The Annals of Statistics
1979, Vol. 7, No. 4, 795–800

A CONVERGENCE THEOREM FOR RANDOM LINEAR COMBINATIONS OF INDEPENDENT NORMAL RANDOM VARIABLES[1]

BY N. CHRISTOPEIT AND K. HELMES

University of Bonn

It is shown that under fairly mild conditions linear combinations of independent normally distributed random variables with random coefficients tend to zero almost everywhere. The result is applied to parameter estimation in linear regression models.

0. Introduction. Let $\varepsilon(n)$ $(n = 1, 2, \cdots)$ be a sequence of random variables and $x^{(N)}(n)$ $(n = 1, 2, \cdots, N; N = 1, 2, \cdots)$ a triangular array of $K \times 1$ random vectors defined on some probability space $(\Omega, \mathfrak{F}, P)$. Consider the linear regression model

$$y(N) = X(N)\beta + u(N), \qquad N = 1, 2, \cdots, \tag{0.1}$$

where

$$u(N) = (\varepsilon(1), \cdots, \varepsilon(N))', \qquad X(N) = (x^{(N)}(1), \cdots, x^{(N)}(N))'$$

and β is a $K \times 1$ parameter vector. Model (0.1) describes the situation of 'sampling without repetition', i.e., when the sample size increases not only are new observations added but also former observations may change in value. Mathematically speaking: the observation matrix $X(N)$ may in general not be considered as a submatrix of $X(N')$ for $N' > N$. This situation arises typically when the error terms are autocorrelated (compare Section 2).

The sampling error of the ordinary least squares (OLS) estimate $\hat{\beta}(N)$ of β (based upon the first N observations) is given by

$$\hat{\beta}(N) - \beta = [X(N)'X(N)]^{-1}X(N)'u(N). \tag{0.2}$$

The question of strong consistency of the OLS-estimator then leads to the investigation of the limiting behavior of sums $z(N) = \sum_{n=1}^{N} x^{(N)}(n)\varepsilon(n)$ in the sense of convergence almost everywhere (a.e.).

Sums of this type have been investigated by Chow [2] and Tomkins [6] for the case of nonstochastic coefficients $x^{(N)}(n)$. On the basis of Chow's results Anderson and Taylor [1] have shown strong consistency of the OLS-estimator in linear regression models with deterministic regressors and independent identically distributed error terms which are generalized Gaussian. Our Theorem 1 may be regarded as an extension of Chow's Theorem 2 to the case of stochastic coefficients. If specialized to the nonstochastic case the hypotheses made in [1] concerning the

Received August 1977; revised April 1978.

[1]This work was supported by the Sonderforschungsbereiche 21 and 72 at the University of Bonn.

AMS 1970 *subject classifications.* Primary 60F15, 60G50; secondary 62J05.

Key words and phrases. Strong law of large numbers, regression analysis.

regressor matrices are obtained. The case of stochastic regressors and 'sampling with repetition' has been dealt with in [1] and [3]. However, the techniques used there do not seem to carry over to the setting of 'sampling without repetition'.

In Section 1 convergence of $z(N)$ to 0 a.e. is established for the case of independent $N(0, \sigma^2)$ distributed disturbances. The proof is based on the observation that almost sure convergence of $z(N)$ to a constant—say 0—is a property modulo stochastic equivalence. This is to say: if $\tilde{z}(N)$ $(N = 1, 2, \cdots)$ is a process defined on some probability space $(\tilde{\Omega}, \tilde{\mathfrak{F}}, \tilde{P})$ whose finite dimensional distributions are the same as those of $z(N)$ $(N = 1, 2, \cdots)$ then $\tilde{z}(N) \to 0$ a.e. implies $z(N) \to 0$ a.e. and vice versa. This follows immediately from the definition of almost sure convergence. In Section 2 this result is applied to show strong consistency of the OLS estimator in linear regression models with 'sampling without repetition'.

1. A strong law of large numbers. Consider a sequence $\varepsilon(n)$ $(n = 1, 2, \cdots)$ together with a triangular array $a^{(N)}(n)$ $(n = 1, 2, \cdots, N; N = 1, 2, \cdots)$ of random variables defined on some probability space $(\Omega, \mathfrak{F}, P)$. Our object of study is the sequence of sums

$$z(N) = \Sigma_{n=1}^{N} a^{(N)}(n)\varepsilon(n), \qquad N = 1, 2, \cdots. \tag{1.1}$$

The assumptions are as follows:

(A) $\varepsilon(1), \varepsilon(2), \cdots$ are independent $N(0, \sigma^2)$ distributed.

(B) $a^{(N)}(n)$ is measurable with respect to $\mathfrak{F}\{\varepsilon(0), \cdots, \varepsilon(n-1)\}$ (the σ-algebra generated by $\varepsilon(0), \cdots, \varepsilon(n-1)$, with the convention that $\varepsilon(0) = 0$), $n = 1, 2, \cdots, N$; $N = 1, 2, \cdots$.

(C) For every $\delta > 0$ there exists a sequence of positive numbers δ_N, $N = 1, 2, \cdots$, such that $\Sigma_{N=1}^{\infty} \exp(-\delta^2/\delta_N^2) < \infty$ and $\Sigma_{N=1}^{\infty} P[A(N)^2 > \delta_N^2] < \infty$, where we have put $A(N)^2 = \Sigma_{n=1}^{N} a^{(N)}(n)^2$.

In the proof of the theorem we shall need a powerful bound for the integral of a simple nonanticipating functional with respect to Brownian motion. Let $\{w(t)\}_{0 \leqslant t \leqslant 1}$ be a Brownian motion defined on $(\tilde{\Omega}, \{\tilde{\mathfrak{F}}_t\}_{0 \leqslant t \leqslant 1}, \tilde{P})$ where $\{\tilde{\mathfrak{F}}_t\}$ is an increasing sequence of σ-fields. For a simple nonanticipating functional f let $\int_0^t f\,dw$ denote the stochastic integral in the sense of Ito (cf. McKean [5] for these notions and for a proof of the following lemma).

LEMMA 1. *For every nonanticipating simple functional f and arbitrarily chosen real numbers α, β*

$$P\left[\max_{0 \leqslant t \leqslant 1}\left(\int_0^t f\,dw - \frac{\alpha}{2}\int_0^t f^2\,ds\right) > \beta\right] \leqslant e^{-\alpha\beta}. \tag{1.2}$$

The main result is contained in

THEOREM 1. *Under assumptions* (A) – (C) $z(N) \to 0$ a.e. *as* $N \to \infty$.

PROOF. As pointed out in the introduction it suffices to show that $\tilde{z}(N) \to 0$ a.e. for some stochastically equivalent process $\tilde{z}$. Consider therefore a probability space

$(\tilde{\Omega}, \{\tilde{\mathfrak{F}}_t\}_{0 \leq t \leq 1}, \tilde{P})$ carrying a Brownian motion $\{w(t)\}_{0 \leq t \leq 1}$ with $E\{w(t)\} = 0$ and $E\{w(t)^2\} = \sigma^2 t$ for all t and put

$$\tilde{\varepsilon}(n) = (n(n+1))^{\frac{1}{2}}\left[w\left(1 - \frac{1}{n+1}\right) - w\left(1 - \frac{1}{n}\right)\right].$$

Since $a^{(N)}(n)$ is measurable with respect to $\mathfrak{F}\{\varepsilon(0), \cdots, \varepsilon(n-1)\}$ there exist Baire functions $g_n^{(N)} : \mathbb{R}^n \to \mathbb{R}$ such that

$$a^{(N)}(n) = g_n^{(N)}(\varepsilon(0), \cdots, \varepsilon(n-1)),$$
$$n = 1, 2, \cdots, N; \ N = 1, 2, \cdots,$$

(cf. [4]). Defining

$$\tilde{a}^{(N)}(n) = g_n^{(N)}(\tilde{\varepsilon}(0), \cdots, \tilde{\varepsilon}(n-1))$$

we find that the process $\{\tilde{z}(N)\}$ with

$$\tilde{z}(N) = \Sigma_{n=1}^N \tilde{a}^{(N)}(n)\tilde{\varepsilon}(n)$$

is equivalent to $\{z(N)\}$. Relabel the tilded variables and define simple nonanticipating functionals

$$f_N(t) = (n(n+1))^{\frac{1}{2}} a^{(N)}(n) \quad \text{for} \quad t \in \left[1 - \frac{1}{n}, 1 - \frac{1}{n+1}\right),$$
$$n = 1, 2, \cdots, N;$$
$$= 0 \quad \text{for} \quad t \geq 1 - \frac{1}{N+1}.$$

Then

$$\int_0^1 f_N^2(s)\, ds = A(N)^2 \tag{1.3}$$

and

$$\int_0^t f_N(s)\, dw\,(s) = z(N) \quad \text{for} \quad t \geq 1 - \frac{1}{N+1}. \tag{1.4}$$

Fix $\delta > 0$ and define a sequence

$$\alpha_N = \delta/\delta_N^2, \qquad N = 1, 2, \cdots$$

of positive numbers and sets

$$\Omega_N = \left[A(N)^2 \leq \delta_N^2\right].$$

Then

$$P\left(\left[\max_{0 \leq t \leq 1} \int_0^t f_N\, dw \geq 2\delta\right] \cap \Omega_N\right) \leq P\left[\max_{0 \leq t \leq 1} \int_0^t f_N\, dw \geq \delta + \frac{\alpha_N}{2} A(N)^2\right]$$
$$\leq P\left[\max_{0 \leq t \leq 1}\left(\int_0^t f_N\, dw - \frac{\alpha_N}{2}\int_0^t f_N^2\, ds\right) \geq \delta\right] \leq e^{-\delta\alpha_N} = e^{-\delta^2/\delta_N^2}.$$

Repeating with $-a^{(N)}(n)$ in place of $a^{(N)}(n)$ we obtain the estimate

$$P\left(\left[\max_{0 \leq t \leq 1}\left|\int_0^t f_N\, dw\right| \geq 2\delta\right] \cap \Omega_N\right) \leq 2e^{-\delta^2/\delta_N^2}. \tag{1.5}$$

Since

$$P\left[\max_{0\leqslant t\leqslant 1}\left|\int_0^t f_N\,dw\right| \geqslant 2\delta\right] \leqslant P\left(\left[\max_{0\leqslant t\leqslant 1}\left|\int_0^t f_N\,dw\right| \geqslant 2\delta\right] \cap \Omega_N\right) + P\left(\overline{\Omega}_N\right),$$

where $\overline{\Omega}_N$ denotes the complement of Ω_N, and by virtue of (C) the right hand side of the inequality can be estimated by the general terms of convergent sums, it follows from the first Borel-Cantelli lemma that

$$P\left(\limsup_{N\to\infty}\left[\max_{0\leqslant t\leqslant 1}\left|\int_0^t f_N\,dw\right| \geqslant 2\delta\right]\right) = 0.$$

Hence

$$\max_{0\leqslant t\leqslant 1}\left|\int_0^t f_N\,dw\right| \to 0 \text{ a.e. as } N\to\infty,$$

and the assertion follows from (1.4).

For nonstochastic coefficients assumption (C) reduces to the condition

$$\sum_{N=1}^{\infty} e^{-\delta/A(N)^2} < \infty \text{ for all } \delta > 0$$

in Chow [2]. A sufficient condition for (C) (and often easier to check) is

(C′) For every $\delta > 0$ $\sum_{N=1}^{\infty} P[A(N)^2\log(N) > \delta] < \infty$. For nonstochastic $a^{(N)}(n)$ this simply means that $A(N)^2\log(N) \to 0$.

2. Linear regression models. Let us return to the model (0.1) in the introduction. Suppose first that the $x^{(N)}(n)$ are nonstochastic. Following the idea of proof in [1] we write (0.2) in the form

$$\hat{\beta}(N) - \beta = \sum_{n=1}^{N} a^{(N)}(n)\varepsilon(n) \tag{2.1}$$

with

$$a^{(N)}(n) = \left[X(N)'X(N)\right]^{-1}x^{(N)}(n). \tag{2.2}$$

Then

$$\sum_{n=1}^{N} a_j^{(N)}(n)^2 = \left[X(N)'X(N)\right]_{jj}^{-1}$$

for $j = 1, 2, \cdots, K$, and the nonstochastic version of (C′) yields

PROPOSITION 1. *Suppose that in model* (0.1) *the regressors are nonstochastic and assumption* (A) *holds. If* $\operatorname{tr}([X(N)'X(N)]^{-1}) = o(\log(N)^{-1})$, *then the* OLS-*estimator of the parameter vector* β *is strongly consistent.*

As pointed out in the introduction a typical situation in which sampling without repetition arises is the transformation of a model with autocorrelated disturbances into a standard white noise model. Suppose that in the linear regression model (0.1) the disturbances at stage N are of the form

$$u(N) = \left(\varepsilon^{(N)}(1), \cdots, \varepsilon^{(N)}(N)\right)' \tag{2.3}$$

with nonsingular covariance matrix $\Omega(N)$, and consider the transformed model

$$P(N)y(N) = P(N)X(N) + v(N), \tag{2.4}$$

where $P(N)$ comes from

$$P(N)'P(N) = \Omega(N)^{-1} \tag{2.5}$$

and

$$v(N) = P(N)u(N).$$

Assumption (A) written down for the transformed model becomes

(A′) $\Omega(N)$ is nonsingular, and the transformed disturbance vector has the form $v(N) = (\zeta(1), \cdots, \zeta(N))'$, where $\zeta(n)$ $(n = 1, 2, \cdots)$ is a sequence of independent $N(0, 1)$ distributed random variables.

Since the OLS-estimator of β in the transformed model is just the generalized least squares (GLS-) estimator in the original model we find

COROLLARY 1. *Suppose that the regressors are nonstochastic and* (A′) *holds. If* $\operatorname{tr}([X(N)'\Omega(N)^{-1}X(N)]^{-1}) = o(\log(N)^{-1})$, *then the* GLS-*estimator of* β *is strongly consistent.*

Let us now turn to the case where the regressors $x^{(N)}(n)$ are stochastic. The attempt to handle the sampling error in the same way as above, i.e., by considering the sums (2.1) with coefficients $a^{(N)}(n)$ defined by (2.2), will not prove successful due to the measurability condition (B) (e.g., we would have to require that $a^{(N)}(1)$ be constant, which would practically lead us back to the case of constant coefficients since $a^{(N)}(1)$ involves $X(N)$). Instead, we take a sequence $\theta(N)$ $(N = 1, 2, \cdots)$ of real numbers and write the sampling error in the form

$$\hat{\beta}(N) - \beta = \left[\theta(N)X(N)'\Omega(N)^{-1}X(N)\right]^{-1}z(N)$$

with

$$z(N) = \theta(N)X(N)'P(N)'v(N) = \Sigma_{n=1}^{N}a^{(N)}(n)\zeta(n),$$

$$a_j^{(N)}(n) = \theta(N)\left[X(N)'P(N)'\right]_{jn}.$$

Suppose now that P as defined by (2.4) is chosen lower diagonal—which is always possible—and that for this specific P the transformed disturbance vector v satisfies (A′). Then

$$a_j^{(N)}(n) = \theta(N)\Sigma_{i=1}^{n}P(N)_{ni}x_j^{(N)}(i), \qquad j = 1, 2, \cdots, K.$$

Hence, in order to comply with the measurability condition (B), it suffices to require that $x^{(N)}(n)$ be measurable with respect to $\mathfrak{F}\{\zeta(1), \cdots, \zeta(n-1)\}$. Assumption (C′) takes the form

(i) $\Sigma_{N=1}^{\infty}P\{\theta(N)^2[X(N)'\Omega(N)^{-1}X(N)]_{jj}\log(N) > \delta\} < \infty$ for every $\delta > 0$.

By Tschebyscheff's inequality a sufficient condition for (i) to hold is:

(i′) The sequence $\theta(N)^2E\{[X(N)'\Omega(N)^{-1}X(N)]_{jj}^2\}$ $(N = 1, 2, \cdots)$ is bounded and

$$\Sigma_{N=1}^{\infty}\theta(N)^2\log(N)^2 < \infty.$$

The last property is trivially satisfied for sequences of the form $\theta(N) = N^{-\beta}$ with $\beta > 1/2$. With the final assumption

(ii) The sequence $[\theta(N)X(N)'\Omega(N)^{-1}X(N)]^{-1}$ $(N = 1, 2, \cdots)$ is a.e. bounded, we can formulate

PROPOSITION 2. *Suppose that for the linear model* (0.1) *with autocorrelated error terms* (2.3) *a lower diagonal matrix P satisfying* (2.5) *can be found such that* (A′) *holds and $x^{(N)}(n)$ is measurable with respect to $\mathfrak{F}\{\zeta(1), \cdots, \zeta(n-1)\}$. If* (i) *and* (ii) *are fulfilled for some sequence $\theta(N)$ $(N = 1, 2, \cdots)$ of real numbers, then the* GLS-*estimator of β is strongly consistent.*

An important example where this set of assumptions is satisfied is provided by models emerging from autoregressive processes whose error process $\{\varepsilon(n)\}$ is itself generated by an autoregressive scheme with independent $N(0, 1)$ distributed disturbances $\zeta(n)$ (and with the characteristic polynomials of both processes having all roots inside the unit disk). In this case the matrix P transforming $u(N) = (\varepsilon(1), \cdots, \varepsilon(N))'$ into $v(N) = (\zeta(1), \cdots, \zeta(N))'$ is lower diagonal, and the regressors $x^{(N)}(n)$, which are just the lagged values of the observation process, are $\mathfrak{F}\{\zeta(1), \cdots, \zeta(n-1)\}$-measurable provided that the initial values are constant.

REFERENCES

[1] ANDERSON, T. W. and TAYLOR, J. B. Strong consistency of least squares estimates in linear models. *Ann. Statist.* 7 484–489.

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

[3] CHRISTOPEIT, N. and HELMES, K. (1977). Strong consistency of generalized least squares estimators in linear regression models. *Ann. Statist.* To appear.

[4] GIHMAN, I. I. and SKOROHOD, A. V. (1974). *The Theory of Stochastic Processes* I. Springer-Verlag, Berlin, Heidelberg, New York.

[5] MCKEAN, H. P. (1969). *Stochastic Integrals*. Academic Press, New York.

[6] TOMKINS, R. J. (1976). Strong limit theorems for certain arrays of random variables. *Ann. Probability* 4 444–452.

INSTITUT FÜR ÖKONOMETRIE
UND OPERATIONS RESEARCH
UNIVERSITY OF BONN
ADENAUERALLEE 24-42
53 BONN
WEST GERMANY

INSTITUT FÜR ANGEWANDTE
MATHEMATIK
UNIVERSITY OF BONN
WEGELERSTR. 6
53 BONN
WEST GERMANY