The Annals of Probability
1978, Vol. 6, No. 4, 695-698

LIMIT PROCESSES FOR SEQUENCES OF PARTIAL SUMS OF REGRESSION RESIDUALS[1]

BY IAN B. MACNEILL

University of Western Ontario

Linear regression of a random variable against several functions of time is considered. Limit processes are obtained for the sequences of partial sums of residuals. The limit processes, which are functions of Brownian motion, have covariance kernels of the form:

$$K(s, t) = \min(s, t) - \int_0^t \int_0^s g(x, y)\, dx\, dy\ .$$

The limit process and its covariance kernel are explicitly stated for each of polynomial and harmonic regression.

1. Introduction and summary. The first test for change of regression at unknown time, obtained by Quandt (1960), was based on the likelihood ratio. Distribution theory for this test statistic has been shown by Feder (1975) to be complicated and to depend upon the configuration of the observations on the independent variables. An approach that does not have this dependence and which yields statistics whose properties are well known has been proposed by Brown, Durbin and Evans (1975). This approach requires the computation of recursive residuals, a computation rarely included in a standard regression analysis. MacNeill (1977) proposes a test for change of polynomial regression based on the sequence of partial sums of raw regression residuals. Derivation of large sample distribution theory for this test depends upon first ascertaining the limit process for the sequence of partial sums of regression residuals. In this paper we consider linear regression of a random variable against several regressor functions of time. The class of regressor functions considered is wide enough to include most functions used in practice. Limit processes are found for the sequences of partial sums of regression residuals. These limit processes, which are functions of Brownian motion, have covariance kernels that are simply defined in terms of the regressor functions. Examples given include polynomial and harmonic regression.

We first define the basic model. Let $\{\varepsilon_j, j \geqq 1\}$ be a sequence of independent and identically distributed random variables possessing zero means and variances $\sigma^2 < \infty$. Also let $\{f_k(\cdot), 0 \leqq k \leqq p\}$ be a collection of regressor functions defined on $[0, 1]$ and define the triangular array $\{Y_{nj}, 1 \leqq j \leqq n, n \geqq 1\}$ of dependent variables as follows:

$$Y_{nj} = \sum_{i=0}^{p} \beta_i f_i(j/n) + \varepsilon_j\ .$$

Received January 10, 1977.

[1] This research was supported by a grant from the National Research Council of Canada.

AMS 1970 *subject classifications.* Primary 60F05; Secondary 62J05.

Key words and phrases. Brownian motion, harmonic regression, polynomial regression, regression residuals, weak convergence.

For convenience the total time of observation has been compressed to the interval [0, 1] and observations are taken at equi-spaced time points. In the usual matrix formulation we have

$$\mathbf{Y}_n = \mathbf{X}_n \boldsymbol{\beta}_p + \boldsymbol{\varepsilon}_n$$

where the (s, r)th component of the design matrix is $f_r(s/n)$. The Gauss–Markov estimator for $\boldsymbol{\beta}_p$ is denoted by $\hat{\boldsymbol{\beta}}_{pn}$ and is defined by

$$\hat{\boldsymbol{\beta}}_{pn} = (\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{X}_n'\mathbf{Y}_n ,$$

subject to the existence of the inverse.

2. Limit processes. Sequences of partial sums of regression residuals are denoted by $\{(Sf_{nj}, 1 \leqq j \leqq n), n \geqq 1\}$ where $Sf_{nj} = \sum_{i=1}^{j} (Y_{ni} - \hat{Y}_{ni})$, $\hat{Y}_{ni} = \hat{\boldsymbol{\beta}}'_{pn}\mathbf{f}(i/n)$ and $\mathbf{f}'(i/n) = \{f_0(i/n), f_1(i/n), \cdots, f_p(i/n)\}$. $Sf_{n0} \equiv 0$. These sequences of partial sums define a sequence of stochastic processes $[\{\theta_{fn}(t), t \in [0, 1]\}, n \geqq 1]$ possessing continuous sample paths as follows:

$$\sigma n^{\frac{1}{2}}\theta_{fn}(t) = Sf_{n[nt]} + (nt - [nt])(Y_{[nt]+1} - \hat{Y}_{[nt]+1}) .$$

Then if $\mathbf{e}_{nj}$ is an $n \times 1$ vector whose first j components are 1 and the remainder zero one can write:

$$\sigma n^{\frac{1}{2}}\theta_{fn}(j/n) = Sf_{nj} = \mathbf{e}'_{nj}\{I_n - \mathbf{X}_n(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{X}_n'\}\boldsymbol{\varepsilon}_n .$$

Provided the Riemann integrals on [0, 1] of $f_r^2(\cdot)$ $(r = 0, \cdots, p)$ exist one finds the (r, s)th component of $\lim n^{-1}(\mathbf{X}_n'\mathbf{X}_n) \equiv \mathbf{F}$ to be $\int_0^1 f_r(t)f_s(t)\,dt$. We then let $g(s, t) \equiv \mathbf{f}'(s)\mathbf{F}^{-1}\mathbf{f}(t)$, assuming, here and in the sequel, that the inverse in this bilinear form exists.

Now denote by $\{B(t), t \in [0, 1]\}$ the standard Brownian motion process with continuous sample paths. Such a process is Gaussian with $E\{B(t)\} = B(0) = 0$ and $E\{B(t)B(s)\} = \min(s, t)$. Also let $S_j = \sum_{i=1}^{j} \varepsilon_i$ and define another sequence of stochastic processes $[\{\theta_n(t), t \in [0, 1]\}, n \geqq 1]$ possessing continuous sample paths by

$$n^{\frac{1}{2}}\sigma\theta_n(t) = S_{[nt]} + (nt - [nt])\varepsilon_{[nt]+1} .$$

We then have the following result.

THEOREM. *If $f_r(t)$ $(r = 0, \cdots, p)$ are continuously differentiable on* [0, 1] *then $\{\theta_{fn}(t), t \in [0, 1]\}$ converges weakly to the Gaussian process $\{B_f(t), t \in [0, 1]\}$ defined by*:

$$B_f(t) = B(t) - B(1) \int_0^t g(x, 1)\,dx + \int_0^1 B(x) \left\{\int_0^t \frac{\partial}{\partial x} g(x, y)\,dy\right\} dx .$$

This process has mean and $B_f(0)$ equal to zero and covariance kernel, $K_f(s, t)$, given by:

$$K_f(s, t) = E\{B_f(s)B_f(t)\} = \min(s, t) - \int_0^s \int_0^t g(x, y)\,dx\,dy .$$

PROOF. Define a sequence of functions, $\{h_{fn}(\cdot), n \geqq 1\}$, from the space of continuous functions, $C[0, 1]$, into itself by the following relation:

$$h_{fn}\{\theta_n(t)\} = \theta_n(t) - \mathbf{e}'_{n,nt}\mathbf{X}_n(\mathbf{X}_n'\mathbf{X}_n)^{-1}(\theta_n(1)\mathbf{f}(1) - \sum_{j=1}^{n}[\mathbf{f}(j/n) - \mathbf{f}\{(j-1)/n\}]\theta_n\{(j-1)/n\}) .$$

$\mathbf{e}_{n,nt}$ is defined to be an $n \times 1$ vector whose first $[nt]$ components are one, the next is $nt - [nt]$ and the remainder are zero. Abel's partial summation formula implies that $h_{fn}\{\theta_n(t)\} = \theta_{fn}(t)$. Also define $h_f(\cdot)$, a function from $C[0, 1]$ into itself, by: $h_f\{B(t)\} = B_f(t)$. $\{h_{fn}(\cdot), n \geqq 1\}$ and $h_f(\cdot)$ are continuous in the uniform topology on $C[0, 1]$. Since $\lim n^{-1}\mathbf{e}'_{n[nt]}\mathbf{X}_n = \int_0^t \mathbf{f}'(s)\,ds$ and $\lim n(\mathbf{X}_n'\mathbf{X}_n)^{-1} = \mathbf{F}^{-1}$ it follows that $h_{fn}(\cdot)$ converges to $h_f(\cdot)$ in the sense that if $\{Z_n, n \geqq 1\}$ and Z are elements in $C[0, 1]$ and Z_n converges uniformly to Z then $h_{fn}(Z_n)$ converges uniformly to $h_f(Z)$. If $\{P_{fn}, n \geqq 1\}$ and P_f are the measures generated in $C[0, 1]$ by $[h_{fn}\{\theta_n(\cdot)\}, n \geqq 1]$ and $B_f(\cdot)$, and $\{W_n, n \geqq 1\}$ and W are the measure generated by $\{\theta_n(\cdot), n \geqq 1\}$ and $B(\cdot)$ we have $P_{fn} = W_n h_{fn}^{-1}$ and $P_f = W h_f^{-1}$. Theorem 5.5 Billingsley (1968) implies that P_{fn} converges weakly to P_f which completes the proof.

Embodied in this proof is the result that $\sum_{j=1}^{n} f_r(j/n)\varepsilon_j/(n^{\frac{1}{2}}\sigma)$ converges in distribution to $f_r(1)B(1) - \int_0^1 B(t)(d/dt)f_r(t)\,dt$, which is integration by parts. It can also be observed that

$$B_f(t) = B(t) - \int_0^t \{\int_0^1 g(x, y)\,dB(y)\}\,dx ,$$

a formula that facilitates computation of the covariance kernel.

The time of observation need not be restricted to $[0, 1]$ nor must the sampling be equi-spaced. Suppose the total sampling period is $[0, T]$ and the rate of sampling is described by a nonconstant positive function $\{r(t), t \in [0, T]\}$. Then, with $R(t) = \int_0^t r(s)\,ds / \int_0^T r(s)\,ds$, the limit process $B_{fRT}(\cdot)$ may be related to that of the theorem by the relation:

$$B_{fRT}(t) = B_f\{R(t)\} .$$

3. Examples. First consider the case of fitting a mean to a set of data, i.e., $p = 0$ and $f_0(t) \equiv 1$. Then $B_f(t) = B(t) - B(1)t$, which is the Brownian bridge with $K_f(s, t) = \min(s, t) - st$.

Next consider the case of fitting a polynomial of degree p to a set of data. Then $f_r(t) = t^r$ $(r = 0, 1, \cdots, p)$. Exactly the same fit is provided by using the orthogonal polynomials up to order p. See Allan (1930) for the general form of these polynomials. Orthogonal polynomials can be used to show that

$$B_f(t) = B(t) - \sum_{m=0}^{p}(2m+1)\left[B(1)g_m(1)\int_0^t g_m(s)\,ds - \{\int_0^t g_m(s)\,ds\}\left\{\int_0^1 B(s)\frac{d}{ds}g_m(s)\,ds\right\}\right]$$

where

$$g_m(t) = \sum_{q=0}^{[m/2]}(-1)^q \frac{\binom{2m}{m,q,q,m-2q}}{2^{4q}\binom{m-\frac{1}{2}}{q}}(t - \tfrac{1}{2})^{m-2q} .$$

See MacNeill (1977) for a derivation of this result. The covariance kernel is given by:

$$K_f(s, t) = \min(s, t) - \sum_{m=0}^{p} (2m + 1)\{\int_0^s g_m(x)\, dx\}\{\int_0^t g_m(y)\, dy\} .$$

Although it is not true in general it can be seen in this case that 0 and 1 are zeros of $K_f(t, t)$ so $B_f(t)$ can be thought of as a generalized Brownian bridge in the sense that its paths are tied to 0 at $t = 0$ and $t = 1$. The bridge property will hold if one of the $f_j(\cdot)$ is constant.

Finally, consider the case of fitting an harmonic polynomial of degree p to a set of data. Let $f_k(t) = \cos 2\pi kt$ $(k = 0, 1, \cdots, p)$ and $f_{p+k}(t) = \sin 2\pi kt$ $(k = 1, 2, \cdots, p)$. The theorem implies that

$$\begin{aligned} B_f(t) = B(t) - B(1)\{t + \textstyle\sum_{k=1}^{p} (\pi k)^{-1} \sin 2\pi kt\} \\ + 2 \int_0^1 B(s)[\textstyle\sum_{k=1}^{p} \{\cos 2\pi ks - \cos 2\pi k(t - s)\}]\, ds \end{aligned}$$

and that

$$\begin{aligned} K_f(s, t) = \min(s, t) - st - \textstyle\sum_{j=1}^{p} (2\pi^2 j^2)^{-1}\{(1 - \cos 2\pi jt)(1 - \cos 2\pi js) \\ + \sin 2\pi js \sin 2\pi jt\} . \end{aligned}$$

Again it can be seen that the paths are tied at both $t = 0$ and $t = 1$.

Acknowledgment. The author would like to thank the referee for helpful comments and suggestions.

REFERENCES

[1] ALLAN, F. E. (1930). The general form of the orthogonal polynomials for simple series, with proofs of their simple properties. *Proc. Roy. Soc. Edinburgh Sect. A* **50** 310–320.

[2] BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York,

[3] BROWN, R. L., DURBIN, J. and EVANS, J. M. (1975). Techniques for testing the constancy of regression relationships over time. *J. Roy. Statist. Soc. Ser. B* **37**, 149–192.

[4] FEDER, P. I. (1975). The log likelihood ratio in segmented regression. *Ann. Statist.* **3** 84–97.

[5] MACNEILL, I. B. (1978). Properties of sequences of partial sums of polynomial regression residuals with applications to tests for change in regression at unknown times. *Ann. Statist.* **6** 422–433.

[6] QUANDT, R. E. (1960). Tests of the hypothesis that a linear regression system obeys two separate regimes. *J. Amer. Statist. Assoc.* **55** 324–330.

DEPARTMENT OF APPLIED MATHEMATICS
THE UNIVERSITY OF WESTERN ONTARIO
LONDON, CANADA