Bias of S^2 in Linear Regressions with Dependent Errors

Author(s): Jean-Marie Dufour
Source: *The American Statistician*, Vol. 40, No. 4 (Nov., 1986), pp. 284-285
Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association
Stable URL: http://www.jstor.org/stable/2684604
Accessed: 02-06-2016 04:44 UTC

gives $k = E\{Z^4\} = V\{Z^2\} + [E\{Z^2\}]^2$, immediately implying the central formula

$$k = V\{Z^2\} + 1. \quad (1)$$

The inequality $k \geq 1$ follows at once; the minimum is attained for a symmetric two-point distribution. Further, (1) implies that k can be viewed as measure of the dispersion of Z^2 around its expectation 1, or equivalently, the dispersion of Z around the values -1 and $+1$. This was essentially Darlington's reasoning, who next concluded that k is a measure of bimodality. This last conclusion does not follow, however. Bimodal distributions can have large kurtosis; this occurs if the modes are not close to the points $Z = \pm 1$. This last phenomenon was illustrated by Hildebrand (1971) by means of a family of double gamma distributions, among others. As a consequence, Darlington's result did not receive the attention it deserves.

A valid interpretation may be formulated as follows: kurtosis measures the dispersion around the *two* values $\mu \pm \sigma$; it is an inverse measure for the concentration in these two points. High kurtosis, therefore, may arise in two situations: (a) concentration of probability mass near μ (corresponding to a peaked unimodal distribution) and (b) concentration of probability mass in the tails of the distribution. The existence of these two possibilities explains the present confusion about the interpretation of kurtosis. Since $V\{Z^2\}$ may be unbounded if heavy tails occur, (1) also indicates clearly why k is strongly affected by the tail behavior of the distribution. Darlington (1970) gave an additional argument by considering the contribution to k of additional point masses.

Sometimes, $k^*: = k - 3$ is presented as an alternative definition of kurtosis, since $k^* = 0$ for normal distributions. There is, however, nothing special about the kurtosis of normal distributions. A better alternative in the spirit of this note would be $k': = k - 1$. Of course, one could as well define a measure of concentration in $\mu \pm \sigma$ as $1 - k$ or $1/k$, the latter having the advantage of taking values between 0 and 1.

Since the derivation of (1) is a nice little exercise in dispersion, I expect to see it in any forthcoming textbook.

[*Received October 1985. Revised January 1986.*]

REFERENCES

Ben-Horim, Moshe, and Levy, Haim (1984), *Statistics: Decisions and Applications in Business and Economics* (2nd ed.), New York: Random House.

Darlington, Richard B. (1970), "Is Kurtosis Really 'Peakedness?'," *The American Statistician*, 24, No. 2, 19–22.

Hildebrand, David K. (1971), "Kurtosis Measures Bimodality?," *The American Statistician*, 25, No. 1, 42–43.

Kendall, M. G., and Stuart, A. (1969), *The Advanced Theory of Statistics* (Vol. 1), London: Charles W. Griffin.

Bias of S^2 in Linear Regressions With Dependent Errors

JEAN-MARIE DUFOUR*

Bounds are given for the expected value of the estimator of the error variance in linear regressions, when the errors are dependent or heteroscedastic. The bounds are valid irrespective of the covariance structure between the errors. Necessary and sufficient conditions to attain the bounds are supplied.

KEY WORDS: Variance estimation; Heteroscedastic errors; Bound on bias.

Recently, David (1985) showed the following interesting result: if $x_1, x_2, \ldots, x_n$ are random variables each with mean μ and variance σ^2, then

$$0 \leq E(s^2) \leq [n/(n-1)]\sigma^2, \quad (1)$$

where $s^2 = \Sigma_{i=1}^n (x_i - \bar{x})^2/(n-1)$. Of course, $E(s^2) = \sigma^2$ when $x_1, \ldots, x_n$ are uncorrelated. It is especially remarkable that the upper bound is valid for any form of dependence between $x_1, \ldots, x_n$. Does a corresponding result hold in a more general setup like the linear regression model? In this article a similar inequality is given for the usual estimator of the error variance in linear regressions when the errors have an arbitrary covariance structure.

Let

$$\mathbf{y} = X\boldsymbol{\beta} + \mathbf{u},$$

where $\mathbf{y}$ is an $n \times 1$ vector of observations on a dependent variable, X is an $n \times k$ fixed regressor matrix such that $1 \leq \text{rank}(X) = k < n$, $\boldsymbol{\beta}$ is a $k \times 1$ vector of coefficients, and $\mathbf{u} = (u_1, \ldots, u_n)'$ is an $n \times 1$ vector of disturbances (or errors) such that $E(\mathbf{u}) = 0$ and $E(\mathbf{uu}') = \Omega$. Ω is an arbitrary covariance matrix. Let

$$\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'\mathbf{y}, \quad \hat{\mathbf{u}} = \mathbf{y} - X\hat{\boldsymbol{\beta}}, \quad s^2 = \hat{\mathbf{u}}'\hat{\mathbf{u}}/(n-k).$$

It is well known that $E(s^2) = \sigma^2$ when $\Omega = \sigma^2 I_n$, but this does not usually hold otherwise. Since $\hat{\mathbf{u}} = M\mathbf{u}$, where $M = I_n - P$ and $P = X(X'X)^{-1}X'$, we have

$$E(\hat{\mathbf{u}}'\hat{\mathbf{u}}) = E(\mathbf{u}'M\mathbf{u}) = \text{tr}(M\Omega) = \text{tr}(\Omega) - \text{tr}(P\Omega). \quad (2)$$

*Jean-Marie Dufour is Associate Professor of Statistics, Département de Sciences Économiques, Université de Montréal, C.P. 6128, Succursale A, Montréal, Québec, H3C 3J7, Canada, and Research Fellow at CORE, Université Catholique de Louvain, 34 voie du Roman Pays, 1348 Louvain-la-Neuve, Belgium. This work was supported by the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, the Fondation FCAR (Government of Québec) and the Centre de Recherche et Développement en Economique (Université de Montréal). The author thanks Herbert David, an associate editor, and two anonymous referees for several useful comments and suggestions.

Since P and Ω are nonnegative definite matrices, we have $\mathrm{tr}(P\Omega) \geq 0$, with $\mathrm{tr}(P\Omega) = 0$ iff $P\Omega = 0$ (see Graybill 1983, pp. 305–306). Hence

$$0 \leq E(\hat{\mathbf{u}}'\hat{\mathbf{u}}) \leq \mathrm{tr}(\Omega) = \sum_{i=1}^{n} \sigma_i^2 = n\bar{\sigma}^2, \qquad (3)$$

where $\sigma_i^2 = \mathrm{var}(u_i) \geq 0$ $(i = 1, \ldots, n)$ and $\bar{\sigma}^2 = \Sigma_{i=1}^n \sigma_i^2/n$. Thus

$$0 \leq E(s^2) \leq [n/(n - k)]\bar{\sigma}^2. \qquad (4)$$

For any given matrix X, the bounds are attainable. From (2), we see that $E(s^2) = 0$ iff $\mathrm{tr}(M\Omega) = 0$, whereas $E(s^2) = [n/(n - k)]\bar{\sigma}^2$ iff $\mathrm{tr}(P\Omega) = 0$. Since P, M, and Ω are nonnegative definite matrices, necessary and sufficient conditions to attain the lower and upper bounds are, respectively, $M\Omega = 0$ and $P\Omega = 0$. Further, the condition for the lower bound is equivalent to the condition

$$\Omega = PAP, \text{ for some nonnegative definite matrix } A, \qquad (5)$$

and the condition for the upper bound is equivalent to

$$\Omega = MCM, \text{ for some nonnegative definite matrix } C. \qquad (6)$$

If the variances of the disturbances are equal ($\sigma_i^2 = \sigma^2$, $i = 1, \ldots, n$), the inequality (4) becomes

$$0 \leq E(s^2) \leq [n/(n - k)]\sigma^2. \qquad (7)$$

The inequalities (4) and (7) hold irrespective of the correlation structure between the disturbances $u_1, \ldots, u_n$. Of course, the inequality (1) is a special case of (7). In the homoscedastic case, we see from (7) that the bias of s^2 is never greater than $\sigma^2 \max\{1, k/(n - k)\}$ (in absolute value). For k/n small, the upper bound on $E(s^2)$ is very close to σ^2 and thus, in most cases, we can expect that s^2 tends to underestimate σ^2. In the heteroscedastic case, the same properties hold with σ^2 replaced by the average variance $\bar{\sigma}^2$.

Another interesting special case is the one in which the disturbances have equal variances and equal covariances: $\Omega = \sigma^2[(1 - \rho)I_n + \rho\mathbf{i}\mathbf{i}']$, where $\mathbf{i} = (1, 1, \ldots, 1)'$ is the $n \times 1$ unit vector, $\sigma^2 \geq 0$, and $-1/(n - 1) \leq \rho \leq 1$. We then have

$$E(s^2) = \sigma^2[(1 - \rho)(n - k) + \rho\mathbf{i}'M\mathbf{i}]/(n - k). \qquad (8)$$

Suppose now that X has $\mathbf{i}$ as one of its columns (which is true in all of the usual applications of the classical linear model). Then $M\mathbf{i} = 0$ and $E(s^2) = \sigma^2(1 - \rho)$. From this last result, we deduce that

$$0 \leq E(s^2) \leq [n/(n - 1)]\sigma^2. \qquad (9)$$

For any value of σ^2, the lower and upper bounds are attained when $\rho = 1$ and $\rho = -1/(n - 1)$, respectively. Finally, if the elements of $\mathbf{u}$ belong to an infinite process of exchangeable random variables, Ω has the same form as above but $0 \leq \rho \leq 1$ so that $0 \leq E(s^2) \leq \sigma^2$.

Further generalizations of these results (e.g., to nonlinear and autoregressive models) are available in Dufour (1985).

[*Received October 1985. Revised February 1986.*]

REFERENCES

David, H. A. (1985), "Bias of S^2 Under Dependence," *The American Statistician*, 39, 201.

Dufour, J.-M. (1985), "On Estimators of the Disturbance Variance in Econometric Models: Some General Small-Sample Results on Bias and the Existence of Moments," working paper, CORE (Université Catholique de Louvain) and Département de Sciences Économiques, Université de Montréal.

Graybill, F. A. (1983), *Matrices With Applications in Statistics* (2nd ed.), Belmont, CA: Wadsworth.