The Annals of Statistics
1976, Vol. 4, No. 4, 788–790

STRONG CONSISTENCY OF LEAST SQUARES ESTIMATES IN NORMAL LINEAR REGRESSION[1]

BY T. W. ANDERSON AND JOHN B. TAYLOR

Stanford University, London School of Economics and Columbia University

In the usual linear regression model the sample regression coefficients converge with probability one to the population regression coefficients when the dependent variables are normally distributed and the inverse of the second-order moment matrix of the independent variables converges to the zero matrix.

In the usual linear model the least squares estimates of the regression coefficients are weakly consistent if the independent variables grow as the number of observations increases in such a way that the covariance matrix of the estimates converges to the matrix consisting of zeros. Here we prove that this condition implies strong consistency when the dependent variables are normally and independently distributed.

THEOREM 1. *Let $y_1, y_2, \cdots$ be a sequence of independently normally distributed random variables with variances σ^2 and expected values*

$$\mathscr{E} y_t = \boldsymbol{\beta}' \mathbf{x}_t\,, \qquad t = 1, 2, \cdots, \tag{1}$$

where $\boldsymbol{\beta}$ is a p-component vector of parameters and $\mathbf{x}_t$ is a p-component nonstochastic vector. Let

$$\mathbf{A}_T = \textstyle\sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t'\,, \tag{2}$$

and suppose $\mathbf{A}_p$ is nonsingular. Define

$$\mathbf{b}_T = \mathbf{A}_T^{-1} \textstyle\sum_{t=1}^T \mathbf{x}_t y_t\,, \qquad T = p, p+1, \cdots. \tag{3}$$

Then $\mathbf{b}_T \to \boldsymbol{\beta}$ as $T \to \infty$ with probability 1 if and only if

$$\mathbf{A}_T^{-1} \to \mathbf{0}\,. \tag{4}$$

PROOF. Let

$$u_t = y_t - \boldsymbol{\beta}' \mathbf{x}_t\,, \tag{5}$$

which has expected value 0 and variance σ^2. Then

$$\mathbf{b}_T - \boldsymbol{\beta} = \mathbf{A}_T^{-1} \textstyle\sum_{t=1}^T \mathbf{x}_t u_t\,. \tag{6}$$

We shall show (6) converges to $\mathbf{0}$ with probability 1 if (4) holds. Consider first the case of $p = 1$. Then the scalar $\sum_{t=1}^T x_t u_t$ is a martingale with $\sigma^2 A_T = \sum_{t=1}^T \mathscr{E}[(x_t u_t)^2 \mid u_{t-1}, u_{t-2}, \cdots, u_1]$ diverging to infinity. Under these conditions

Received June 1975; revised December 1975.

[1] This work was supported by National Science Foundation Grant SOC73-05453 at the Institute for Mathematical Studies in the Social Sciences, Stanford University.

AMS 1970 *subject classifications.* Primary 62J05; Secondary 60F15.

Key words and phrases. Least squares estimates, linear regression, strong consistency.

(e.g., Neveu (1965), page 150) we have that $A_T^{-\frac{1}{2}}(\log A_T)^{-\frac{1}{2}-\varepsilon} \sum_{t=1}^T x_t u_t$ converges to 0 for every $\varepsilon > 0$ with probability 1. Then equation (6) converges to 0 with probability 1, since $A_T^{-\frac{1}{2}}(\log A_T)^{\frac{1}{2}+\varepsilon} \to 0$ as $T \to \infty$ for $\varepsilon = \frac{1}{2}$, say. Next, take $p > 1$ and consider the first component of $\mathbf{b}_T - \boldsymbol{\beta}$. Let

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \boldsymbol{\beta}^{(2)} \end{pmatrix}, \qquad \mathbf{x}_t = \begin{pmatrix} x_{1t} \\ \mathbf{x}_t^{(2)} \end{pmatrix}, \tag{7}$$

$$\mathbf{b}_T = \begin{pmatrix} b_{1T} \\ \mathbf{b}_T^{(2)} \end{pmatrix}, \qquad \mathbf{A}_T = \begin{pmatrix} a_{11T} & \mathbf{A}_{12T} \\ \mathbf{A}_{21T} & \mathbf{A}_{22T} \end{pmatrix}. \tag{8}$$

Then

$$b_{1T} - \beta_1 = \frac{Y_T}{S_T}, \tag{9}$$

where

$$Y_T = \sum_{t=1}^T (x_{1t} - \mathbf{A}_{12T}\mathbf{A}_{22T}^{-1}\mathbf{x}_t^{(2)})u_t, \tag{10}$$

$$S_T = \sum_{t=1}^T (x_{1t} - \mathbf{A}_{12T}\mathbf{A}_{22T}^{-1}\mathbf{x}_t^{(2)})^2. \tag{11}$$

(See, for example, Section 2.3 of Anderson (1971).) The variance of Y_T is $\sigma^2 S_T$. Define $\gamma_1^2 = S_p$, $v_1 = Y_p/\gamma_1$. For $T \geqq p$ consider

$$\begin{aligned} Y_{T+1} = Y_T &+ (\mathbf{A}_{12T}\mathbf{A}_{22T}^{-1} - \mathbf{A}_{12,T+1}\mathbf{A}_{22,T+1}^{-1}) \textstyle\sum_{t=1}^T \mathbf{x}_t^{(2)}u_t \\ &+ (x_{1,T+1} - \mathbf{A}_{12,T+1}\mathbf{A}_{22,T+1}^{-1}\mathbf{x}_{T+1}^{(2)})u_{T+1}. \end{aligned} \tag{12}$$

The second term on the right-hand side of (12) is uncorrelated with Y_R, $R = p, p+1, \cdots, T$, because

$$\sum_{t=1}^R (x_{1t} - \mathbf{A}_{12R}\mathbf{A}_{22R}^{-1}\mathbf{x}_t^{(2)})\mathbf{x}_t^{(2)\prime} = \mathbf{0}. \tag{13}$$

The third term is uncorrelated with Y_R, $R = p, p+1, \cdots, T$, because u_{T+1} is uncorrelated with $u_1, \cdots, u_T$. Thus the increment $Y_{T+1} - Y_T$ is uncorrelated with $Y_p, Y_{p+1}, \cdots, Y_T$. Define

$$\gamma_t^2 = S_{t+p-1} - S_{t+p-2}, \qquad t = 2, 3, \cdots, \tag{14}$$

$$v_t = \frac{Y_{t+p-1} - Y_{t+p-2}}{\gamma_t}, \qquad t = 2, 3, \cdots. \tag{15}$$

Then $v_1, v_2, \cdots$ constitute a sequence of independent random variables, each with distribution $N(0, \sigma^2)$. Since

$$\sum_{t=1}^{T-p+1} \gamma_t^2 = S_T = \sum_{t=1}^T (x_{1t} - \mathbf{A}_{12T}\mathbf{A}_{22T}^{-1}\mathbf{x}_t^{(2)})^2 = a_{11T} - \mathbf{A}_{12T}\mathbf{A}_{22T}^{-1}\mathbf{A}_{21T} \tag{16}$$

is the reciprocal of the upper left hand corner of $\mathbf{A}_T^{-1}$, it diverges to ∞ as $T \to \infty$. Then the argument for $p = 1$ above shows that

$$b_{1T} - \beta_1 = \frac{\sum_{t=1}^{T-p+1} \gamma_t v_t}{\sum_{t=1}^{T-p+1} \gamma_t^2} \tag{17}$$

converges to 0 with probability 1.

On the other hand $b_{1T} - \beta_1$ converges to 0 in probability only if its variance

σ^2/S_T converges to 0 since it is normally distributed. Because $\mathbf{A}_T^{-1}$ is positive definite, (4) holds if and only if every diagonal element of $\mathbf{A}_T^{-1}$ converges to 0. □

For $p = 1$ the theorem holds under more general conditions.

THEOREM 2. *Let* $u_1, u_2, \cdots$ *and* $x_1, x_2, \cdots$ *be sequences of scalar random variables such that* $\mathscr{E}(u_t \mid u_1, \cdots, u_{t-1}, x_1, \cdots, x_t) = 0$ *and* $\mathscr{E}(u_t^2 \mid u_1, \cdots, u_{t-1}, x_1, \cdots, x_t) = \sigma^2 < \infty$. *Let* $y_t = \beta x_t + u_t$, $t = 1, 2, \cdots$. *If* x_1 *is bounded away from* 0 *with probability* 1 *and if* $\sum_{t=1}^T x_t^2 \to \infty$ *with probability* 1, *then* $\hat{\beta}_T = \sum_{t=1}^T x_t y_t / \sum_{t=1}^T x_t^2 \to \beta$ *with probability* 1.

This theorem was implicit in Taylor (1974)

Alternative forms of (4) are (i) the maximum characteristic root of $\mathbf{A}_T^{-1}$ converges to 0, (ii) the minimum characteristic root of A_T diverges to ∞ and (iii) $\boldsymbol{\gamma}'\mathbf{A}_T\boldsymbol{\gamma} \to \infty$ for every $\boldsymbol{\gamma} \neq \mathbf{0}$.

Brown, Durbin and Evans (1975) have defined "recursive residuals," which are equivalent to $v_1, v_2, \cdots$ here, and used them in a different context.

REFERENCES

ANDERSON, T. W. (1971). *The Statistical Analysis of Time Series*. Wiley, New York.

BROWN, R. L., DURBIN, J. and EVANS, J. M. (1975). Techniques for testing the consistency of regression relationships over time. *J. Roy. Statist. Soc. Ser. B* **37** 149-163.

NEVEU, J. (1965). *Mathematical Foundations of the Calculus of Probability*. Holden-Day, San Francisco.

TAYLOR, J. B. (1974). Asymptotic properties of multiperiod control rules in the linear regression model. *Internat. Econom. Rev.* **15** 472-484.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305

DEPARTMENT OF ECONOMICS
COLUMBIA UNIVERSITY
NEW YORK, NEW YORK 10027