International Economic Review*, (40), 509–533.
- RIPLEY, B. D. (1981): *Spatial Statistics*. New York: Wiley.
- (1988): *Statistical Inference for Spatial Processes*. Cambridge: Cambridge University Press.
- SAJJAN, S. G. (2000): “A note on central limit theorems for lattice models,” *Journal of Statistical Planning and Inference*, (83), 283–290.
- SERFLING, R. (1980): *Approximation Theorems of Mathematical Statistics*. New York: Wiley.
- WHITE, H. (1984): *Asymptotic Theory for Econometricians*. London: Academic Press.
- ZILIAK, J. P., B. A. WILSON, AND J. A. STONES (1999): “Spatial Dynamics and Heterogeneity in The Cyclicity of Real Wages,” *The Review of Economics and Statistics*, (81-2), 227–236.