OLS-based estimation of the disturbance variance under spatial autocorrelation

Walter Krämer[*] and Christoph Hanck[†]

This version: October 2006

Abstract

We investigate the OLS-based estimator s^2 of the disturbance variance in the standard linear regression model with cross section data when the disturbances are homoskedastic, but spatially correlated. For the most popular model of spatially autoregressive disturbances, we show that s^2 can be severely biased in finite samples, but is asymptotically unbiased and consistent for most types of spatial weighting matrices as sample size increases.

Keywords: regression, spatial error correlation, bias, variance

1 Introduction

We consider the standard linear regression model

$$y = X\beta + u,$$

where y is $N \times 1$, X is nonstochastic $N \times K$ with rank K and β is unknown $K \times 1$. The components of u have expected value $E(u) = 0$ and a common variance $E(u_i^2) = \sigma^2$. The OLS estimate for β is $\hat{\beta} = (X'X)^{-1}X'y$, and the OLS-based estimate for σ^2 is

$$s^2 = \frac{1}{N-K}(y - X\hat{\beta})'(y - X\hat{\beta}) = \frac{1}{N-K}u'Mu, \tag{1}$$

where $M = I - X(X'X)^{-1}X'$. It has long been known that s^2 is in general (and contrary to $\hat{\beta}$) biased whenever $V := Cov(u)$ is no longer a multiple of the identity matrix. Krämer [1991] and Krämer and Berghoff [1991] show that this problem disappears asymptotically for certain types of temporal correlation such as stationary AR(1)-disturbances, although it is clear from Kiviet and Krämer [1992] that the relative bias of s^2 might still be substantial for any finite sample size. The present paper extends these analyses to the case of spatial correlation, where we allow the disturbance vector u to be generated by the spatial autoregressive scheme

$$u = \rho W u + \epsilon, \tag{2}$$

where ϵ is a $N \times 1$ random vector with mean zero and scalar covariance matrix $\sigma_\epsilon^2 I$ and W is some known $N \times N$-matrix of nonnegative spatial weights with $w_{ii} = 0$ ($i = 1, \ldots, N$). Such patterns of dependence are often entertained when the objects under study are positioned in some "space," whether geographical or sociological (in some social network, say) and account for spillovers from one unit to its neighbors, whichever way "neighborhood" may be defined. They date back to Whittle [1954] and have become quite popular in econometrics recently. See Anselin and Florax [1995] or Anselin [2001] for surveys of this literature.

The coefficient ρ in (2) measures the degree of correlation, which can be both positive and negative. Below we focus on the empirically more relevant case of positive disturbance correlation, where

$$0 \leqslant \rho \leqslant \frac{1}{\lambda_{\max}} \tag{3}$$

and where $\lambda_{\max}$ is the Frobenius-root of W (i.e. the unique positive real eigenvalue such that $\lambda_{\max} \geqslant |\lambda_i|$ for arbitrary eigenvalues λ_i). The disturbances are then given by

$$u = (I - \rho W)^{-1}\epsilon, \tag{4}$$

so

$$V = Cov(u) = \sigma_\epsilon^2[(I - \rho W)'(I - \rho W)]^{-1} \tag{5}$$

and

$$V = \sigma_\epsilon^2 I \tag{6}$$

whenever $\rho = 0$.

Of course, for our analysis to make sense, the main diagonal of V should be constant, i.e.

$$V = \sigma^2 \Sigma, \tag{7}$$

where Σ is the correlation matrix of the disturbance vector.[1] It is therefore important to clarify that many, though not all, spatial autocorrelation schemes are compatible with homoskedasticity. Consider for instance the following popular specification for the weight matrix known as "one ahead and one behind:"

$$\tilde{W} := \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 1 \\ 1 & 0 & \ddots & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & 1 & 0 & 1 \\ 1 & 0 & \cdots & 0 & 1 & 0 \end{pmatrix}$$

---

[1]Note that $\sigma^2 = Var(u_i)$ need not be equal to $\sigma_\epsilon^2 = Var(\epsilon_i)$, unless $\Sigma = I$. In the sequel, we keep σ_ϵ^2 fixed, so σ^2 will in general vary with W and N.

and renormalize the rows such that the row sums are 1. Then it is easily seen that $E(u_i^2)$ is independent of i, and analogous results hold for the more general "j ahead and j behind" weight matrix W which has non-zero elements in the j entries before and after the main diagonal, with the non-zero entries equal to $j/2$. This specification has been considered by, for instance, Kelejian and Prucha [1999] and Krämer and Donninger [1987].

As another example, consider the equal-weight matrix (see, e.g., Kelejian and Prucha [2002], Lee [2004], Case [1992] or Kelejian et al. [2006]), defined by

$$W^{EW} = (w_{ij}^{EW}) = \begin{cases} \frac{1}{N-1} & \text{for } i \neq j \\ 0 & \text{for } i = j \end{cases}. \tag{8}$$

One easily verifies that, for $|\rho| < 1$,

$$(I - \rho W^{EW})^{-1} = \delta_1 J_N + \delta_2 I_N, \tag{9}$$

where

$$\delta_1 = \frac{\rho}{(N-1+\rho)(1-\rho)}, \quad \delta_2 = \frac{N-1}{N-1+\rho} \tag{10}$$

and J_N is an $(N \times N)$ matrix of ones. Without loss of generality, let $\sigma_\epsilon^2 = 1$. We then have, using symmetry of W,

$$\begin{aligned} V &= [(I - \rho W^{EW})'(I - \rho W^{EW})]^{-1} \\ &= (I - \rho W^{EW})^{-1}(I - \rho W^{EW})^{-1} \\ &= (\delta_1 J_N + \delta_2 I_N)^2. \end{aligned} \tag{11}$$

Carrying out the multiplication, it is seen that

$$E(u_i^2) = (\delta_1^2 + \delta_2^2)^2 + (N-1)\delta_1^2 \quad \text{for} \quad i = 1, \dots, N.$$

So V is homoskedastic. It is straightforward to extend this result to the case where W is block-diagonal with B blocks of dimension $(R \times R)$, defined as

$$W_R^{EW} = (w_{R,ij}^{EW}) = \begin{cases} \frac{1}{R-1} & \text{for } i \neq j \\ 0 & \text{for } i = j, \end{cases} \tag{12}$$

where $N = BR$. We therefore conclude that our analysis is applicable in many relevant spatial econometric specifications.

2 The relative bias of s^2 in finite samples

We have

$$\begin{aligned} E\left(\frac{s^2}{\sigma^2}\right) &= E\left(\frac{1}{\sigma^2(N-K)}u'Mu\right) \\ &= \frac{1}{\sigma^2(N-K)}tr(MV) \\ &= \frac{1}{N-K}tr(M\Sigma). \end{aligned} \tag{13}$$

Watson [1955] and Sathe and Vinod [1974] derive the (attainable) bounds

$$\text{mean of } N-K \text{ smallest eigenvalues of } \Sigma$$
$$\leqslant E\left(\frac{s^2}{\sigma^2}\right) \leqslant \tag{14}$$
$$\text{mean of } N-K \text{ largest eigenvalues of } \Sigma,$$

which shows that the bias can be both positive and negative, depending on the regressor matrix X, whatever Σ may be. Finally, Dufour [1986] points out that the inequalities (14) amount to

$$0 \leqslant E\left(\frac{s^2}{\sigma^2}\right) \leqslant \frac{N}{N-K} \tag{15}$$

when no restrictions are placed on X and Σ. Again, these bounds are sharp and show that underestimation of σ^2 is much more of a threat in practise than overestimation.

The problem with Dufour's bounds is that they are unnecessarily wide when extra information on V is available. Here we assume a disturbance covariance matrix V as in (5) and show first that the relative bias of s^2 depends crucially on the interplay between X and W. In particular, irrespective of sample size and of the weighting matrix W, we can always produce a regressor matrix X such that $E(s^2/\sigma^2)$ becomes as close to zero as desired. To see this, let W be symmetric[2] and let

$$W = \sum_{i=1}^{N} \lambda_i \omega_i \omega_i' \tag{16}$$

be the spectral decomposition of W, with the eigenvalues λ_i in increasing order and ω_i the corresponding orthonormal eigenvectors. Now it is easily seen that

$$\lim_{\rho \to 1/\lambda_N} E\left(\frac{s^2}{\sigma^2}\right) = 0 \tag{17}$$

whenever

$$M\omega_N = 0. \tag{18}$$

This follows from

$$V = \sigma_\epsilon^2 \left[\sum_{i=1}^{N} \frac{1}{(1 - \rho\lambda_i)^2} \omega_i \omega_i' \right] \tag{19}$$

and

$$\Sigma = \frac{1}{\sigma^2} V = \frac{1}{\sum_{i=1}^{N} \frac{1}{(1-\rho\lambda_i)^2}\omega_{i1}^2} \sum_{i=1}^{N} \frac{1}{(1 - \rho\lambda_i)^2} \omega_i \omega_i', \tag{20}$$

where ω_{i1}^2 is the $(1,1)$-element of $\omega_i\omega_i'$ (under homoscedasticity, we could select any diagonal element of $\omega_i\omega_i'$) and

$$\sigma^2 = \sigma_\epsilon^2 \sum_{i=1}^{N} \frac{1}{(1 - \rho\lambda_i)^2} \omega_{i1}^2. \tag{21}$$

---

[2]Notice that for all the homoskedastic examples considered above, row-normalization does not destroy symmetry of W.

Multiplying the numerator and denominator of (20) by $(1 - \rho\lambda_N)^2$, we obtain

$$\Sigma = \frac{1}{\sigma^2}V = \frac{1}{\sum_{i=1}^{N} \frac{(1-\rho\lambda_N)^2}{(1-\rho\lambda_i)^2}\omega_{i1}^2} \sum_{i=1}^{N} \frac{(1-\rho\lambda_N)^2}{(1-\rho\lambda_i)^2}\omega_i\omega_i', \qquad (22)$$

which tends to

$$\frac{1}{\omega_{N1}^2}\omega_N\omega_N' \qquad (23)$$

as $\rho \to 1/\lambda_N$. Given W, one can therefore choose X to be $(N \times 1)$ and equal to ω_N. Then, M is by construction orthogonal to ω_N, which implies that $tr(M\Sigma)$ and therefore also $E(s^2/\sigma^2)$ tend to zero as $\rho \to 1/\lambda_N$.
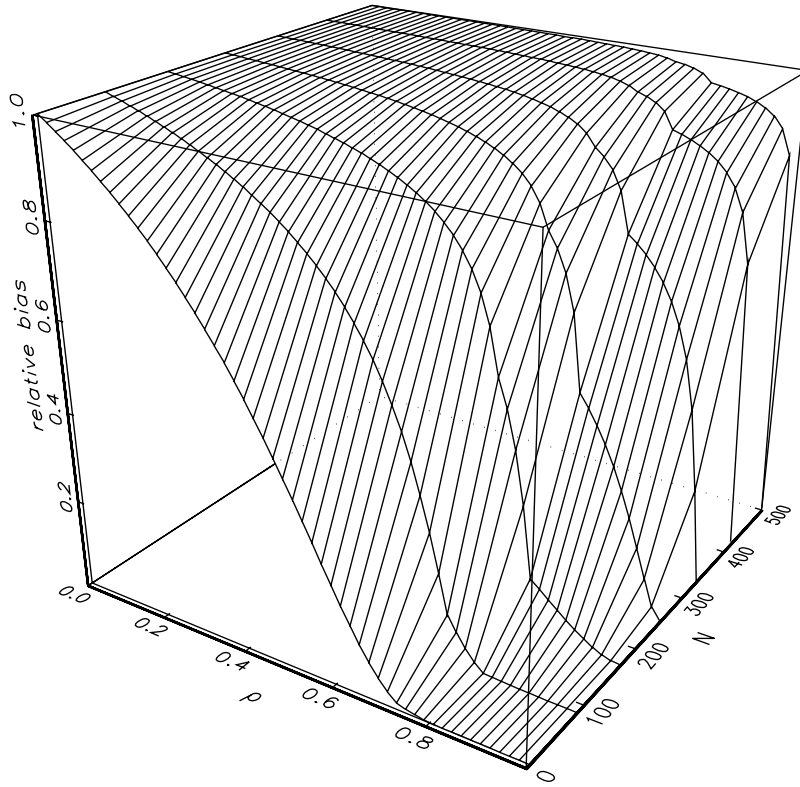

FIGURE I—THE RELATIVE BIAS OF s^2 AS A FUNCTION OF ρ AND N

For illustration, consider the following example. The largest eigenvalue λ_N of a row-normalized matrix such as $\tilde{W}/2$ is 1. (This follows immediately from Theorem 8.1.22

6

of Horn and Johnson [1985].) It is then readily verified that $\omega_N = \iota := (1, \ldots, 1)'$ is (up to the usual multiple) the eigenvector corresponding to λ_N. Now, if $X = \iota$, $M\omega_N = (I - \frac{1}{N}\iota\iota')\iota = 0$. Figure I shows the behaviour of the relative bias as $\rho \to 1/\lambda_N = 1$. We see that (17) holds for any given N. Also, pointwise in ρ, the relative bias vanishes as $N \to \infty$, as one would expect. We now rigorously establish the latter property.

3   Asymptotic bias and consistency

From (15), it is clear that, for any V, the relative upward bias of s^2 must vanish as $N \to \infty$. A sufficient condition for the relative downward bias to disappear as well is that the largest eigenvalue of Σ, μ_N, is

$$\mu_N = o(N). \tag{24}$$

This is so because, using $\sum_{i=1}^{N} \mu_i = \sum_{i=1}^{N-K} \mu_i + \sum_{i=1}^{K} \mu_{i+N-K} = N$, we have

$$
\begin{aligned}
\text{mean of } N-K \text{ smallest eigenvalues of } \Sigma \;\; &= \;\; \frac{N}{N-K} - \frac{1}{N-K}\sum_{i=1}^{K}\mu_{i+N-K} \\
&\geqslant \;\; \frac{N}{N-K} - \frac{K}{N-K}\mu_N
\end{aligned}
$$

and the right-hand side tends to 1 when (24) holds as $N \to \infty$.

Condition (24) also guarantees consistency. From (1), we have

$$s^2 = \frac{1}{N}u'Mu = \frac{1}{N}u'u - \frac{1}{N}u'Hu, \tag{25}$$

where $H = X(X'X)^{-1}X'$. Since $u'u/N \overset{p}{\longrightarrow} \sigma^2$, it remains to show that

$$\frac{1}{N}u'Hu \overset{p}{\longrightarrow} 0. \tag{26}$$

To this purpose, consider

$$
\begin{aligned}
E\left(\frac{1}{N}u'Hu\right) &= E\left(\frac{1}{N}\varepsilon'\Sigma^{1/2}H\Sigma^{1/2}\varepsilon\right) \qquad \text{(where } \varepsilon = \Sigma^{-1/2}u) \\
&= \frac{\sigma^2}{N}tr(\Sigma^{1/2}H\Sigma^{1/2}) \\
&= \frac{\sigma^2}{N}tr(H\Sigma) \\
&\leqslant \frac{\sigma^2}{N}K\cdot\mu_N, \qquad\qquad\qquad\qquad\qquad (27)
\end{aligned}
$$

where the inequality follows from the fact that $H\Sigma$ has rank K (since rank $(H) = K$). Since no eigenvalue of $H\Sigma$ can exceed μ_N, and $H\Sigma$ has exactly K nonzero eigenvalues, the inequality follows from the well known fact that the trace of a matrix equals the sum of its eigenvalues. By assumption, $\mu_N/N \to 0$ as $N \to \infty$, so in view of (27), $E(u'Hu/N) \to 0$. As $u'Hu$ is nonnegative, this in turn implies $u'Hu/N \xrightarrow{p} 0$ and therefore the consistency of s^2.

The crucial condition (24) is a rather mild one; in the present context, it obviously depends on the weighting matrix W. From (7) and (19), we have

$$
\mu_N = \frac{\sigma_\epsilon^2}{\sigma^2(1-\rho\lambda_N)^2}, \qquad\qquad\qquad\qquad (28)
$$

so the condition (24) obtains whenever

$$
\sigma^2(1-\rho\lambda_N)^2N \to \infty \qquad\qquad\qquad\qquad (29)
$$

For row-normalized weight matrices, $\lambda_N \equiv 1$ irrespective of N, so (29) holds trivially, provided σ^2 remains bounded away from zero. This in turn follows from the fact that, in view of (21),

$$
\sigma^2 \geqslant \frac{\sigma_\epsilon^2}{(1-\rho\lambda_N)^2}\sum_{i=1}^{N}\omega_{i1}^2, \qquad\qquad\qquad\qquad (30)
$$

where

$$
\sum_{i=1}^{N}\omega_{i1}^2 = 1 \qquad\qquad\qquad\qquad (31)
$$

8

as $\Omega = (\omega_1, \ldots, \omega_N)$ satisfies $\Omega\Omega' = I$.

As another example, consider the "one ahead and one behind" matrix adapted to a "non-circular world" where the $(1, N)$ and $(N, 1)$ entries of $\tilde{W}$ are set to zero, such that after row-normalization,

$$
W' := \begin{pmatrix}
0 & 0.5 & 0 & \cdots & 0 & 0 \\
0.5 & 0 & \ddots & 0 & \cdots & 0 \\
0 & \ddots & \ddots & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & 0 & \ddots & 0 \\
0 & \cdots & 0 & 0.5 & 0 & 0.5 \\
0 & 0 & \cdots & 0 & 0.5 & 0
\end{pmatrix}.
$$

Ord [1975] shows that the eigenvalues of W' are then given by

$$
\lambda'_i = \cos\left(\frac{\pi i}{N+1}\right), \ i = 1, \ldots, N,
$$

so

$$
\lambda'_i \in [-1, 1], \ i = 1, \ldots, N.
$$

References

Anselin, L.: 2001, 'Rao's score test in spatial econometrics'. *Journal of Statistical Planning and Inference* **97**(1), 113–139.

Anselin, L. and R. Florax: 1995, 'Small sample properties of tests for spatial dependence in regression models: Some further results'. In: L. Anselin and R. Florax (eds.): *New Directions in Spatial Econometrics*. Berlin: Springer, pp. 21–74.

Case, A.: 1992, 'Neighborhood Influence and Technological Change'. *Regional Science and Urban Economics* **22**(2), 491–508.

Dufour, J.-M.: 1986, 'Bias of s^2 in Linear Regression with Dependent Errors'. *The American Statistican* **40**(4), 284–285.

Horn, R. A. and C. R. Johnson: 1985, *Matrix Analysis*. Cambridge: Cambridge University Press.

Kelejian, H. H. and I. R. Prucha: 1999, 'A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model'. *International Economic Review* **40**(2), 509–533.

Kelejian, H. H. and I. R. Prucha: 2002, '2SLS and OLS in a spatial autoregressive model with equal spatial weights'. *Regional Science and Urban Economics* **32**(6), 691–707.

Kelejian, H. H., I. R. Prucha, and Y. Yuzefovich: 2006, 'Estimation Problems in Models with Spatial Weighting Matrices Which Have Blocks of Equal Elements'. *Journal of Regional Science* **46**(3), 507–515.

Kiviet, J. and W. Krämer: 1992, 'Bias of s^2 in the Linear Regression Model with Autocorrelated Errors'. *The Review of Economics and Statistics* **74**(2), 362–365.

Krämer, W.: 1991, 'The Asymptotic Unbiasedness of S^2 in the Linear Regression Model with AR(1)-Disturbances'. *Statistical Papers* **32**(1), 71–72.

Krämer, W. and S. Berghoff: 1991, 'Consistency of s^2 in the Linear Regression Model with Correlated Errors'. *Empirical Economics* **16**(3), 375–377.

Krämer, W. and C. Donninger: 1987, 'Spatial Autocorrelation Among Errors and the Relative Efficiency of OLS in the Linear Regression Model'. *Journal of the American Statistical Association* **82**(398), 577–579.

Lee, L.-F.: 2004, 'Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models'. *Econometrica* **72**(6), 1899–1925.

Ord, K.: 1975, 'Estimation Methods for Models of Spatial Interaction'. *Journal of the American Statistical Association* **70**(349), 120–126.

Sathe, S. and H. Vinod: 1974, 'Bounds on the Variance of Regression Coefficients due to Heteroscedastic or Autoregressive Errors'. *Econometrica* **42**(2), 333–340.

Watson, G.: 1955, 'Serial Correlation in Regression Analysis I'. *Biometrika* **42**(3/4), 327–341.

Whittle, P.: 1954, 'On Stationary Processes in the Plane'. *Biometrika* **41**(3/4), 434–449.