The Annals of Statistics
1979, Vol. 7, No. 3, 484–489

STRONG CONSISTENCY OF LEAST SQUARES ESTIMATES IN DYNAMIC MODELS[1]

BY T. W. ANDERSON AND JOHN B. TAYLOR

Stanford University and Columbia University

The least squares estimate of the parameter matrix $\mathbf{B}$ in the model $\mathbf{y}_t = \mathbf{B}'\mathbf{x}_t + \mathbf{u}_t$, where $\mathbf{u}_t$ is an m-component vector of unobservable disturbances and $\mathbf{x}_t$ is a p-component vector, converges to $\mathbf{B}$ with probability one under certain conditions on the behavior of $\mathbf{x}_t$ and $\mathbf{u}_t$. When $\mathbf{x}_t$ is stochastic and the conditional expectation of $\mathbf{u}_t$ given $\mathbf{x}_s$ for $s \leq t$ and $\mathbf{u}_s$ for $s < t$ is zero, then the least squares estimates are strongly consistent if the inverse of $\mathbf{A}_T = \Sigma_{t=1}^T \mathbf{x}_t\mathbf{x}'_t$, where T is the sample size, converges to the zero matrix and if the ratio of the largest to the smallest characteristic root of $\mathbf{A}_T$ is bounded with probability one.

Many statistical problems are concerned with estimating the parameter matrix $\mathbf{B}$ in the model

$$\mathbf{y}_t = \mathbf{B}'\mathbf{x}_t + \mathbf{u}_t, \tag{1}$$

where $\mathbf{u}_t$ is an m-component vector of unobservable disturbances and $\mathbf{x}_t$ is a p-component vector, $t = 1, 2, \cdots$. The model (1) is a first-order vector autoregressive model when $\mathbf{x}_t = \mathbf{y}_{t-1}$, and a sequential control or sequential design model when $\mathbf{x}_t$ is a function of $\mathbf{x}_s$ and $\mathbf{y}_s$ for $s < t$. In the special case where $\mathbf{x}_t$ is a set of nonstochastic regressors and the $\mathbf{u}_t$'s are uncorrelated, the model reduces to the classical linear regression model. The least squares estimate of $\mathbf{B}$ based on T observations is given by $\hat{\mathbf{B}}_T = \mathbf{A}_T^{-1}\Sigma_{t=1}^T \mathbf{x}_t\mathbf{y}'_t$, where $\mathbf{A}_T = \Sigma_{t=1}^T \mathbf{x}_t\mathbf{x}'_t$.

Here we prove that $\hat{\mathbf{B}}_T \to \mathbf{B}$ with probability one under a specified set of conditions on the asymptotic behavior of $\mathbf{x}_t$ which covers many of the dynamic applications mentioned above. The conditions are that $\mathbf{A}_T^{-1} \to \mathbf{0}$ with probability one and the ratio of the largest to the smallest characteristic root of $\mathbf{A}_T$ is bounded with probability one.

THEOREM 1. *Let* $\mathbf{y}_t = \mathbf{B}'\mathbf{x}_t + \mathbf{u}_t$, *where* $\mathbf{B}$ *is a* $p \times m$ *matrix of parameters*, $\mathbf{x}_t$ *is a* p-*component vector, and* $\mathbf{u}_t$ *is an* m-*component* (*unobservable*) *stochastic vector*, $t = 1, 2, \cdots$. *Let* $\mathcal{F}_t$ *be the* σ-*algebra generated by* $(\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_t, \mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_{t+1})$, $t = 1, 2, \cdots$, *and let* $\mathcal{F}_0$ *be the* σ-*algebra generated by* $\mathbf{x}_1$. *Let* $\mathbf{A}_t = \Sigma_{s=1}^t \mathbf{x}_s\mathbf{x}'_s$ *and suppose that* $\mathbf{A}_q$ *is nonsingular with probability one and that* $\mathcal{E}\,\mathrm{tr}\,\mathbf{A}_q^{-1} < \infty$ *for some* $q \geq p$. *Define* $\hat{\mathbf{B}}_T = \mathbf{A}_T^{-1}\Sigma_{t=1}^T \mathbf{x}_t\mathbf{y}'_t$. *If*

(i) $\mathcal{E}(\mathbf{u}_t|\mathcal{F}_{t-1}) = \mathbf{0}$ *and* $\mathcal{E}(\mathbf{u}_t\mathbf{u}'_t|\mathcal{F}_{t-1}) = \mathbf{\Sigma}$, $t = 1, 2, \cdots$, *with probability one*,

---

Received September 1976; revised April 1978.

[1]This work was supported by National Science Foundation Grants SOC73-05453 at the Institute for Mathematical Studies in the Social Sciences, Stanford University and SOC74-13376 at Columbia University.

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

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

(ii) $\lim_{T\to\infty}\mathbf{A}_T^{-1} = \mathbf{0}$ *with probability one,*

(iii) *the ratio of the largest to the smallest characteristic roots of* $\mathbf{A}_T$ *is bounded uniformly in* T *with probability one,*

then $\lim_{T\to\infty}\hat{\mathbf{B}}_T = \mathbf{B}$ *with probability one.*

PROOF. We want to show that $\hat{\mathbf{B}}_T - \mathbf{B} = \mathbf{A}_T^{-1}\Sigma_{t=1}^T \mathbf{x}_t\mathbf{u}_t'$ converges to $\mathbf{0}$ with probability 1. The proof for $p > 1$ and $m > 1$ is based on the scalar case $p = 1$ and $m = 1$. Hence we consider the scalar case first.

The sum $z_T = \Sigma_{t=1}^T A_t^{-1}x_t u_t$ is a martingale because $\mathcal{E}[A_t^{-1}x_t u_t|\mathcal{F}_{t-1}] = A_t^{-1}x_t\mathcal{E}(u_t|\mathcal{F}_{t-1}) = 0$ with probability one since $A_t^{-1}x_t$ is bounded with probability one. The variance of $z_T - z_q$ (which is a martingale) is

$$
\begin{aligned}
(2)\qquad \mathcal{E}(z_T - z_q)^2 &= \mathcal{E}\Sigma_{r,s=q+1}^T (A_r A_s)^{-1}x_r x_s u_r u_s \\
&= \mathcal{E}\left[\Sigma_{s=q+1}^T A_s^{-2}x_s^2u_s^2 + \Sigma_{r=q+2}^T\Sigma_{s=q+1}^{r-1}(A_rA_s)^{-1}x_rx_su_ru_s \right.\\
&\qquad \left. + \Sigma_{s=q+2}^T\Sigma_{r=q+1}^{s-1}(A_rA_s)^{-1}x_rx_su_ru_s\right] \\
&= \mathcal{E}\Sigma_{s=q+1}^T A_s^{-2}x_s^2\Sigma \leqslant \mathcal{E}A_q^{-1}\Sigma.
\end{aligned}
$$

The third equality follows from the fact that, for $s < r$

$$(3)\qquad \mathcal{E}(A_rA_s)^{-1}x_rx_su_ru_s = \mathcal{E}(A_rA_s)^{-1}x_rx_su_s\mathcal{E}(u_r|\mathcal{F}_{r-1}) = 0.$$

The inequality in (2) follows from Lemma 1 of Taylor (1974). (See also Neveu (1965), page 150.). Since $\mathcal{E}A_q^{-1} < \infty$ by assumption, the sum $z_T - z_q$ is a martingale with a bounded variance, and, by the martingale convergence theorem, converges with probability one (Feller (1966) page 236). The convergence of $A_T^{-1}\Sigma_{t=1}^T x_t u_t$ to zero with probability one, therefore, follows from Kronecker's lemma.

For $p > 1$ and $m \geqslant 1$ we write the ith column of $\hat{\mathbf{B}}_T - \mathbf{B}$ as

$$(4)\qquad \mathbf{A}_T^{-1}\Sigma_{t=1}^T\mathbf{x}_t u_{it} = \left[(\operatorname{tr}\mathbf{A}_T)^{-1}\mathbf{A}_T\right]^{-1}\left[(\operatorname{tr}\mathbf{A}_T)^{-1}\Sigma_{t=1}^T\mathbf{x}_t u_{it}\right].$$

The jth component of the vector $(\operatorname{tr}\mathbf{A}_T)^{-1}\Sigma_{t=1}^T\mathbf{x}_t u_{it}$ on the right hand side of (5) is

$$(5)\qquad \frac{\Sigma_{t=1}^T x_{jt}u_{it}}{\Sigma_{h=1}^p\Sigma_{t=1}^T x_{ht}^2},$$

whose absolute value is less than the absolute value of $\Sigma_{t=1}^T x_{jt}u_{it}/\Sigma_{t=1}^T x_{jt}^2$. Since the reciprocal of the jth diagonal element of $\mathbf{A}_q$ is less than the jth diagonal element of $\mathbf{A}_q^{-1}$ (Anderson and Taylor (1976a), for example), the assumption $\mathcal{E}\operatorname{tr}A_q^{-1} < \infty$ implies $\mathcal{E}\left(\Sigma_{t=1}^q x_{jt}^2\right)^{-1} < \infty$. Hence, by the case of $p = m = 1$ of the theorem $\Sigma_{t=1}^T x_{jt}u_{it}/\Sigma_{t=1}^T x_{jt}^2$ converges to 0 with probability 1 and so does (5).

To complete the proof we must show that the elements of $[(\operatorname{tr}\mathbf{A}_T)^{-1}\mathbf{A}_T]^{-1}$ are bounded with probability one. Let λ_i be the ith characteristic root of $\mathbf{A}_T$, and let λ_s and λ_l be the smallest and largest roots of $\mathbf{A}_T$, respectively. Then the ith characteristic root of $[(\operatorname{tr}\mathbf{A}_T)^{-1}\mathbf{A}_T]^{-1}$ is equal to $\lambda_i^{-1}\Sigma_{j=1}^p\lambda_j \leqslant p\lambda_i^{-1}\lambda_l \leqslant p\lambda_s^{-1}\lambda_l$. Hence, by assumption (iii) the roots, and therefore the elements, of $[(\text{tr}\, \mathbf{A}_T)^{-1}\mathbf{A}_T]^{-1}$ are bounded with probability one. □

The following lemma shows that condition (ii) in Theorem 1 can be replaced by alternative conditions.

LEMMA 1. *Let* $\mathbf{A}_1, \mathbf{A}_2 \cdots$ *be a sequence of positive semidefinite matrices such that* $\mathbf{A}_t - \mathbf{A}_{t-1}$ *is positive semidefinite and*

$$\gamma'\mathbf{A}_t\gamma \to \infty \tag{6}$$

for every $\gamma \neq 0$. *Then*

$$\lim_{t\to\infty}\mathbf{A}_t^{-1} = \mathbf{0}. \tag{7}$$

PROOF. Condition (6) implies $\mathbf{A}_t$ is nonsingular for sufficiently large t since $\mathbf{A}_t\gamma \neq \mathbf{0}$ for a linearly independent set of γ. Let $f_t(\gamma) = (\gamma'\mathbf{A}_t\gamma)^{-1}$ for $\gamma'\gamma = 1$, Then, for every $\gamma, f_t(\gamma)$ is a non-increasing sequence such that

$$\lim_{t\to\infty} f_t(\gamma) = 0. \tag{8}$$

By Dini's theorem the convergence (8) is uniform and hence

$$\max_{\gamma'\gamma=1} f_t(\gamma) = \max_{\gamma'\gamma=1}(\gamma'\mathbf{A}_t\gamma)^{-1} \to 0; \tag{9}$$

that is, the maximum characteristic root of $\mathbf{A}_t^{-1}$ (which is the reciprocal of the minimum characteristic root of $\mathbf{A}_t$) converges to 0, which implies (7). □

Lemma 1 and its proof imply that condition (ii) of Theorem 1 can be replaced by: (iia) $\gamma'\mathbf{A}_T\gamma \to \infty$ for every $\gamma \neq \mathbf{0}$ with probability one; (iib) the smallest characteristic root of $\mathbf{A}_T$ diverges to infinity with probability one; or (iic) the largest characteristic root of $\mathbf{A}_T^{-1}$ converges to zero with probability one.

Theorem 1 applies to sequential control or sequential design problems in which the regressors $\mathbf{x}_t$ are chosen according to observations on $\mathbf{x}_s$ and $\mathbf{y}_s$ for $s < t$. The primary advantage of Theorem 1 is that $\mathbf{x}_t$ may be generated by a very general stochastic structure which does not need to be specified as long as the stated conditions, or their alternatives, are satisfied. These conditions will certainly be satisfied if $T^{-1}\mathbf{A}_T$ converges to a nonsingular matrix (with probability one), a frequent assumption for weak consistency in regression models with fixed regressors. However, the conditions will also be satisfied with more general asymptotic behavior of $\mathbf{x}_t$.

Another implication of Theorem 1 is that the least squares estimates in the stable vector autoregressive model are consistent under more general assumptions than have been made previously (for example, Anderson (1971), Section 5.5), and further that these estimates are consistent in the stronger sense. In this case consistency of least squares follows directly from the strong consistency of the moment matrix. To see this let $\mathbf{x}_t = \mathbf{y}_{t-1}$ in equation (1). Then the following lemma shows that conditions (ii) and (iii) of Theorem 1 are satisfied. For convenience we consider the case where $\mathbf{y}_0$ has the same first and second moments as the stationary distribution of $\mathbf{y}_t$. This implies that all of $\mathbf{y}_0, \mathbf{y}_1, \mathbf{y}_2, \cdots$ have the same mean vectors and covariance matrices.

LEMMA 2. *Let* $\mathbf{y}_t = \mathbf{B}'\mathbf{y}_{t-1} + \mathbf{u}_t$, $t = 1, 2, \cdots$, *and assume that the characteristic roots of* $\mathbf{B}$ *are less than one in absolute value. Let* $\mathcal{F}_t$ *be the* σ*-algebra generated by* $(u_1, u_2, \cdots, u_t, y_0, y_1, \cdots, y_t)$, $t = 1, 2, \cdots$, *and let* $\mathcal{F}_0$ *be the* σ*-algebra generated by* $\mathbf{y}_0$. *Assume that* $\mathcal{E}(\mathbf{u}_t|\mathcal{F}_{t-1}) = \mathbf{0}$, $\mathcal{E}(\mathbf{u}_t\mathbf{u}_t'|\mathcal{F}_{t-1}) = \mathbf{\Sigma}$, $\mathcal{E}\mathbf{y}_0 = \mathbf{0}$, *and* $\mathcal{E}(\mathbf{y}_0\mathbf{y}_0') = \mathbf{\Gamma}$, *where* $\mathbf{\Gamma}$ *is the solution of* $\mathbf{\Gamma} - \mathbf{B}'\mathbf{\Gamma}\mathbf{B} = \mathbf{\Sigma}$. *Further assume that* $\sum_{t=1}^{\infty} t^{-2}\mathcal{E}y_{it}^4 < \infty$ *and* $\sum_{t=1}^{\infty} t^{-2}\mathcal{E}u_{it}^4 < \infty$, $i = 1, \cdots, p$, *where* y_{it} *is the* ith *component of* $\mathbf{y}_t$ *and* u_{it} *is the* ith *component of* $\mathbf{u}_t$. *Then* $\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T}\mathbf{y}_t\mathbf{y}_t' = \mathbf{\Gamma}$ *with probability one.*

PROOF. The fourth-order conditions on $\{\mathbf{u}_t\}$ imply $\sum_{t=1}^{\infty} t^{-2}\mathcal{E}u_{it}^2u_{jt}^2 < \infty$ (by the Cauchy-Schwarz inequality). Hence by the law of large numbers for martingales, (Feller (1966), page 238) $\sum_{t=1}^{T} t^{-1}u_{it}u_{jt}$ converges with probability one and $T^{-1}\sum_{t=1}^{T} u_{it}u_{jt} \to \sigma_{ij}$ with probability one, where σ_{ij} is an element of $\mathbf{\Sigma}$. Alternatively stated, $T^{-1}\sum_{t=1}^{T}\mathbf{u}_t\mathbf{u}_t' \to \mathbf{\Sigma}$ with probability one. Further,

(10)

$$\begin{aligned} T^{-1}\textstyle\sum_{t=1}^{T}\mathbf{u}_t\mathbf{u}_t' &= T^{-1}\textstyle\sum_{t=1}^{T}(\mathbf{y}_t - \mathbf{B}'\mathbf{y}_{t-1})(\mathbf{y}_t - \mathbf{B}'\mathbf{y}_{t-1})' \\ &= T^{-1}\left[\textstyle\sum_{t=1}^{T}\mathbf{y}_t\mathbf{y}_t' - \mathbf{B}'\sum_{t=1}^{T}\mathbf{y}_{t-1}\mathbf{y}_t' - \sum_{t=1}^{T}\mathbf{y}_t\mathbf{y}_{t-1}'\mathbf{B} + \mathbf{B}'\sum_{t=1}^{T}\mathbf{y}_{t-1}\mathbf{y}_{t-1}'\mathbf{B}\right] \\ &= T^{-1}\textstyle\sum_{t=1}^{T}\mathbf{y}_t\mathbf{y}_t' - \mathbf{B}'\left(T^{-1}\sum_{t=1}^{T}\mathbf{y}_t\mathbf{y}_t'\right)\mathbf{B} + \mathbf{B}'\left(T^{-1}\sum_{t=1}^{T}\mathbf{y}_{t-1}\mathbf{u}_t'\right) \\ &\quad + \left(T^{-1}\textstyle\sum_{t=1}^{T}\mathbf{u}_t\mathbf{y}_{t-1}'\right)\mathbf{B} - \mathbf{B}'T^{-1}(\mathbf{y}_0\mathbf{y}_0' - \mathbf{y}_T\mathbf{y}_T')\mathbf{B}. \end{aligned}$$

Each element in the third and fourth terms on the right-hand side of (10) is a linear combination of terms of the form $T^{-1}\sum_{t=1}^{T} y_{i,t-1}u_{jt}$. Under the conditions of the lemma $\sum_{t=1}^{T} t^{-1}y_{i,t-1}u_{jt}$ is a martingale with variance equal to $\gamma_{ii}\sigma_{ii}\sum_{t=1}^{T} t^{-2}$, where γ_{ii} is the ith diagonal element of $\mathbf{\Gamma}$. Therefore $\sum_{t=1}^{T} t^{-1}y_{i,t-1}u_{jt}$ converges with probability one, and, by Kronecker's lemma, $T^{-1}\sum_{t=1}^{T} y_{i,t-1}u_{jt}$ converges to zero with probability one. The condition $\sum_{t=1}^{\infty} t^{-2}\mathcal{E}y_{it}^4$ implies $\sum_{t=1}^{\infty} t^{-2}\mathcal{E}y_{it}^2y_{jt}^2$, and hence $\sum_{t=1}^{T} t^{-1}y_{it}y_{jt}$ converges with probability one and $y_{it}y_{jt}/t$ converges to zero with probability one. These results imply that the last three terms in (10) converge to zero with probability one. Thus

$$(11) \qquad \left(\lim_{T\to\infty} T^{-1}\textstyle\sum_{t=1}^{T}\mathbf{y}_t\mathbf{y}_t'\right) - \mathbf{B}'\left(\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T}\mathbf{y}_t\mathbf{y}_t'\right)\mathbf{B} = \mathbf{\Sigma}$$

with probability one. Hence, $\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T}\mathbf{y}_t\mathbf{y}_t' = \mathbf{\Gamma}$ with probability one. □

If $\mathbf{\Gamma}$ is positive definite, Lemma 2 implies that $\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T}\mathbf{y}_t\mathbf{y}_t'$ is positive definite, and therefore $\hat{\mathbf{B}}_T - \mathbf{B} = (T^{-1}\sum_{t=1}^{T}\mathbf{y}_{t-1}\mathbf{y}_{t-1}')^{-1}T^{-1}\sum_{t=1}^{T}\mathbf{y}_{t-1}\mathbf{u}_t'$ converges to $\mathbf{0}$ with probability one. The matrix $\mathbf{\Gamma}$ will be positive definite if $\mathbf{\Sigma}$ is positive definite or if model (1) represents a pth order scalar stochastic difference equation (see Anderson (1971), pages 196–197). The conclusions for the scalar process of arbitrary order also follow from the results of Hannan and Heyde (1972). Note that Lemma 2 holds if the fourth-order moments of $\mathbf{y}_t$ and $\mathbf{u}_t$ exist and do not depend on t.

For $p = 1$, Theorem 1 was implicit in Taylor (1974). Drygas (1976) proved the strong convergence of $\hat{\mathbf{B}}_T$ in the multivariate regression model by assuming that $\mathbf{A}_T^{-1}\mathbf{D}_T$ is bounded with probability one, where $\mathbf{D}_T$ is a diagonal matrix with $\Sigma_{t=1}^T x_{it}^2$, $i = 1, \cdots, p$, on the diagonal. Scalar theorems can then be used to prove that $\mathbf{D}_T^{-1}\Sigma_{t=1}^T \mathbf{x}_t\mathbf{u}_t' \to \mathbf{0}$ with probability one, if $\mathbf{D}_T \to \infty$ with probability one. With $\mathbf{A}_T^{-1}\mathbf{D}_T$ bounded with probability one, $\hat{\mathbf{B}}_T$ must therefore converge to $\mathbf{B}$ with probability one.

Anderson and Taylor (1976a) showed that if $\mathbf{x}_t$ is nonstochastic and the $\mathbf{u}_t$ are independently normally distributed with zero mean, then $\mathbf{A}_T^{-1} \to \mathbf{0}$ is necessary and sufficient for the strong consistency of $\hat{\mathbf{B}}_T$. Lai and Robbins (1977) showed that normality could be replaced by the boundedness of $\mathcal{E}[u_t^2(\log(1 + |u_t|)^r]$ for some $r > 1$ as a sufficient condition for the case where $p = 2$ and $x_{1t} = 1$. An unsettled question is whether conditions (i) and (ii) are sufficient for Theorem 1.

Alternative conditions on the sequences $\{\mathbf{u}_t\}$ and $\{\mathbf{x}_t\}$ when $\mathbf{x}_t$ is nonstochastic yield other results.

THEOREM 2. *Let* $\mathbf{y}_t = \mathbf{B}'\mathbf{x}_t + \mathbf{u}_t$, $t = 1, 2, \cdots$, *where* $\mathbf{B}'$ *is an* $m \times p$ *matrix of parameters, the* $\mathbf{x}_t$'s *are p-component nonstochastic vectors, and the* $\mathbf{u}_t$'s *are m-component independently and identically distributed stochastic vectors with zero mean and finite covariance matrix. If the smallest characteristic root of* $T^{-1}\Sigma_{t=1}^T \mathbf{x}_t\mathbf{x}_t'$ *is bounded away from zero, then* $\hat{\mathbf{B}}_T = (\Sigma_{t=1}^T \mathbf{x}_t\mathbf{x}_t')^{-1}\Sigma_{t=1}^T \mathbf{x}_t\mathbf{y}_t' \to \mathbf{B}$ *with probability one.*

The theorem follows from a result of Chow (1966), pages 1484–1485, to the effect that if $u_1, u_2, \cdots$ are independently and identically distributed with $\mathcal{E}u_t = 0$ and $\mathcal{E}u_t^2 = \sigma^2 < \infty$, $t = 1, 2, \cdots$ and a_{Tt}, $t = 1, \cdots, T$ is an array of real numbers such that $\Sigma_{t=1}^T a_{Tt}^2 = 1$, $T = 1, 2, \cdots$, then

$$\lim_{T\to\infty} T^{-1/2}\Sigma_{t=1}^T a_{Tt}u_t = 0 \tag{12}$$

with probability one. Gleser (1966) used this result to show that the usual estimate of a diagonal element of $\mathbf{\Sigma}$ is strongly consistent.

THEOREM 3. *Let* $\mathbf{y}_t = \mathbf{B}'\mathbf{x}_t + \mathbf{u}_t$, *where* $\mathbf{B}'$ *is an* $m \times p$ *matrix of parameters,* $\mathbf{x}_t$ *is a p-component nonstochastic vector and* $\mathbf{u}_t$ *is an m-component stochastic vector which is independently and identically distributed and whose components are generalized Gaussian. If* $\operatorname{tr}(\Sigma_{t=1}^T \mathbf{x}_t\mathbf{x}_t')^{-1} = o(\log^{-1} T)$, *then* $\lim_{T\to\infty}\hat{\mathbf{B}}_T = \mathbf{B}$ *with probability one.*

A random variable u is said to be *generalized Gaussian* if there exists an $\alpha \geqslant 0$ such that for every real number r, $\mathcal{E}e^{ru} \leqslant e^{\alpha^2 r^2/2}$. Theorem 3 follows from Chow's result (1966), page 1484, that if $\{u_t\}$ is a sequence of independent generalized Gaussian random variables with $\mathcal{E}u_t = 0$, w_{Tt}, $t = 1, \cdots, T$, form an array such that $\Sigma_{t=1}^T w_{Tt}^2 = o(\log^{-1} T)$, $T = 1, 2, \cdots$, then $\lim_{T\to\infty}\Sigma_{t=1}^T w_{Tt}u_t = 0$ with probability one. These two theorems were proved by Anderson and Taylor (1976b). They and other similar results are also consequences of theorems in Stout (1974).

Acknowledgment. We thank the referee for suggesting the proof of Theorem 1 reported above. An alternative proof appears in Anderson and Taylor (1976b). The

alternative proof is based on a multivariate generalization of Kronecker's lemma which was proved and shown, through a counter-example, to depend on a special multivariate condition by Anderson and Taylor (1974) and, independently, by B. D. O. Anderson and J. B. Moore (1976).

REFERENCES

ANDERSON, B. D. O. and MOORE, J. B. (1976). A matrix Kronecker lemma. *Linear Algebra Appl.* **15** 227–234.

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

ANDERSON, T. W. and TAYLOR, J. B. (1974). Some theorems concerning strong consistency of least squares estimates in linear regression models. Working Paper No. 47, Inst. Math. Studies Social Sci. Stanford Univ.

ANDERSON, T. W. and TAYLOR, J. B. (1976a). Strong consistency of least squares estimates in normal linear regression. *Ann. Statist.* **4** 788–790.

ANDERSON, T. W. and TAYLOR, J. B. (1976b). Conditions for strong consistency of least squares estimates in linear models. Technical Report No. 213, Inst. Math. Studies Social Sci., Stanford Univ.

CHOW, Y. S. (1966). Some convergence theorems for independent random variables. *Ann. Math. Statist.* **37** 1482–1493.

DRYGAS, H. (1976). Weak and strong consistency of the least squares estimators in regression models. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **34** 119–127.

FELLER, W. (1966). *An Introduction to Probability Theory and Its Applications II*. Wiley, New York.

GLESER, L. J. (1966). Correction to 'On the asymptotic theory of fixed size sequential confidence bounds for linear regression parameters'. *Ann. Math. Statist.* **37** 1053–1055.

HANNAN, E. J. and HEYDE, C. C. (1972). On limit theory for quadratic functions of discrete time series. *Ann. Math. Statist.* **43** 2058–2066.

LAI, T. L. and ROBBINS, H. (1977). Strong consistency of least squares estimates in regression models. *Proc. Nat. Acad. Sci.* **74** 2667–2669.

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

STOUT, W. F. (1974). *Almost Sure Convergence*. Academic Press, New York.

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

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305

DEPARTMENT OF ECONOMICS
COLUMBIA UNIVERSITY
NEW YORK, N.Y. 10027