Netherlands School of Economics

ECONOMETRIC INSTITUTE

Report 7003

NOTE ON CONSISTENT ESTIMATION OF THE VARIANCE OF THE DISTURBANCES IN THE LINEAR MODEL

by T. Kloek

February 2, 1970

Preliminary and Confidential

NOTE ON CONSISTENT ESTIMATION OF THE VARIANCE OF THE DISTURBANCES IN THE LINEAR MODEL

by T. Kloek[1]

We consider the linear model $y = X\beta + \varepsilon$ with the classical assumptions: (i) X is an $n \times r$ matrix of rank r with nonstochastic known real elements; (ii) β is an unknown real r-vector; (iii) ε is an n-vector whose elements are independent, identically distributed real-valued random variables with zero mean and unknown variance σ^2; (iv) y is an observed real n-vector which has been generated by $y = X\beta + \varepsilon$. The problem is to estimate β and σ^2, given y and X. The least-squares estimator of β is $b = (X'X)^{-1}X'y$. Well-known estimators of σ^2 are

$$s^2 = (y - Xb)'(y - Xb)/n \tag{1}$$

and

$$\hat{s}^2 = ns^2/(n - r) \tag{2}$$

In this note we shall prove that neither normality of ε has to be assumed, nor additional assumptions on the elements of the matrix X have to be made[2] in order to prove that both s^2 amd $\hat{s}^2$ converge in probability to σ^2. In addition we shall show that under a rather weak condition the residual variance $\bar{s}^2$ obtained from the application of least sqaures to an incorrect model either tends in probability to a number greater than σ^2 or is divergent.

Since it is obvious that s^2 and $\hat{s}^2$ converge to the same probability limit if they converge at all, we shall concentrate our attention on the probability limit of s^2 when the number of observations n tends to infinity.

[1] The author gratefully acknowledges that he received useful comments from Messrs. A.P.J. Abrahamse, S.P. Burley, R. Harkema, and J. Sandee, all of the Econometric Institute.

[2] Proofs which make use either of a normality assumption on ε or of certain restrictions on the X-matrix can be found in several textbooks.

We observe that, if X has rank r for some $n = n_0$, it has rank r for all $n > n_0$, so that for any $n \geqq n_0$ there exists a real nonsingular $r \times r$ matrix P such that

$$(3) \qquad PP' = (X'X)^{-1}$$

or, equivalently,

$$(4) \qquad P'X'XP = I$$

With respect to the disturbance vector ε, we assume that Assumption (iii) above can be extended for the infinite sequence of random variables $\{\varepsilon_n\}_{n=1}^{\infty}$; that is, we assume that $\varepsilon_1, \varepsilon_2, \ldots$ are independent, identically distributed real-valued random variables with zero mean and variance σ^2. This implies that $\varepsilon_1^2, \varepsilon_2^2, \ldots$ are independent, identically distributed random variables with mean σ^2. It then follows from the law of large numbers that

$$(5) \qquad \operatorname*{plim}_{n\to\infty} \frac{\varepsilon'\varepsilon}{n} = \sigma^2$$

Proofs can be found in Feller (1957), Section X.2 or Loève (1955), Section 20.1.

We now return to s^2. It is well known that

$$(6) \qquad s^2 = \frac{1}{n}\varepsilon'\varepsilon - \frac{1}{n}\varepsilon'X(X'X)^{-1}X'\varepsilon$$

Since we have already considered the first term of this expression in (5), we now turn to the second term. It can be written as $-\eta'\eta$ where η is a random r-vector defined by

$$(7) \qquad \eta = P'X'\varepsilon . n^{-\frac{1}{2}}$$

compare (3). It is easily seen that $E\eta = 0$ and that

$$(8) \qquad E(\eta\eta') = \frac{\sigma^2}{n} P'X'XP = \frac{1}{n}\sigma^2 I$$

Thus, the sequence of η vectors converges in the squared mean to zero, which implies $\operatorname*{plim}_{n\to\infty} \eta = 0$, $\operatorname*{plim}_{n\to\infty} \eta'\eta = 0$, and[3]

$$\operatorname*{plim}_{n\to\infty} s^2 = \sigma^2 \tag{9}$$

Next we consider the consequences of a specification error with respect to X. Suppose that one has replaced the true X-matrix by a nonstochastic $n \times p$ matrix Z with rank p and that one has computed $\bar{b} = (Z'Z)^{-1}Z'y$ and

$$\bar{s}^2 = \frac{1}{n}(y - Z\bar{b})'(y - Z\bar{b}) \tag{10}$$

while, in fact, y has been generated by $y = X\beta + \varepsilon$. We assume that there does not exist a p-vector α such that $Z\alpha = X\beta$; otherwise the specification with $Z\alpha$ would be a correct alternative to that with $X\beta$. Upon defining

$$M = I - Z(Z'Z)^{-1}Z' \tag{11}$$

we can write

$$\bar{s}^2 = \frac{1}{n}y'My = \frac{1}{n}\beta'X'MX\beta + \frac{2}{n}\varepsilon'MX\beta + \frac{1}{n}\varepsilon'M\varepsilon \tag{12}$$

or

$$\bar{s}^2 = \frac{1}{n}u'u + \frac{2}{n}\varepsilon'u + \frac{1}{n}\varepsilon'M\varepsilon = v + \frac{1}{n}\varepsilon'M\varepsilon \tag{13}$$

where u and v are defined by

$$u = MX\beta \qquad\qquad v = \frac{1}{n}u'u + \frac{2}{n}\varepsilon'u \tag{14}$$

---

[3] Note that $n^{\frac{1}{2}}\eta$ need not be asymptotically normally distributed, since under the assumptions made the elements of $P'X'$ are not necessarily uniformly asymptotically negligible; compare Loève (1955). A counter-example is the case $X' = [I \quad 0]$, where the identity matrix is of order $r \times r$ and the zero matrix is of order $r \times (n - r)$.

Note that u can be interpreted as the vector of least-squares residuals which results when $X\beta$ is "explained" by the columns of Z. So, our assumption $Z\alpha \neq X\beta$ for all α implies that[4] $u \neq 0$. Now we introduce the additional assumption that, when the sample size increases, from some n_0 onward the mean square of the elements of u is bounded below by a certain positive number g. More precisely, we assume that g exists such that

$$\frac{u'u}{n} > g > 0 \qquad (n \geqq n_0) \tag{15}$$

The random variable v, defined in (14) has mean $u'u/n$ and variance $4\sigma^2 u'u/n^2$. Its standard deviation is $2\sigma(u'u)^{\frac{1}{2}}/n$. From Chebyshev's inequality we have for any positive k

$$P[v > \frac{u'u}{n} - k\,\frac{2\sigma(u'u)^{\frac{1}{2}}}{n}] > 1 - \frac{1}{k^2} \tag{16}$$

or

$$P[v > \frac{u'u}{n}\{1 - \frac{2\sigma k}{(u'u)^{\frac{1}{2}}}\}] > 1 - \frac{1}{k^2} \tag{17}$$

It follows from (15) that $u'u \to \infty$ as $n \to \infty$, so that the factor in braces tends to unity. Hence, for any given $\delta > 0$

$$\lim_{n\to\infty} P[v \geqq g - \delta] = 1 \tag{18}$$

The last term of (13) $\varepsilon' M\varepsilon/n$ converges in probability to σ^2 [the proof is analogous to that of (9)] so that for any given $\delta > 0$

$$\lim_{n\to\infty} P[\bar{s}^2 \geqq g + \sigma^2 - \delta] = 1 \tag{19}$$

This result includes both the case that plim $\bar{s}^2 > \sigma^2$ and the case that $\bar{s}^2$ does not converge at all. We conclude that, as the sample size increases, we may have more confidence that the model with the greater s^2 (that is with the smaller squared correlation coefficient R^2) was incorrectly specified. This

---

[4] It follows from (13) and (14) that $n\bar{s}^2/(n-p)$ has expectation $\sigma^2 + u'u/(n-p) > \sigma^2$. This result has already been given in Theil (1958), Section 6.2.4.

positive result on R^2 for large samples contrasts with recent negative results for small samples found by Koerts and Abrahamse (1969).

REFERENCES

Feller, W. (1957), An Introduction to Probability Theory and Its Applications, Vol. I, Second Edition, Wiley, New York.

Koerts, J., and A.P.J. Abrahamse (1969), On the Theory and Application of the General Linear Model, Rotterdam University Press, Rotterdam.

Loève, M. (1955), Probability Theory, Van Nostrand, New York.

Theil, H. (1958), Economic Forecasts and Policy, North Holland Publishing Company, Amsterdam.