SFB
823

The exact bias of S^2 in linear panel regressions with spatial autocorrelation

Christoph Hanck, Walter Krämer

Discussion Paper

Nr. 18/2010

The exact bias of S^2 in linear panel regressions with spatial autocorrelation

Christoph Hanck* and Walter Krämer†

May 2010

Abstract

We investigate the OLS-based estimator s^2 of the disturbance variance in an error component linear panel regression model when the disturbances are homoskedastic, but spatially correlated. Although consistent (Song and Lee, Econ. Lett. 2008), s^2 can be arbitrarily biased towards zero in finite samples.

Keywords: panel data, spatial error correlation, bias

1 Introduction and Model

We consider the following standard linear panel regression model

$$y_{it} = X_{it}'\beta + u_{it}, \qquad i = 1, \ldots, N; \qquad t = 1, \ldots, T \tag{1}$$

where i indexes units and t indexes time. The dependent variable y_{it} is affected by the (for simplicity fixed) $K \times 1$ regressor vector X_{it} through the unknown $K \times 1$ vector β and the error u_{it}. The errors follow the one-way error component structure

$$u_{it} = \mu_i + \epsilon_{it}, \qquad i = 1, \ldots, N; \qquad t = 1, \ldots, T$$

where the individual-specific effect μ_i is i.i.d. with mean 0 and variance σ_μ^2. In matrix notation, (1) can be written as $y = X\beta + u$, where the observations are stacked such that y, $u = \mu \otimes \iota_T + \epsilon$ and $\epsilon = (\epsilon_1', \ldots, \epsilon_T')'$ are $NT \times 1$ (ι_T is a $T \times 1$-vector of ones), while X is $NT \times K$ with rank K and $\mu = (\mu_1, \ldots, \mu_N)'$ is $N \times 1$. The idiosyncratic error $\epsilon_t = (\epsilon_{1t}, \ldots, \epsilon_{Nt})'$ is generated by a first-order spatial autoregressive process

$$\epsilon_t = \rho W \epsilon_t + \nu_t, \tag{2}$$

*Corresponding author. Rijksuniversiteit Groningen, Department of Economics, Econometrics and Finance, Nettelbosje 2, 9747AE Groningen, Netherlands. Tel. +31 50 363 3836, Fax+31 50 363 7337. c.h.hanck@rug.nl.

†Universität Dortmund, Fakultät Statistik, Vogelpothsweg 78, 44227, Dortmund, Germany. Research supported by DFG under Sonderforschungsbereich 823. Tel. +49 231 755 3125. walterk@statistik.uni-dortmund.de.

where ρ is the scalar spatial autoregressive coefficient and the elements of $\nu = (\nu_1', \dots, \nu_T')' = (\nu_{11}, \dots, \nu_{N1}, \dots, \nu_{NT})'$ are i.i.d. with $(0, \sigma_\nu^2)$. The $N \times N$-matrix W has known nonnegative spatial weights with $w_{ii} = 0$ $(i = 1, \dots, N)$. Such patterns of dependence are often entertained when the objects under study are positioned in some "space," whether geographical or sociological, 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 [2001] for a survey 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 $0 \leqslant \rho \leqslant 1/\lambda_{\max}$ 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).

Under (2), u satisfies $u = (I_N \otimes \iota_T)\mu + (B^{-1} \otimes I_T)\nu$, with I_M the identity matrix of dimension M and $B = I_N - \rho W$. Furthermore, the variance-covariance matrix of u becomes

$$E(uu') =: \Omega = \sigma_\mu^2 \big[I_N \otimes J_T\big] + \sigma_\nu^2 \big[(B'B)^{-1} \otimes I_T\big], \tag{3}$$

where J_T is a $T \times T$ matrix of ones. The pooled OLS estimate for β is $\hat{\beta} = (X'X)^{-1}X'y$, and the OLS-based estimate for $\sigma^2 := \mathrm{Var}(u_{it})$ is

$$s^2 = \frac{1}{NT - K}(y - X\hat{\beta})'(y - X\hat{\beta}) = \frac{1}{NT - K} u'Mu, \tag{4}$$

where $M = I - X(X'X)^{-1}X'$. The present paper shows that, while consistent, s^2 can still be arbitrarily biased. Of course, for our analysis to make sense, the main diagonal of Ω should be constant, i.e. $\Omega = \sigma^2\Sigma$, where Σ is the correlation matrix of u. It is therefore important to clarify that many, though not all, spatial autocorrelation schemes imply homoskedasticity of Ω. Consider for instance the following popular specification for W known as "one ahead and one behind:"

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

and renormalize the rows such that the row sums are 1, i.e. take $\tilde{W}/2$. Then it is easily seen that $E(u_{it}^2)$ is independent of i and t, 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]. The condition is also met if W is the equal-weight matrix (see, e.g., Kelejian and Prucha [2002], Lee [2004] or Case [1992]) W^{EW}, defined by

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

Our results therefore hold for both $W = W^{EW}$ and $W = \tilde{W}/2$, among others. In the sequel, we work with $W = W^{EW}$ for brevity, following Song and Lee [2008].

It has long been known that s^2 is in general (and contrary to $\hat{\beta}$) biased for σ^2 whenever Ω 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 in a standard linear regression model, 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. Recently, Song and Lee [2008] prove consistency of s^2 for σ^2 also under the panel model (1), for $W = W^{EW}$. We extend their results to the finite-sample case by showing that s^2 can nevertheless be arbitrarily biased for σ^2.

2 The relative bias of s^2 in finite samples

We have

$$\begin{aligned} E\left(\frac{s^2}{\sigma^2}\right) &= E\left(\frac{1}{\sigma^2(NT-K)} u'Mu\right) \\ &= \frac{1}{\sigma^2(NT-K)} \operatorname{tr}(M\Omega) \\ &= \frac{1}{NT-K} \operatorname{tr}(M\Sigma). \end{aligned} \tag{5}$$

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

$$\begin{aligned} &\text{mean of } NT-K \text{ smallest eigenvalues of } \Sigma \\ &\qquad \leqslant E\left(\frac{s^2}{\sigma^2}\right) \leqslant \\ &\text{mean of } NT-K \text{ largest eigenvalues of } \Sigma, \end{aligned} \tag{6}$$

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 (6) amount to

$$0 \leqslant E\left(\frac{s^2}{\sigma^2}\right) \leqslant \frac{NT}{NT-K} \tag{7}$$

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

The problem with Dufour's bounds is that they are unnecessarily wide when extra information on Ω is available. Here we assume Ω to satisfy (3) and show that the relative bias of s^2 depends crucially on the interplay between X and W. In particular, irrespective of N and of the weighting matrix W, there is always a regressor matrix X such that $E(s^2/\sigma^2)$ becomes as close to zero as desired. To see this, note that the above W are symmetric. Hence, we can write $W = \sum_{i=1}^{N} \lambda_i \omega_i \omega_i'$, the spectral decomposition of W, with the eigenvalues λ_i in increasing order and ω_i the corresponding orthonormal eigenvectors. We directly obtain from (5) that

$$\lim_{\rho \to 1/\lambda_N} E\big(s^2/\sigma^2\big) = 0 \tag{8}$$

whenever $\lim_{\rho \to 1/\lambda_N} \mathrm{tr}(M\Sigma) = 0$. Since M is constant w.r.t. ρ and the trace is continuous, we need $\lim_{\rho \to 1/\lambda_N} \Sigma$ to investigate the limiting bias.

Lemma 1. Let $\tilde{\omega}_i = \omega_i \otimes e_k$, with e_k an eigenvector of I_T, and $\tilde{\omega}_{i1}^2$ the $(1,1)$-element of $\tilde{\omega}_i \tilde{\omega}_i'$ (under homoscedasticity, we could select any diagonal element of $\tilde{\omega}_i \tilde{\omega}_i'$). When $\rho \to 1/\lambda_N$, Σ tends to

$$\tilde{\Sigma} := \frac{\sum_{i=T(N-1)+1}^{NT} \tilde{\omega}_i \tilde{\omega}_i'}{\sum_{i=T(N-1)+1}^{NT} \tilde{\omega}_{i1}^2}$$

Proof. Using symmetry of W, write

$$\Omega = \sigma_\mu^2 \big[I_N \otimes J_T\big] + \sigma_\nu^2 \big[(I_N - \rho W)^{-2} \otimes I_T\big]$$

From e.g. Lütkepohl [1996, Sec. 5.2.1], $(I_N - \rho W)^{-2} \otimes I_T$ has eigenvalues $(1-\rho\lambda_i)^{-2}$ with multiplicity T each, as $\lambda_j(I_T) = 1$, $t = 1, \ldots, T$ and using standard results on eigenvalues of Kronecker products [e.g. Abadir and Magnus, 2005, 10.10]. Hence, we can write

$$\Omega = \sigma_\mu^2 \big[I_N \otimes J_T\big] + \sigma_\nu^2 \left[\sum_{i=1}^{NT} \frac{1}{(1-\rho\lambda_i)^2} \tilde{\omega}_i \tilde{\omega}_i'\right],$$

again using standard results on eigenvectors of Kronecker products [e.g. Abadir and Magnus, 2005, 10.11]. Note $\sigma^2 = \sigma_\mu^2 + \sigma_\nu^2 \left[\sum_{i=1}^{NT} \frac{1}{(1-\rho\lambda_i)^2} \tilde{\omega}_{i1}^2\right]$. Hence

$$\Sigma = \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_\nu^2 \left[\sum_{i=1}^{NT} \frac{1}{(1-\rho\lambda_i)^2} \tilde{\omega}_{i1}^2\right]} \big[I_N \otimes J_T\big] + \frac{\sigma_\nu^2}{\sigma_\mu^2 + \sigma_\nu^2 \left[\sum_{i=1}^{NT} \frac{1}{(1-\rho\lambda_i)^2} \tilde{\omega}_{i1}^2\right]} \left[\sum_{i=1}^{NT} \frac{1}{(1-\rho\lambda_i)^2} \tilde{\omega}_i \tilde{\omega}_i'\right]$$

Multiplying numerator and denominator by $(1-\rho\lambda_N)^2$ yields

$$\begin{aligned}
\Sigma \quad = \quad & \frac{\sigma_\mu^2 (1-\rho\lambda_N)^2}{\sigma_\mu^2 (1-\rho\lambda_N)^2 + \sigma_\nu^2 \left[\sum_{i=1}^{NT} \frac{(1-\rho\lambda_N)^2}{(1-\rho\lambda_i)^2} \tilde{\omega}_{i1}^2\right]} \big[I_N \otimes J_T\big] \\
& + \frac{\sigma_\nu^2}{\sigma_\mu^2 (1-\rho\lambda_N)^2 + \sigma_\nu^2 \left[\sum_{i=1}^{NT} \frac{(1-\rho\lambda_N)^2}{(1-\rho\lambda_i)^2} \tilde{\omega}_{i1}^2\right]} \left[\sum_{i=1}^{NT} \frac{(1-\rho\lambda_N)^2}{(1-\rho\lambda_i)^2} \tilde{\omega}_i \tilde{\omega}_i'\right]
\end{aligned}$$

Now, as the T largest eigenvalues of $\big[(I_N - \rho W)^{-2} \otimes I_T\big]$ are equal to λ_N, we have $\lim_{\rho \to 1/\lambda_N} = \frac{(1-\rho\lambda_N)^2}{(1-\rho\lambda_i)^2} = 1$ for $i = i = T(N-1)+1, \ldots, NT$ and 0 else. Thus,

$$\begin{aligned}
\lim_{\rho \to 1/\lambda_N} \Sigma \quad = \quad & \frac{0}{0 + \sigma_\nu^2 \left[\sum_{i=T(N-1)+1}^{NT} \tilde{\omega}_{i1}^2\right]} \big[I_N \otimes J_T\big] \\
& + \frac{\sigma_\nu^2}{0 + \sigma_\nu^2 \left[\sum_{i=T(N-1)+1}^{NT} \tilde{\omega}_{i1}^2\right]} \left[\sum_{i=T(N-1)+1}^{NT} \tilde{\omega}_i \tilde{\omega}_i'\right]
\end{aligned}$$

□

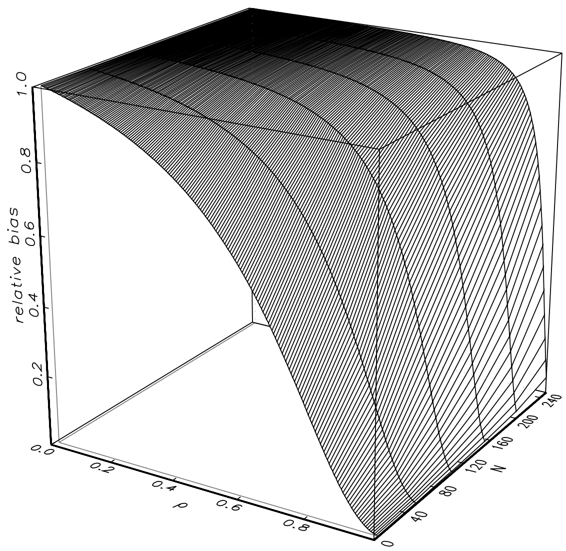

FIGURE I—THE RELATIVE BIAS OF s^2 AS A FUNCTION OF ρ AND N, $X = \iota_N \otimes I_2$

Remark 1. The result, and by extension the Proposition below, also holds if $\sigma_\mu^2 = 0$, i.e. if the if the individual-specific effects are homogenous.

We now demonstrate that s^2 can be arbitrarily biased for σ^2.

Proposition 1. For any W such that σ^2 is constant over i and t, there always exists an X such that $\lim_{\rho\to 1/\lambda_N} E\left(s^2/\sigma^2\right) = 0$.

Proof. Choose I_T as eigenvectors for I_T. The largest eigenvalue λ_N of a row-normalized matrix such as $\tilde{W}/2$ or W^{EW} is 1. (This follows immediately from Theorem 8.1.22 of Horn and Johnson [1985].) Hence, $\omega_N = a\iota_N$ for some $a \in \mathbb{R} \setminus \{0\}$. ω_N then also is the eigenvector corresponding to the largest eigenvalue of B^{-2}, $(1-\rho)^{-2}$. (This follows because, for e.g. W^{EW}, $B^{-2} = (\delta_1 J_N + \delta_2 I_N)^2 = (N\delta_1^2 + 2\delta_1\delta_2)J_N + \delta_2^2 I_N$, where $\delta_1 = \rho/[(N-1+\rho)(1-\rho)]$, $\delta_2 = (N-1)/(N-1+\rho)$, is a matrix with constant on- and off-diagonal elements. Then, $aB^{-2}\iota_N$ is a vector with identical elements $aN[N\delta_1^2 + 2\delta_1\delta_2)] + \delta_2^2$. Some algebra shows this to be equal to $a(1-\rho)^{-2}$.) Hence, $(\tilde{\omega}_i)_{i=T(N-1)+1,\ldots,NT} = (a\iota_N) \otimes I_T$. Thus, the numerator of $\tilde{\Sigma}$, $\tilde{\Omega}$, can be written as a matrix in which the jth column, $j = 1, T+1, \ldots, T(N-1)+1$ is $(a^2, 0'_{T-1}, a^2, 0'_{T-1}, \ldots, a^2, 0'_{T-1})'$. The adjacent $t' = 1, \ldots, T-1$ columns start on t' zeros before the sequence $a^2, 0'_{T-1}, a^2, 0'_{T-1}, \ldots$ sets in. E.g., for $T = 2$, $\tilde{\Omega}$ is

$$\tilde{\Omega} := \sum_{i=T(N-1)+1}^{NT} \tilde{\omega}_i \tilde{\omega}_i' = \begin{pmatrix} a^2 & 0 & a^2 & \cdots & a^2 & 0 \\ 0 & a^2 & 0 & & & a^2 \\ a^2 & 0 & \ddots & \ddots & & 0 \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ a^2 & & & \ddots & a^2 & 0 \\ 0 & a^2 & 0 & \cdots & 0 & a^2 \end{pmatrix}$$

Hence $\operatorname{tr}(\tilde{\Omega}) = NTa^2$. Now, take $X = \iota_N \otimes I_T$. Then, $I_{NT} - M = (\iota_N \otimes I_T)[(\iota_N \otimes I_T)'(\iota_N \otimes I_T)]^{-1}(\iota_N \otimes I_T)' = N^{-1}(J_N \otimes I_T)$. We have $N^{-1}\operatorname{diag}\big((J_N \otimes I_T)\tilde{\Omega}\big) = N^{-1}(NT/Ta^2, \ldots, NT/Ta^2)' = (a^2, \ldots, a^2)'$, whence $N^{-1}\operatorname{tr}\big((J_N \otimes I_T)\tilde{\Omega}\big) = NTa^2$. Thus, $\operatorname{tr}(M\tilde{\Omega}) = \operatorname{tr}(M\tilde{\Sigma}) = 0$, which, by continuity of tr, completes the proof. $\square$

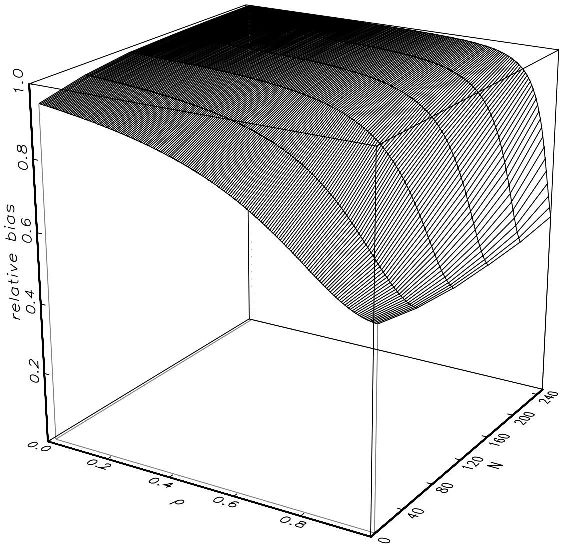

FIGURE II—THE RELATIVE BIAS OF s^2 AS A FUNCTION OF ρ AND N, $X = \iota_{NT}$

Figure I illustrates Theorem 1 for $X = \iota_N \otimes I_2$ for $N = 5, 10, \ldots, 250$ and $W = W^{EW}$. We see that (8) holds for any given N. Also, pointwise in ρ, the relative bias vanishes as $N \to \infty$, as one would expect. To highlight that not all X matrices produce a limiting relative bias of zero, Figure II reports the case $X = \iota_N \otimes \iota_T$. Although the bias still is substantial, it is clearly bounded away from zero. (One can show that $\mathrm{E}(s^2/\sigma^2) \to \frac{NT}{2(NT-K)\tilde{\omega}} a^2$ in this case, where $\tilde{\omega} := \sum_{i=T(N-1)+1}^{NT} \tilde{\omega}_{i1}^2$. As $\frac{NT}{NT-K}$ is slightly larger than one and $\tilde{\omega} = a^2$, a limiting bias slightly above 1/2 obtains for $N < \infty$.)

References

Abadir, Karim M., and Jan R. Magnus, 2005, Matrix Algebra number 1. In 'Econometric Exercises.' (Cambridge University Press, Cambridge).

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

Case, Ann, 1992, Neighborhood influence and technological change, Regional Science and Urban Economics 22(2), 491–508.

Dufour, Jean-Marie, 1986, Bias of s^2 in linear regression with dependent errors, The American Statistican 40(4), 284–285.

Horn, Roger A., and Charles R. Johnson, 1985, Matrix Analysis (Cambridge University Press, Cambridge).

Kelejian, Harry H., and Ingmar R. Prucha, 1999, A generalized moments estimator for the autoregressive parameter in a spatial model, International Economic Review 40(2), 509–533.

— 2002, 2SLS and OLS in a spatial autoregressive model with equal spatial weights, Regional Science and Urban Economics 32(6), 691–707.

Kiviet, Jan, and Walter 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, Walter, 1991, The asymptotic unbiasedness of s^2 in the linear regression model with AR(1)-disturbances, Statistical Papers 32(1), 71–72.

Krämer, Walter, and Christian 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.

Krämer, Walter, and Sonja Berghoff, 1991, Consistency of s^2 in the linear regression model with correlated errors, Empirical Economics 16(3), 375–377.

Lee, Lung-Fei, 2004, Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models, Econometrica 72(6), 1899–1925.

Lütkepohl, Helmut, 1996, Handbook of Matrices (John Wiley & Sons, New York).

Sathe, S.T., and H.D. Vinod, 1974, Bounds on the variance of regression coefficients due to heteroscedastic or autoregressive errors, Econometrica 42(2), 333–340.

Song, Seuck Heun, and Jaejun Lee, 2008, A note on s^2 in a spatially correlated error components regression model for panel data, Economics Letters 101, 41–43.

Watson, Geoff S., 1955, Serial correlation in regression analysis I, Biometrika 42(3/4), 327–341.

Whittle, Peter, 1954, On stationary processes in the plane, Biometrika 41(3/4), 434–449.