〈Research Paper〉

Journal of the Korean
Statistical Society
Vol. 12, No. 2, 1983

The Bias of the Least Squares Estimator of Variance, the Autocorrelation of the Regressor Matrix, and the Autocorrelation of Disturbances

Ki-Jun Jeong*

ABSTRACT

The least squares estimator of disturbance variance in a regression model is biased under a serial correlation. Under the assumption of an AR(1), Theil(1971) crudely related the bias with the autocorrelation of the disturbances and the autocorrelation of the explanatory variable for a simple regression. In this paper we derive a relation which relates the bias with the autocorrelation of disturbances and the autocorrelation of explanatory variables for a multiple regression with improved precision.

1. Introduction

Consider the model, $y=X\beta+\varepsilon$, $\varepsilon\sim N(0,\sigma^2P)$. The regressor matrix X is of order $n\times k$ with full column rank. The first column of X is the $n\times l$ vector of 1's. The $n\times n$ matrix P is the correlation matrix of the disturbance vector ε such that

$$P=\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{n-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{n-3} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \cdots & 1 \end{bmatrix} \tag{1.1}$$

or,

$$P=I+2\rho_1D_1+2\rho_2D_2+\cdots+2\rho_{n-1}D_{n-1}, \tag{1.2}$$

with D_s the $n\times n$ matrix with $(i+s,i)$th and $(i,i+s)$th elements $\frac{1}{2}$ for $i=1,2,\cdots,$

*Seoul National University

$n-s$, and $s=1, 2, \cdots, n-1$, and the other elements 0. For example, for $s=1$, D_s becomes

$$(1.3)\qquad D_1=\frac{1}{2}\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 1 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix}.$$

If ε is generated by an $AR(1)$ process, then

$$(1.4)\qquad \rho_s=\rho^s,\quad s=1, 2, \cdots, n-1,$$

for some ρ bounded by -1 and $+1$.

The least squares estimator of the disturbance variance denoted by s^2 is defined by

$$(1.5)\qquad s^2=e'e/(n-k),$$

where

$$(1.6)\qquad e=\{I-X(X'X)^{-1}X'\}\varepsilon=M\varepsilon.$$

It is well known that s^2 is a biased estimator of σ^2 under a serial correlation. Under the assumption of an $AR(1)$, Theil(1971, p. 256) approximately related the bias with the autocorrelation of the disturbances and the autocorrelation of the explanatory variable for $k=2$. But he did not make any attempt to generalize it, or improve the approximation. Under the assumption of the $AR(1)$, Neudecker(1978) established the bounds for the bias which depended on n, k, and ρ. But he did not establish any relation between the regressor matrix X and the bias.

In this paper we will derive a relation which relates the bias with the autocorrelation of disturbances and the autocorrelation of explanatory variables for arbitrary k. This may be seen as a generalization of Theil's attempt with improved precision so that it may be a practial guide for researchers.

2. The Bias of s^2 and the Autocorrelations of ε and X

As is well known, the expectation of s^2 is given by

$$(2.1)\qquad E(s^2)=\frac{\sigma^2}{n-k}trPM.$$

If we use the minor diagonal decomposition of P given by (1.2), then we have

$$(2.2)\qquad trPM=tr(I+2\rho_1 D_1+2\rho_2 D_2+\cdots+2\rho_{n-1}D_{n-1})M$$

$$=trM+2\rho_1 trD_1M+2\rho_2 trD_2M+\cdots+2\rho_{n-1}trD_{n-1}M$$
$$=(n-k)-2\rho_1t_1-2\rho_2t_2-\cdots-2\rho_{n-1}t_{n-1},$$

where

$$t_s=-trD_sM=trD_sX(X'X)^{-1}X',\quad s=1,2,\cdots,n-1. \tag{2.3}$$

In Appendix we show that under a certain set of assumptions t_s is equal to the sum of the autocorrelation coefficients of k explanatory variables at lag s, or,

$$t_s=r_1(s)+r_2(s)+\cdots+r_k(s), \tag{2.4}$$

where $r_1(s)=a_s$, and $r_j(s)$ $(j=2,3,\cdots,k)$ are bounded by $-a_s$ and a_s, with $a_s=(n-s)/n$ by definition. If we define q_s as

$$q_s=\frac{1}{k-1}(t_s-a_s), \tag{2.5}$$

then under the same assumptions q_s is the average of autocorrelation coefficients at lag s of the $k-1$ explanatory variables, and is bounded by $-a_s$ and a_s. Hereafter, we call q_s the sth autocorrelation coefficient of X.

Combining (2.1), (2.2), and (2.5), we get

$$E(s^2)=\frac{\sigma^2}{n-k}\{(n-k)-2\sum\rho_sa_s-2(k-1)\sum\rho_sq_s\} \tag{2.6}$$
$$=\sigma^2(1+c),$$

where

$$c=\frac{-2}{n-k}\{\sum\rho_sa_s+(k-1)\sum\rho_sq_s\} \tag{2.7}$$

with summations over $s=1,2,\cdots,n-1$. We may call c the relative bias of s^2.

3. A Useful Approximation of the Relative Bias

For further useful results we assume the followings:

Assumption 1 : $\rho_s=\rho^s,\quad s=1,2,\cdots,n-1;$

Assumption 2 : $q_s=q^s,\ s=1,2,\cdots,n-1.$

Assumption 1 is nothing but an assumption of $AR(1)$ process for ε. Assumption 2 is a similar assumption for the explanatory variables. In this context we remind the numerical results obtained by Ames and Reiter(1961). These two authors considered 100 annual statistical series of 25 observations referring to the period 1929—1953, and taken at random from the annual abstract of statistics of the United States. On average, the first five autocorrelation coefficients came out at 0.84, 0.71, 0.60, 0.53, and 0.45. If we multiply each coefficient by $(25-s)/25$, $s=1,2,\cdots,5$, we get q_s-equivalents as 0.81, 0.65, 0.53, 0.44, and 0.36. If we calculate q^s for $q=0.81$, we get 0.81, 0.66, 0.53, 0.43, and 0.35. The correspondence of these figures strongly supports our Assumption 2 for economic time series data.

Under Assumptions 1 and 2, we can approximate the relative bias c by

$$(3.1) \qquad c^*=\frac{-2}{n-k}\left\{\frac{\rho a}{1-\rho a}+(k-1)\frac{\rho q}{1-\rho q}\right\},$$

with $a=a_1=(n-1)/n$, $\rho=\rho_1$, and $q=q_1$. Equation (3.1) shows that the relative bias of s^2 may be calculated approximately using n, k, ρ, and q only.

An acid test of the approximation c^* for the relative bias c may be a direct comparison of the bounds calculated by Neudecker(1978) and the theoretical bounds of c^*. For positive ρ, c^* is a decreasing function of q. Therefore the lower and upper bounds of c^* are reached at $q=a$ and $q=-a$, respectively, or,

$$(3.2) \qquad c_L{}^*=\frac{-2k}{n-k}\;\frac{\rho a}{1-\rho a}$$

$$(3.3) \qquad c_U{}^*=\frac{-2}{n-k}\left\{\frac{\rho a}{1-\rho a}-(k-1)\frac{\rho a}{1+\rho a}\right\}=c_L{}^*\frac{1}{1+\rho a}\left(\rho a-\frac{k-2}{k}\right).$$

Table 1 compares the lower bounds of c^* and c, and Table 2 the upper bounds of them. The bounds of c, c_L, and c_U, are calculated from Table 1 of Neudecker(1978). Table 1 shows that the approximation error of $c_L{}^*$ becomes small as ρ and/or k decreases, or n increases. For each combination of (k, ρ), Table 3 shows the range of n for which the approximation error is less than 0.01. According to Table 2, the

Table 1. Comparison of the Lower Bounds of $c*$ and c

ρ	n	$k=2$		$k=3$		$k=4$		$k=5$	
		$-c_L*$	$-c_L$	$-c_L*$	$-c_L$	$-c_L*$	$-c_L$	$-c_L*$	$-c_L$
0.3	10	0.185	0.165	0.317	0.243	0.493	0.304	0.740	0.354
	15	0.120	0.114	0.194	0.173	0.283	0.224	0.389	0.271
	20	0.141	0.132	0.199	0.175	0.266	0.215		
	20	0.089	0.087	0.141	0.132	0.199	0.175	0.266	0.215
	25	0.070	0.070	0.110	0.107	0.154	0.142	0.202	0.177
	30	0.058	0.058	0.091	0.089	0.126	0.119	0.163	0.149
	50	0.035	0.035	0.053	0.053	0.072	0.072	0.093	0.090
	70	0.025	0.025	0.038	0.03	0.051	0.051	0.065	0.064

ρ	n	$k=2$ $-c_L{}^*$	$-c_L$	$k=3$ $-c_L{}^*$	$-c_L$	$k=4$ $-c_L{}^*$	$-c_L$	$k=5$ $-c_L$"	$-c_L$
0.5	10	0.409	0.327	0.701	0.443	1.091	0.517	1.636	0.570
	15	0.269	0.241	0.438	0.344	0.636	0.420	0.875	0.481
	20	0.201	0.189	0.319	0.276	0.452	0.348	0.603	0.409
	25	0.161	0.155	0.252	0.230	0.352	0.294	0.462	0.352
	30	0.134	0.131	0.208	0.195	0.288	0.254	0.374	0.307
	50	0.080	0.080	0.123	0.121	0.167	0.160	0.214	0.199
	70	0.057	0.057	0.087	0.087	0.118	0.116	0.150	0.145
0.8	10	1.286	0.688	2.204	0.780	3.429	0.822	5.143	0.848
	15	0.907	0.594	1.474	0.709	2.144	0.767	2.947	0.804
	20	0.704	0.519	1.118	0.645	1.583	0.715	2.111	0.761
	25	0.576	0.459	0.903	0.589	1.261	0.667	1.655	0.720
	30	0.487	0.409	0.758	0.540	1.050	0.622	1.365	0.682
	50	0.302	0.281	0.463	0.399	0.631	0.482	0.807	0.551
	70	0.219	0.212	0.334	0.306	0.452	0.386	0.574	0.453

approximation error of $c_U{}^*$ also becomes small as ρ and/or k decreases, or n increases, with minor exceptions. For each combination of (k, ρ), Table 4 shows the range of n

Table 2. Comparison of the Upper Bounds of c_* and c

ρ	n	$k=2$ $-c_U{}^*$	$-c_U$	$k=3$ $-c_U{}^*$	$-c_U$	$k=4$ $-c_U{}^*$	$-c_U$	$k=5$ $-c_U{}^*$	$-c_U$
0.3	10	0.035	0.035	−0.012	−0.012	−0.060	−0.087	−0.107	−0.166
	15	0.024	0.024	−0.007	−0.011	−0.038	−0.050	−0.069	−0.095
	20	0.019	0.019	−0.005	−0.007	−0.028	−0.035	−0.051	−0.065
	25	0.015	0.015	−0.004	−0.005	−0.022	−0.026	−0.041	−0.050
	30	0.013	0.013	−0.003	−0.004	−0.018	−0.021	−0.034	−0.040
	50	0.008	0.008	−0.002	−0.002	−0.011	−0.012	−0.020	−0.022
	70	0.006	0.006	−0.001	−0.001	−0.008	−0.008	−0.014	−0.015
0.5	10	0.113	0.118	0.044	0.043	−0.025	−0.049	−0.094	−0.165
	15	0.080	0.082	0.034	0.035	−0.011	−0.020	−0.057	−0.083
	20	0.061	0.063	0.027	0.028	−0.006	−0.011	−0.040	−0.055
	25	0.050	0.051	0.023	0.023	−0.004	−0.007	−0.031	−0.040
	30	0.042	0.043	0.020	0.010	−0.003	−0.005	−0.025	−0.031
	50	0.026	0.026	0.012	0.013	−0.001	−0.002	−0.014	−0.016
	70	0.019	0.019	0.009	0.009	−0.001	−0.001	−0.010	−0.011
0.8	10	0.478	0.443	0.385	0.381	0.292	0.301	0.199	0.194
	15	0.360	0.349	0.299	0.305	0.238	0.253	0.177	0.191
	20	0.288	0.285	0.242	0.250	0.197	0.210	0.152	0.166
	25	0.240	0.240	0.203	0.210	0.167	0.178	0.131	0.143
	30	0.205	0.206	0.175	0.181	0.145	0.154	0.115	0.125
	50	0.130	0.131	0.112	0.115	0.094	0.099	0.076	0.081
	70	0.095	0.096	0.083	0.084	0.070	0.072	0.057	0.060

Table 3. Range of n for which $|c_L^* - c_L| < 0.01$

k	$\rho=0.3$	$\rho=0.5$	$\rho=0.8$
2	15 or more	25 or more	70 or more
3	20 〃	50 〃	more than 70
4	30 〃	50 〃	〃 〃
5	50 〃	70 〃	〃 〃

Table 4. Range of n for which $|c_U^* - c_U^*| < 0.01$

k	$\rho=0.3$	$\rho=0.5$	$\rho=0.8$
2	10 or more	15 or more	20 or more
3	10 〃	15 〃	20 〃
4	20 〃	20 〃	30 〃
5	25 〃	25 〃	30 〃

for which the approximation error of c_U^* is less than 0.01. According to Table 1 or 3, approximation c_L^* is poor for $\rho=0.8$ if n is less than 70. But for $\rho=0.5$ or less, the approximation may be excellent for moderate n according as the combination (k, ρ). According to Table 2 or 4, the approximation c_U^* is excellent for all cases considered if $n=30$ or more.

4. Some Properties of the Relative Bias of s^2

In Section 3 we have got an approximation of the relative bias of s^2, and examined the accuracy of it. Based on this approximation c^* defined by (3.1), we may get some propositions on the relative bias of s^2:

(1) Under Assumptions 1 and 2, the relative bias depends almost exclusively on the autocorrelations of disturbances and of the regressor matrix, ρ, and q.

(2) If both ρ and q are positive, the bias is definitely negative.

(3) If ρ is positive, the lower bound of the bias is definitely negative.

(4) If ρ is positive, the upper bound of the bias is positive or negative according as $k(1-\rho a)-2$ is positive or negative.

(5) If ρ is positive and $q=0$, the bias is definitely negative.

Proposition (1) follows from equation (3.1) directly. Proposition (2) is well known at least for $k=2$. (See, for example, Theil(1971, p. 257).) Proposition (5) shows an asymmetry of ρ and q. It follows from equation (3.2). Proposition (4) follows from equation (3.3). It is interesting to note that the sign of the upper bound of the bias depends essentially on k and ρ only. Specifically for n not too small (say less than 10), the sign is positive if and only if

$$k \geqq \frac{2}{1-\rho}. \tag{4.1}$$

For example, for $\rho=0.3$, the sign is positive in case $k \geqq 3$; for $\rho=0.5$, in case $k \geqq 4$; and for $\rho=0.8$, in case $k \geqq 10$. These are totally consistent with Neudecker's(1978) results. Neudecker(1978, p. 1223) said, "It appears that for $\rho=0.8$, s^2 is biased toward zero for all values of k considered by us." Note that he did not consider $k>5$. He also said, "Increasing the number of parameters k clearly tends to undermine the conclusion about the sign of the bias [being negative] for low or intermediate values of ρ. An obvious remedy for this is increasing the number of observations n." In the context of this passage, it seems that he did not clearly understand that what is important in connection with the sign of the upper bounds of the bias is k, not n.

APPENDIX

In this Appendix we will show that

$$t_s = trD_sX(X'X)^{-1}X', \quad s=1,2,\cdots,n-1, \tag{A.1}$$

is related to the autocorrelation coefficients of the explanatory variables in X. To this end we need to transform the X matrix into a normalized form.

First, define the $n \times n$ nonsingular matrix A^* as

$$A^* = \frac{1}{\sqrt{n}} \begin{bmatrix} 1 & -\bar{x}_2/s_2 & \bar{x}_3/s_3 & \cdots & \bar{x}_k/s_k \\ 0 & 1/s_2 & 0 & \cdots & 0 \\ 0 & 0 & 1/s_3 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1/s_k \end{bmatrix}, \tag{A.2}$$

where $\bar{x}_j$ and s_j are the mean and the standard deviation, respectively, of the jth explanatory variable in X. Then we can show that X^*, defined by $X^*=XA^*$ is the normalized form of X in the sense that $X^{*\prime}X^*$ is the correlation matrix of X, or

$$X^{*\prime}X^* = R = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & r_{23} & \cdots & r_{2k} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & r_{2k} & r_{3k} & \cdots & 1 \end{bmatrix}, \tag{A.3}$$

where

(A. 4) $$r_{ij}=\boldsymbol{x}_i^{*\prime}\boldsymbol{x}_j^{*}=\frac{\sum_{t=1}^{n}(x_{ti}-\bar{x}_i)(x_{tj}-\bar{x}_j)}{ns_is_j},\quad i,j=2,3,\cdots,k.$$

Further, using the normalized matrix X^*, we can express the sth order serial correlation matrix of the explanatory variables $R(s)$ in a compact form as

(A. 5) $$X^{*\prime}D_sX^*=R(s)=\begin{bmatrix} r_1(s) & r_{12}(s) & \cdots & r_{1k}(s) \\ r_{12}(s) & r_2(s) & \cdots & r_{2k}(s) \\ \cdots & \cdots & \cdots & \cdots \\ r_{1k}(s) & r_{2k}(s) & \cdots & r_k(s) \end{bmatrix},$$

where

$$r_{ij}(s)=\boldsymbol{x}_i^{*\prime}D_s\boldsymbol{x}_j^{*};\quad i,j=1,2,\cdots,k;$$

(A. 6) $$r_{ij}(s)=\boldsymbol{x}_i^{*\prime}D_s\boldsymbol{x}_j^{*}=\frac{\sum_{t=1}^{n-s}(x_{ti}-\bar{x}_i)(x_{t+s,\,j}-\bar{x}_j)}{ns_is_j};\quad i,j=2,3,\cdots,k$$

$$r_1(s)=r_{11}(s)=\frac{n-s}{n}$$

$$r_i(s)=r_{ii}(s),\quad i=1,2,\cdots,k.$$

If we divide $r_i(s)$ by $r_1(s)$, then it is bounded by -1 and 1, so that $r_i(s)$ is bounded by $-(n-s)/n$ and $(n-s)/n$. To further our investigation we need

Lemma 1. $X^*(X^{*\prime}X^*)^{-1}X^{*\prime}=X(X'X)^{-1}X'$.

Proof. Noting that A^* is a nonsingular matrix, the result follows trivially. According to Lemma 1, we can express the matrix $X(X'X)^{-1}X'$ as

(A. 7) $$X(X'X)^{-1}X'=X^*R^{-1}X^{*\prime}.$$

Lemma 2. If $R=I$, then $t_s=trR(s)$.

Proof. $t_s=trD_sX(X'X)^{-1}X'=trD_sX^*R^{-1}X^{*\prime}=trX^{*\prime}D_sX^*=trR(s)$. Lemma 2 says that if the explanatory variables are mutually uncorrelated, then t_s is the sum of the autocorrelation coefficients of the columns of X at lag s, or,

(A. 8) $$t_s=\textstyle\sum_{i=1}^{k}r_i(s).$$

Lemma 2 can be generalized for cases when R is a block diagonal matrix satisfying the following assumptions.

Assumption A1 :

(1) X and X^* are partitioned in G groups as

$$X=[X_1\ X_2\cdots X_G],\quad X_g \text{ is of order } n\times k_g,\quad g=1,2,\cdots,G;$$

$$X^*=[X_1^*\ X_2^*\cdots X_G^*],\quad X_g^* \text{ is of order } n\times k_g,\quad g=1,\cdots,G.$$

(2) R is a block diagonal matrix such that

(A.9) $$R=\begin{bmatrix} R_1 & & & \\ & R_2 & & \\ & & \ddots & \\ & & & R_G \end{bmatrix},$$

where R_g is $k_g \times k_g$ such that

(A.10) $$R_g=R_g(1,r_g)=\begin{bmatrix} 1 & r_g & r_g & \cdots & r_g \\ r_g & 1 & r_g & \cdots & r_g \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ r_g & r_g & r_g & \cdots & 1 \end{bmatrix},\quad g=2,3,\cdots,G;$$

$$R_1=I_{k_1}.$$

(3) $R(s)$ is a block diagonal matrix with gth block $R_g(s)$ such that

(A.11) $$R_g(s)=c_{gs}R_g,\quad g=1,2,\cdots,G;$$

where c_{gs} may be interpreted as the sth common autocorrelation coefficient in the gth group.

Theorem A. Under Assumption A1, $t_s=trR(s)$.

Proof. Under Assumption A1,

(A.12) $$R^{-1}=\begin{bmatrix} R_1^{-1} & & & \\ & R_2^{-1} & & \\ & & \ddots & \\ & & & R_G^{-1} \end{bmatrix}=I_k+\begin{bmatrix} 0 & & & \\ & b_2R_2^* & & \\ & & \ddots & \\ & & & b_GR_G^* \end{bmatrix},$$

where

(A.13) $$R_g^*=R_g(a_g,1),$$

with

$$a_g=-(k_g-1)r_g,$$

(A.14) $$b_g=\frac{-r_g}{1-(k_g-1)r_g^2+(k_g-2)r_g}.$$

Now, we can express the matrix in (A.7) as follows:

(A.15) $$X^*R^{-1}X^{*\prime}=X^*X^{*\prime}+\sum_{g=2}^{G}b_gX_g^*R_g^*X_g^{*\prime}$$

Therefore $t_s=\mathrm{tr}D_sX^*R^{-1}X^{*\prime}$ becomes

(A.16) $$\begin{aligned} t_s&=trD_sX^*X^{*\prime}+\sum_{g=2}^{G}b_gtrD_sX_g^*R_g^*X_g^{*\prime} \\ &=trR(s)+\sum_{g=2}^{G}b_gtrR_g^*R_g(s) \\ &=trR(s)+\sum_{g=2}^{G}b_gc_{gs}trR_g^*R_g. \end{aligned}$$

The last equality holds due to our assumption (A.11). Now,

(A.17) $$\begin{aligned} trR_g^*R_g&=trR_g(a_g,1)R_g(1,r_g) \\ &=k_ga_g+k_g(k_g-1)r_g=k_ga_g-k_ga_g=0. \end{aligned}$$

The second equality holds due to the definition of trace, and the third due to our definition (A.14). Combining (A.16) and (A.17) we arrive at the conclusion of Theorem A.

REFERENCES

(1) Ames, E. and S. Reiter (1961). Distributions of Correlation Coefficients in Economic Time Series, *Journal of the American Statistical Assooiation*, Vol. 56, 637-656.

(2) Jeong, Ki-Jun (1976). An Investigation of the Power of Some Classical Tests in Econometrics, Unpublished Doctoral Dissertation, Claremont Graduate School.

(3) Malinvaud, E. (1980). *Statistical Methods of Econometrics*, 3rd ed., Amsterdam, North-Holland.

(4) Neudecker, H.(1978). Bounds for the Bias of the Least Squares Estimator of σ^2 in the Case of a First-Order (Positive) Autoregressive Process When the Regressor Contains a Constant Term, *Econometrcia*, Vol. 46, 1223-1226.

(5) Theil, H. (1971). *Principles of Econometrics*, New York, Wiley and Sons.