The Annals of Statistics
1998, Vol. 26, No. 4, 1398–1410

A FUNCTIONAL CENTRAL LIMIT THEOREM FOR REGRESSION MODELS

BY WOLFGANG BISCHOFF

University of Karlsruhe

Dedicated to the memory of Werner Fieger
(*May 31, 1935–December 6, 1997*)

Let a linear regression be given. For detecting change-points, it is usual to consider the sequence of partial sums of least squares residuals whence a partial sums process is defined. Given a sequence of exact experimental designs, we consider for each design the corresponding partial sums process. If the sequence of designs converges to a continuous design, we derive the explicit form of the limit process of the corresponding sequence of partial sums processes. This is a complicated function of the Brownian motion. These results are useful for the problem of testing for change of regression at known or unknown times.

1. Introduction. In the literature on "detecting change-points" in linear regression models, it is usual to consider the sequence of the partial sums of the least squares residuals or variants of it; see, for instance, Gardner (1969), Brown, Durbin and Evans (1975), Sen and Srivastava (1975), MacNeill (1978a), Jandhyala and MacNeill (1991), Jandhyala (1993), Watson (1995) and the references cited there. For solving change-point problems, the so-called residual partial sums processes are useful; these are the limit processes of sequences of stochastic processes defined by partial sums of regression residuals [see MacNeill (1978a), Jandhyala and MacNeill (1989, 1991, 1992) and Tang and MacNeill (1993)]. Under certain conditions, MacNeill (1978b) derived the explicit form of the residual partial sums processes for general linear regression residuals. These processes are complicated functions of the standard Brownian motion.

In this paper we consider a generalized approach by assuming that the observations are taken according to an arbitrary "design" in contrast to their being sampled equidistantly. We derive the corresponding residual partial sums processes for general linear regression residuals. By our approach we can generalize the result of MacNeill (1978b) under weaker assumptions.

There are several points of view from which we are interested in arbitrary "designs." First, for economic, technical or ecological reasons or by prior information, it is possible that the statistician cannot or will not sample equidistantly. Next, our approach enables us to derive Bayes-type statistics by using a design according to the prior information on the change-point. This is the nat-

Received May 1996; revised January 1998.
AMS 1991 *subject classifications*. Primary 60F17, 60G15, 62J05.
Key words and phrases. Linear regression, regression residuals, partial sums process, functional central limit theorem, functions of Brownian motion, change-point problems, tests.

ural way in contrast to the Bayes-type method [following Chernoff and Zacks (1964)] for regression models that weights the sums of the residuals according to a prior; see Jandhyala [(1993), page 324]. Further, let us consider tests for detecting change-points which are based on the sequence of the partial sums of the least squares residuals. Usually, the sample is taken equidistantly. But is this the best design for detecting change-points? This problem can be treated asymptotically with the help of our generalized approach to residual partial sums processes. Note that in the literature this problem of designing has not been handled yet (as far as the author knows). A forthcoming paper, Bischoff and Miller (1998), determines asymptotically optimal tests and asymptotically optimal designs for the problem of testing whether a change-point does or does not occur. There it can be seen that for many problems it is not optimal to sample equidistantly.

To explain the problem in more detail, let us consider a sequence of real random variables $\varepsilon_1, \varepsilon_2, \varepsilon_3, \dots$ being independent and identically distributed with mean 0 and variance $\sigma^2 > 0$. We assign $\boldsymbol{\varepsilon}_n = (\varepsilon_1, \dots, \varepsilon_n)^\top$ to the stochastic process $(\sigma\sqrt{n})^{-1} T_n(\boldsymbol{\varepsilon}_n)$ in $C[0,1]$ where

$$T_n(a)(z) = \sum_{i=1}^{[nz]} a_i + \big(nz - [nz]\big) a_{[nz]+1}, \qquad z \in [0,1],$$

with $a = (a_1, \dots, a_n)^\top \in \mathbb{R}^n$, $[s] = \max\{n \in \mathbb{N}_0 \mid n \le s\}$ and $\sum_{i=1}^{0} a_i = 0$. It is well-known by Donsker's theorem [see Billingsley (1968), Theorem 10.1] that $(\sigma\sqrt{n})^{-1} T_n(\boldsymbol{\varepsilon}_n)$ converges weakly to the Brownian motion in $C[0,1]$; note, as usual, we consider the uniform topology on $C[0,1]$.

MacNeill (1978b) proved a similar result for the least squares residuals of a regression model. Throughout the paper we shall consider the following regression model: let $f_1, \dots, f_m \colon \mathscr{E} \to \mathbb{R}$ be known measurable regression functions where $\mathscr{E} \subseteq \mathbb{R}$ is the experimental region. As usual we write $f(t)$ for $(f_1(t), \dots, f_m(t))^\top$, $t \in \mathscr{E}$. Let us consider a triangular array t_{nj}, $1 \le j \le n$, $n \in \mathbb{N}$, of arbitrary experimental conditions, that is $t_{nj} \in \mathscr{E}$. We assume for each $n \in \mathbb{N}$,

$$t_{n1} \le t_{n2} \le \cdots \le t_{nn}.$$

Each $(t_{n1}, \dots, t_{nn})$ is called an (exact) design (for n observations). Note that we do not assume that $t_{ni} \ne t_{ni+1}$. Corresponding to this array of experimental conditions, we have a triangular array of random variables Y_{nj}, $1 \le j \le n$, $n \in \mathbb{N}$, defined by

$$Y_{nj} = \sum_{i=1}^{m} \beta_i f_i(t_{nj}) + \varepsilon_{nj}$$

where $\boldsymbol{\beta} = (\beta_1, \dots, \beta_m)^\top \in \mathbb{R}^m$ is the unknown parameter vector of interest and $\boldsymbol{\varepsilon}_n = (\varepsilon_{n1}, \dots, \varepsilon_{nn})^\top$ is a vector of stochastic independent and identically distributed real-valued random variables with $E(\varepsilon_{nj}) = 0$ and $\operatorname{var}(\varepsilon_{nj}) = \sigma^2 > 0$.

Let $n \in \mathbb{N}$ be fixed. In the usual matrix formulation we have

$$\mathbf{Y}_n = X_n \boldsymbol{\beta} + \boldsymbol{\varepsilon}_n, \tag{1.1}$$

where X_n is the model matrix corresponding to the design $(t_{n1}, \ldots, t_{nn})$, that is the (s, r)th component of X_n is $f_r(t_{ns})$. Then for $\boldsymbol{\beta}$ being estimable, $\operatorname{rank}(X_n)$ must be equal to m. We assume that $\operatorname{rank}(X_n) = m$ for all $n \geq n_0$ and in the sequel we consider $n \geq n_0$ only. Given model (1.1) the best linear unbiased estimation for $X_n \boldsymbol{\beta}$ is given by the least squares estimation $pr_{X_n} \mathbf{Y}_n = X_n (X_n^\top X_n)^{-1} X_n^\top \mathbf{Y}_n$ and the corresponding least squares residual vector is given by

$$\mathbf{r}_n = (r_{n1}, \ldots, r_{nn})^\top = pr_{X_n^\perp} \mathbf{Y}_n = pr_{X_n^\perp} \boldsymbol{\varepsilon}_n,$$

where $pr_{X_n} = X_n (X_n^\top X_n)^{-1} X_n^\top$ and $pr_{X_n^\perp} = I_n - X_n (X_n^\top X_n)^{-1} X_n^\top$ are the orthogonal projectors onto $\operatorname{range}(X_n)$ and onto the orthogonal complement of $\operatorname{range}(X_n)$, respectively. Then let us consider the stochastic process $(\sigma\sqrt{n})^{-1} T_n(\mathbf{r}_n)(z)$, $z \in [0, 1]$, in $C[0, 1]$ corresponding to $\mathbf{r}_n$.

Under the conditions that

$$\mathscr{E} = [0, 1] \quad \text{and} \quad t_{ni} = \frac{i}{n}, \quad 1 \leq i \leq n, \ n \in \mathbb{N}, \tag{1.2}$$

$$f_1, \ldots, f_m \text{ are continuously differentiable,}$$

$$J^{-1} \text{ exists where } J = \lim_{n \to \infty} \frac{1}{n} (X_n^\top X_n),$$

MacNeill (1978b) showed that $(\sigma\sqrt{n})^{-1} T_n(\mathbf{r}_n)$ converges weakly in $C[0, 1]$ to the Gaussian process B_f defined by

$$B_f(z) = B(z) + \left(\int_0^z f(x)\, dx \right)^\top J^{-1} \left(\int_0^1 B(x) \frac{d}{dx} f(x)\, dx - B(1) f(1) \right), \qquad z \in [0, 1],$$

where $B(t)$ denotes the standard Brownian motion in $C[0, 1]$; see also Remark 3.3. Functionals of $B_f(z)$ can be used as test statistics for change of regression at known or unknown times; see MacNeill (1978a), Jandhyala (1993) and the references cited there. The expression B_f and the generalization of B_f given in this paper are called residual partial sums processes. Further, it is worth mentioning that certain properties of the sample path behavior of residual partial sums processes are shown in Bischoff (1996). In Bischoff and Miller (1998) the problem of testing the null hypothesis

H_0: a change-point occurs

against the alternative:

K: a change-point does not occur

is considered. Based on B_f (and its generalization, respectively) the most powerful test statistic is determined there. Further, the problem of experimental design is investigated for the test problem given above.

There are some practical reasons why we are interested in generalizing the result of MacNeill to a sequence of designs which does not fulfill (1.2):

1. (Bayesian view.) In the literature on change-point problems, Bayes-type methods were introduced by Chernoff and Zacks (1964). Following them, it was suggested for change-point problems in regression models to weight the partial sums of the residuals according to a discrete prior on the unknown change-point; see Jandhyala (1993) and the references cited there. But if some prior information is available on the change-point, that is, if there exists a probability measure P_0 on the experimental region according to which the change-point is distributed, then it seems to be more convenient to sample according to the probability measure P_0. Note that we do not consider Bayes-type statistics further in this paper. For more information we refer the reader to Jandhyala (1993) and Bischoff (1996).

   Here we should mention how we can sample according to a probability measure P_0. That is, how an exact design $(t_{n1}, \dots, t_{nn})$ is chosen according to P_0: choose $t_{ni} = Q_0(i/n - z_0)$ with $z_0 \in [0, 1/n]$ arbitrarily fixed where Q_0 is the quantile function of P_0. Note that throughout the paper our distribution functions and hence our quantile functions are continuous from the right. Further, a probability measure on the experimental region is called continuous design below.
2. (Technical or economic view.) There may be technical or economic reasons preventing sampling equidistantly.
3. (Design of experiments.) Design of experiments can be considered for the problem of testing for change of regression at known or unknown times. For solutions for such problems with the help of the functional central limit theorem given in Section 2, we refer to the forthcoming paper of Bischoff and Miller (1998).

Moreover, there are some theoretical interests in considering a more general approach. Obviously, the residual partial sums process B_f depends on the vector of regression functions $f = (f_1, \dots, f_m)^\top$. However, one wonders whether B_f depends on the sequence of designs $(t_{n1}, \dots, t_{nn})_{n\in\mathbb{N}}$ or only on the "limit" of the designs $(t_{n1}, \dots, t_{nn})$ for $n \to \infty$. To make the last sentence precise we have to explain the meaning of "limit." To this end, note that each design $(t_{n1}, \dots, t_{nn}) \in \mathscr{E}^n$ uniquely corresponds to a discrete probability measure P_n on $\mathscr{E}$ by

$$P_n = \frac{1}{n} \sum_{i=1}^{n} P_{\{t_{ni}\}}, \tag{1.3}$$

where $P_{\{t\}}$ denotes the one point measure in t. In the sequel we identify a design with its representation as a discrete probability measure and we call each probability measure on $\mathscr{E}$ continuous design. Coming back to the result of MacNeill, we see that the sequence of designs $(1/n, 2/n, \dots, 1)$ converges to the Lebesgue-measure on $[0, 1]$. (We will explain later in which sense the convergence is meant.) Now several questions arise; for instance, the following:

1. Does the limit process B_f in the result of MacNeill depend on the sequence of designs or only on the limit of the sequence of designs, that is, only on the Lebesgue-measure on [0, 1]? (This means: does an “invariance principle” hold true?)
2. Can we define a residual partial sums process for each convergent sequence of designs?
3. Can we define a residual partial sums process for each continuous design?

It is worth mentioning that we need not assume for our result that $f_1, \dots, f_m$ are differentiable. It is not even necessary that $f_1, \dots, f_m$ are continuous if in return certain conditions are assumed on P_0. This is important for regression models such as Haar regression [see Herzberg and Traves (1994)], where the regression functions $f_1, \dots, f_m$ are not continuous. However, in the present paper we restrict ourselves to the case that $f_1, \dots, f_m$ are continuous and P_0 is arbitrary.

The paper is organized as follows. In Section 2 we introduce the residual partial sums process for a sequence of designs converging to a continuous design. Moreover, it is discussed there what kind of convergence for the sequence of designs is needed for the functional central limit theorem which is given in Section 2 as well. Note that the sequence of partial sums processes in Section 2 is defined on [0, 1] where $z \in [0, 1]$ corresponds with the observations belonging to the experimental conditions $t_{n1}, \dots, t_{n[nz]}$. It is more convenient, however, to consider these processes on the experimental region. Therefore we transform these processes to the experimental region in Section 3. For further reasons for this transformation, see Section 3. Finally, in the Appendix some technical proofs are given.

2. Residual partial sums processes.

2.1. *Preliminary remarks.* Let us consider the regression model defined in Section 1 with experimental region $\mathscr{E} = [a, b] \subseteq \mathbb{R}$; let $(t_{n1}, \dots, t_{nn})$ be a design and let P_n be its representation as a probability measure according to (1.3). In the sequel we do not distinguish between a design, its representation as a probability measure according to (1.3) and the distribution function F_n, say, corresponding to P_n. We define $F_{n;z}(t) := \min\{F_n(t), [nz]/n\}$, $z \in [0, 1]$. Then we have, with the notation $f(t) = (f_1(t), \dots, f_m(t))^\top$,

$$\frac{1}{n} X_n^\top X_n = \int_{\mathscr{E}} f(t) f(t)^\top P_n(dt)$$

and

$$\frac{1}{n} \mathbf{1}_{n;z}^\top X_n = \frac{1}{n} \sum_{j=1}^{[nz]} f(t_{n_j})^\top = \int_{\mathscr{E}} f(t)^\top F_{n;z}(dt), \qquad z \in [0, 1],$$

where $\mathbf{1}_{n;z} \in \mathbb{R}^n$ is the vector whose first $[nz]$ components are one and the remainder is zero. For our results we need that the sequence of designs F_n

converges uniformly to a continuous design F_0, say,

$$\sup_{t\in[a,b]} \left|F_n(t) - F_0(t)\right| \to 0 \quad \text{for } n \to \infty. \tag{2.1}$$

To avoid misunderstanding we repeat that each probability measure P_0 on the experimental region as well as the corresponding distribution function F_0 is called continuous design, but this does not imply that F_0 is continuous.

Note that condition (2.1) is equivalent to the fact that $F_{n;z}$ converges uniformly to $F_{0;z}$ for each $z \in [0,1]$ where $F_{0;z}(t) = \min\{F_0(t), z\}$. Hence condition (2.1) implies

> F_n converges weakly to F_0 and $F_{n;z}$ converges weakly to $F_{0;z}$ for all $z \in [0,1]$.

Under the additional assumption

$$f_1, \dots, f_m \text{ continuous},$$

we have then, by using the definition of weak convergence pointwise,

$$\lim_{n\to\infty} \frac{1}{n} X_n^\top X_n = \lim_{n\to\infty} \int_{\mathscr{E}} f(t) f(t)^\top P_n(dt) = \int_{\mathscr{E}} f(t) f(t)^\top P_0(dt) =: J,$$

$$\lim_{n\to\infty} \frac{1}{n} \mathbf{1}_{n;z}^\top X_n = \lim_{n\to\infty} \int_{\mathscr{E}} f(t)^\top F_{n;z}(dt) = \int_{\mathscr{E}} f(t)^\top F_{0;z}(dt), \qquad z \in [0,1].$$

Let us assume

$$\operatorname{rank}(J) = m. \tag{2.2}$$

It is obvious that (2.2) is fulfilled if and only if the regression functions $f_1, \dots, f_m$ are linearly independent in $L_2(P_0)$ where $L_2(P_0)$ is the Hilbert space of quadratic integrable functions with respect to P_0.

REMARK 2.1. Given a continuous design P_0, in a natural way we can then construct a sequence of designs (P_n) with $P_n = (1/n)\sum_{i=1}^n P_{\{t_{ni}\}}$ converging uniformly [according to (2.1)] to P_0: for example, we can choose $t_{ni} := Q_0(i/n)$, $1 \le i \le n$, where Q_0 is the quantile function of P_0.

2.2. *Functional central limit theorem.* In Section 2.1 all integrals are Lebesgue integrals. But in the following we also consider Riemann–Stieltjes integrals. Integrals obtained by the Riemann–Stieltjes approach are denoted by $\int^{(R)}$. To understand this in more detail, have a look at the formula for $B_{f,P_0}(z)$ given in Theorem 2.2. Then $\int_{\mathscr{E}} f(t) F_{0;z}(dt)$ and J are obtained as limits via weak convergence as described in Section 2.1 above, but $\int_{\mathscr{E}}^{(R)} B(F_0(t)) f(dt)$ is obtained as the limit of Riemann–Stieltjes sums. Obviously, the last integral coincides with the corresponding measure integral. However, for getting more insight into the proof of Theorem 2.2 and where and why the assumptions are needed, it seems to be more convenient to distinguish between these integral notions.

Now we are able to prove a generalization of the theorem given in MacNeill (1978b).

THEOREM 2.2. *Let the regression functions* $f_1, \dots, f_m$ *be continuous and of bounded variation. Let the conditions* (2.1) *and* (2.2) *be fulfilled for* $f_1, \dots, f_m$, *the sequence of designs* (P_n) *and the continuous design* P_0. *Then* $(\sigma\sqrt{n})^{-1}T_n(\mathbf{r}_n)$ *converges weakly in* $C[0,1]$ *to the Gaussian process* B_{f,P_0}, *defined by*

$$B_{f,P_0}(z) = B(z) + \left(\int_{\mathscr{E}} f(t)F_{0;z}(dt)\right)^{\top} \times J^{-1}\left(\int_{\mathscr{E}}^{(R)} B\big(F_0(t)\big)f(dt) - B(1)f\big(Q_0(1)\big)\right),$$

where

$$\int_{\mathscr{E}}^{(R)} B\big(F_0(t)\big)f(dt) = \left(\int_{\mathscr{E}}^{(R)} B\big(F_0(t)\big)f_1(dt), \dots, \int_{\mathscr{E}}^{(R)} B\big(F_0(t)\big)f_m(dt)\right)^{\top}.$$

PROOF. Let us consider the following mapping for each $n \in \mathbb{N}$:

$$V_n\colon C[0,1] \to \mathbb{R}^n,$$

$$u \mapsto \left(u\left(\frac{1}{n}\right) - u(0), u\left(\frac{2}{n}\right) - u\left(\frac{1}{n}\right), \dots, u(1) - u\left(\frac{n-1}{n}\right)\right)^{\top}.$$

The function

$$\phi_n\colon C[0,1] \to C[0,1], \qquad u \mapsto T_n\big(pr_{X_n^{\perp}} V_n(u)\big)$$

is continuous, linear and idempotent; that is, ϕ_n is a continuous projector. By Lemma A.1 in the Appendix, we have for $u \in C[0,1]$,

$$\begin{aligned} u^* &:= \lim_{n\to\infty} X_n^{\top} V_n(u) \\ &= f\big(Q_0(1)\big)u(1) - f\big(Q_0(0)\big)u(0) - \int_{\mathscr{E}}^{(R)} u\big(F_0(t)\big)f(dt). \end{aligned}$$

Thus by Section 2.1 we get for $u \in C[0,1]$,

$$\begin{aligned} \phi(u)(z) &:= \lim_{n\to\infty} \phi_n(u)(z) \\ &= \lim_{n\to\infty} \left[T_n\big(V_n(u)\big)(z) - T_n\big(X_n(X_n^{\top}X_n)^{-1}X_n^{\top}V_n(u)\big)(z)\right] \\ &= u(z) - u(0) - \lim_{n\to\infty}\left[\frac{1}{n}\mathbf{1}_{n;z}^{\top} X_n \left(\frac{1}{n}X_n^{\top}X_n\right)^{-1} X_n^{\top} V_n(u)\right] \\ &= u(z) - u(0) - \int_{\mathscr{E}} f(t)^{\top} F_{0;z}(dt) J^{-1} u^*, \qquad z \in [0,1]. \end{aligned}$$

Note that ϕ is linear and continuous. Further, the above convergence holds true with respect to the uniform topology on $[0,1]$ for each $u \in C[0,1]$ and

$\limsup_{n\to\infty} \|\phi_n\| < \infty$. However, we do not have $\phi_n \to \phi$ with respect to the operator norm, but we obtain a weaker property of the sequence (ϕ_n). If u_n converges to u in $C[0,1]$, then we get

$$\begin{aligned} \|\phi(u) - \phi_n(u_n)\| &\leq \|(\phi - \phi_n)(u)\| + \|\phi_n\|\,\|u - u_n\| \\ &\to 0 \quad \text{for } n \to \infty. \end{aligned} \tag{2.3}$$

Note that by construction we have

$$\phi_n\big(T_n(\boldsymbol{\varepsilon}_n)\big) = T_n(\mathbf{r}_n).$$

Because (2.3) holds true and $(1/\sigma\sqrt{n})T_n(\boldsymbol{\varepsilon}_n)$ converges weakly to the Brownian motion B, the stochastic process $(1/\sigma\sqrt{n})T_n(\mathbf{r}_n)$ converges weakly to the limit process $\phi(B)$ of $\phi_n((1/\sigma\sqrt{n})T_n(\boldsymbol{\varepsilon}_n))$ by Theorem 5.5 of Billingsley (1968). Hence the assertion follows by noting that $B(0) = 0$ almost surely. □

REMARK 2.3. Theorem 2.2 and Remark 2.1 imply that for each continuous design a residual partial sums process exists under the following assumptions:

> The regression functions $f_1, \dots, f_m$ are continuous and of bounded variation, condition (2.2) is fulfilled.

This process does not depend on the sequence of designs converging uniformly to the "limit" design P_0 but only on P_0. That means: we have an "invariance principle."

REMARK 2.4. If σ^2 is unknown, then without altering the asymptotic distribution theory, σ^2 may be replaced with a consistent estimator, such as the usual variance estimator based on the sum of squares of residuals.

REMARK 2.5. Analogous formulas as given in Corollary 3.2 hold true for B_{f,P_0}.

3. Transformation to the experimental region. In Theorem 2.2 the stochastic processes have paths in $C[0,1]$ but it seems to be more convenient to consider corresponding processes with paths in $D[a,b]$ because $[a,b]$ is the experimental region. (Note that as usual, $D[a,b]$ is the set of all functions f: $[a,b] \to \mathbb{R}$ being continuous from the right on $[a,b)$ and with left-hand limits on $(a,b]$.) For example, let us take exact $k > 1$ experiments with the experimental condition $t = a$; that is, $t_{n1} = \cdots = t_{nk} = a < t_{nk+1}$, and let $r_{n1}, \dots, r_{nk}$ be the corresponding residuals. Then the residual partial sums process $(\sigma\sqrt{n})^{-1}T_n(\mathbf{r}_n)(z)$ for $z = i/n$ is given by $(\sigma\sqrt{n})^{-1}\sum_{j=1}^{i} r_{nj}$, $i = 1, \dots, k$. The values of the partial sums process $(\sigma\sqrt{n})^{-1}T_n(\mathbf{r}_n)(z)$ for $z \in (0, k/n)$ depend on the numbering of the residuals $r_{n1}, \dots, r_{nk}$, but the value for $z = k/n$ is independent of the order and gives the information corresponding to the observations taken in $t = a$. Such problems do not occur for the transformed process; note that the value of the transformed process for $s = a$ is $(\sigma\sqrt{n})^{-1}T_n(\mathbf{r}_n)(k/n)$.

Note that we have to use $D[a, b]$ instead of $C[a, b]$ since F_n, $n \in \mathbb{N}$ and F_0 are elements of $D[a, b]$ in general. We give $D[a, b]$ the uniform topology—the topology given by the uniform metric $\rho(x, y) = \sup_t |x(t) - y(t)|$. Theorem 2.2 and Lemma A.2 imply the following theorem.

THEOREM 3.1. *Let the assumptions of Theorem* 2.2 *be fulfilled. Then* $(\sigma\sqrt{n})^{-1}T_n(\mathbf{r}_n) \circ F_n$ *converges weakly in* $D[a, b]$ *with respect to the uniform topology to the process* $B_{f, P_0} \circ F_0$ *where the limit process may be written as*

$$B_{f, P_0} \circ F_0(s) = B(F_0(s)) + \left(\int_{[a, s]} f(t)F_0(dt)\right)^{\top} \times J^{-1}\left(\int_{\mathscr{E}}^{(R)} B(F_0(t))f(dt) - B(1)f(Q_0(1))\right).$$

Next we give some other useful formulas for $B_{f, P_0} \circ F_0$. These formulas are consequences of well-known results of the theory of integration. Note that for parts (i) and (ii), integration by parts is used.

COROLLARY 3.2. *Under the assumptions of Theorem* 2.2, *the following formulas hold true for the limit process* $B_{f, P_0} \circ F_0$:

(i) $$B_{f, P_0} \circ F_0(s) = B(F_0(s)) + \left(\int_{[a, s]} f(t)F_0(dt)\right)^{\top} J^{-1}\left(\int_{\mathscr{E}}^{(R)} f(t)B(F_0(dt))\right).$$

(ii) $$B_{f, P_0} \circ F_0(s) = B(F_0(s)) - \left(\int_{[a, s]} F_0(t)\, f(dt) - F_0(s)f(s) + F_0(a)f(a)\right)^{\top} \times J^{-1}\left(\int_{\mathscr{E}}^{(R)} B(F_0(t))f(dt) - B(1)f(Q_0(1))\right).$$

(iii) *If additionally* F_0 *and* f *are absolutely continuous, then*

$$B_{f, P_0} \circ F_0(s) = B(F_0(s)) - \left(\int_{[a, s]} f(t)\phi_0(t)\, dt\right)^{\top} \times J^{-1}\left(\int_{\mathscr{E}}^{(R)} B(F_0(t))h(t)\, dt - B(1)f(Q_0(1))\right),$$

where $F_0(s) - F_0(a) = \int_{[a, s]} \phi_0(t)\, dt$ *and* $f(s) - f(a) = \int_{[a, s]} h(t)\, dt$.

REMARK 3.3. The most general result mentioned in MacNeill (1978b) is a special case of Corollary 3.2(iii). MacNeill (1978b) assumed: F_0 is absolutely continuous and strictly increasing and f is continuously differentiable.

APPENDIX

Let us denote the support of P_0 by D. In order to simplify the notation we assume that $a, b \in D$. Then $[a, b]\backslash D$ can be written as a disjoint sum of open intervals:

$$[a, b]\backslash D = \sum_{i=1}^{\infty}(a_i, b_i) \quad \text{with } a_i \leq b_i.$$

Indeed the above assumption can be supposed without loss of generality which can be seen by the same technique as used in part (b) of the proof below.

LEMMA A.1. *Let the conditions of Theorem* 2.2 *be fulfilled. Then for each* $u \in C[0, 1]$ *the following equation holds true*:

$$\lim_{n\to\infty} X_n^{\top} V_n(u) = f\big(Q_0(1)\big)u(1) - f\big(Q_0(0)\big)u(0) - \int_{\mathscr{E}}^{(R)} u(F_0(t))f(dt),$$

where V_n *is defined in the proof of Theorem* 2.2.

PROOF. The proof is divided in two parts depending on the structure of the support D of F_0.

(a) First, let the support D of F_0 be an interval. Then we have $\mathscr{E} = [a, b] = D$ because it is assumed $a, b \in D$.

Let us consider the design $(t_{n1}, \dots, t_{nn})$ and note that F_n is the distribution function corresponding to this design. Because $F_0|_D$ is strictly increasing there exist numbers

$$z_1(n) < \cdots < z_{l(n)}(n) \in F_0(D) = [0, 1]$$

with

$$\big\{Q_0\big(z_1(n)\big), \dots, Q_0\big(z_{l(n)}(n)\big)\big\} = \{t_{n1}, \dots, t_{nn}\},$$

where Q_0 is the quantile function of F_0. Thereby we obtain for fixed $u \in C[0, 1]$,

$$\begin{aligned} f(t_{nn})u(1) &- f(t_{n1})u(0) - X_n^{\top} V_n(u) \\ &= \sum_{i=2}^{n}(f(t_{ni}) - f(t_{ni-1}))u\left(\frac{i-1}{n}\right) \\ &= \sum_{j=2}^{l(n)} \big[f(Q_0(z_j(n))) - f(Q_0(z_{j-1}(n)))\big]\, u(F_n(Q_0(z_{j-1}(n)))) =: A_n. \end{aligned}$$

Consider the sum

$$\begin{aligned} &\sum_{j=2}^{l(n)} \big[f(Q_0(z_j(n))) - f(Q_0(z_{j-1}(n)))\big]\, u(F_0(Q_0(z_{j-1}(n)))) \\ &\qquad + [f(Q_0(z_1(n))) - f(a)]\, u(0) + \big[f(b) - f(Q_0(z_{l(n)}(n)))\big]\, u(1) =: I_n. \end{aligned}$$

In order to show the assertion of this section it suffices to prove (i) $|I_n - A_n| \to 0$ for $n \to \infty$ and (ii) $I_n \to \int_D^{(R)} u(F_0(t))f(dt)$ for $n \to \infty$.

(i) Let $\varepsilon > 0$. Then (2.1) implies

$$\exists\, n_0 \in \mathbb{N} \quad \forall\, n \geq n_0 \quad \forall\, t \in [a, b]\colon \big|u\big(F_0(t)\big) - u\big(F_n(t)\big)\big| < \varepsilon.$$

Further by (2.1) and because F_0 is strictly increasing, we have $Q_0(z_1(n)) \to a$, $Q_0(z_{l(n)}(n)) \to b$ for $n \to \infty$. Thus we obtain for sufficiently great n,

$$\begin{aligned}
|I_n - A_n| &= \Bigg|\sum_{j=2}^{l(n)} \big[f\big(Q_0(z_j(n))\big) - f\big(Q_0(z_{j-1}(n))\big)\big] \\
&\qquad \times \big[u\big(F_0(Q_0(z_{j-1}(n)))\big) - u\big(F_n(Q_0(z_{j-1}(n)))\big)\big] \\
&\qquad + \big[f\big(Q_0(z_1(n))\big) - f(a)\big]\, u(0) + \big[f(b) - f\big(Q_0(z_{l(n)}(n))\big)\big]\, u(1)\Bigg| \\
&\leq \varepsilon\Bigg[\sum_{j=2}^{l(n)} \big|f\big(Q_0(z_j(n))\big) - f\big(Q_0(z_{j-1}(n))\big)\big| + \big|u(0)\big| + \big|u(1)\big|\Bigg];
\end{aligned}$$

whence $|I_n - A_n| \to 0$ for $n \to \infty$ because f has bounded variation on $[a, b]$.

(ii) Let $\varepsilon > 0$. Let $u \in C[0, 1]$ be fixed. Then

$$\exists\, \delta \in (0, 1) \quad \forall\, y, z \in [0, 1]\colon |y - z| < \delta \quad \Rightarrow \quad |u(y) - u(z)| < \varepsilon.$$

Note, for each choice $s_1, \dots, s_m \in [a, b]$ with

$$a = s_1 < s_2 < \cdots < s_m = b,$$

we have

$$\sharp\big\{i \in \{1, \dots, m-1\} \big|\ \big|F_0(s_{i+1}) - F_0(s_i)\big| \geq \delta\big\} \leq \left[\frac{1}{\delta}\right] =: q.$$

Next

$$\exists\, \gamma > 0 \quad \forall\, s, t \in [a, b]\colon |s - t| < \gamma \quad \Rightarrow \quad \big|f(s) - f(t)\big| < \frac{\varepsilon}{q}.$$

Let us choose $s_1, \dots, s_m \in [a, b]$ such that

$$a = s_1 < s_2 < \cdots < s_m = b,$$
$$s_{i+1} - s_i < \gamma \quad \text{for } i = 1, \dots, m-1.$$

Then with

$$A := \big\{i \in \{1, \dots, m\} \big|\big|u(F_0(s_{i+1})) - u(F_0(s_i))\big| \geq \varepsilon\big\}$$

we obtain

$$\sum_{i \in A} \big|f(s_{i+1}) - f(s_i)\big| < q\frac{\varepsilon}{q} = \varepsilon.$$

Hence, condition (1.2.27) of Stroock (1994) is fulfilled, whence $u(F_0(\cdot))$ is Riemann–Stieltjes integrable with respect to f. [Note that the above fact can be shown in another way that we mention shortly. The function f corresponds to a signed measure μ, say. Because the discontinuities of $F_0(\cdot)$ are countable, we recognize that $u(F_0(\cdot))$ is μ-a.s. continuous. Hence, $u(F_0(\cdot))$ is Riemann–Stieltjes integrable with respect to f; see Stroock (1994), Theorem 5.12.]

Further, by (2.1) and because F_0 is strictly increasing we get

$$\exists\, n_0 \in \mathbb{N} \quad \forall\, n \geq n_0 \quad \forall\, j \in \{1, \dots, l(n)+1\}:$$
$$\lambda\big([Q_0(z_{j-1}(n)), Q_0(z_j(n))]\big) = Q_0\big(z_j(n)\big) - Q_0\big(z_{j-1}(n)\big) < \varepsilon,$$

where λ is the Lebesgue-measure on $\mathbb{R}$ and $z_0(n) = 0$, $z_{l(n)+1}(n) = 1$. Thus I_n converges to the Riemann–Stieltjes integral $\int_D^{(R)} u(F_0(t))f(dt)$.

(b) Next let the support D of P_0 be arbitrary. Let $\varepsilon > 0$. By (2.1) we can choose $n_0 \in \mathbb{N}$ such that $\forall\, n \geq n_0\ \forall\, t \in D\ \exists\, j \in \{1, \dots, n\}$: $|t - t_{nj}| < \varepsilon/2$. Next, we choose $s_1, \dots, s_r \in [a, b]\backslash D$ such that for the different elements $v_{n1} < \cdots < v_{nr(n)}$, say, of the set $\{t_{ni} \mid i = 1, \dots, n\} \cup \{s_j \mid j = 1, \dots, r\}$ the following hold true:

$$\forall\, n \geq n_0 \quad \forall\, i \in \big\{1, \dots, r(n)+1\big\}: v_{ni} - v_{ni-1} < \varepsilon \text{ and } v_{nr(n)} \in D,$$

where $v_{n0} = a$, $v_{nr(n)+1} = b$. Further we have then

$$\sum_{i=2}^{n} \big(f(t_{ni}) - f(t_{ni-1})\big) u\big(F_n(t_{ni-1})\big) = \sum_{i=2}^{r(n)} \big(f(v_{ni}) - f(v_{ni-1})\big) u\big(F_n(v_{ni-1})\big).$$

By these considerations we can ensure that the right-hand side of the above formula converges to $\int_{\mathscr{E}}^{(R)} u(F_0(t))f(dt)$ by an analogous argumentation as in (a) if $\max\{v_{ni} - v_{ni-1} \mid i = 1, \dots, r(n)+1\} \to 0$, which completes the proof. □

LEMMA A.2. *Let condition* (2.1) *be fulfilled, let* ξ_n, $n \in \mathbb{N}$, ξ *be random variables with values in* $C[0, 1]$, *and let* ξ_n *converge weakly to* ξ. *Then* $\xi_n(F_n(t))$ *converges weakly to* $\xi(F_0(t))$, $t \in [a, b]$, *as random variables with values in* $D[a, b]$ *which is furnished with the uniform metric.*

PROOF. Let us consider the mapping

$$G_n\colon \begin{cases} C[0, 1] \to D[a, b], \\ \quad x(\cdot) \mapsto x(F_n(\cdot)), \end{cases}$$

for each $n \in \mathbb{N} \cup \{0\}$. If x_k converges to x for $k \to \infty$ in $C[0, 1]$, then we have

$$\begin{aligned} &\sup_{t\in[a,b]} \big|G_n(x_n)(t) - G_0(x)(t)\big| \\ &\quad \leq \sup_{t\in[a,b]} \big|G_n(x_n)(t) - G_n(x)(t)\big| + \sup_{t\in[a,b]} \big|G_n(x)(t) - G_0(x)(t)\big| \\ &\quad = \sup_{t\in[a,b]} \big|x_n(F_n(t)) - x(F_n(t))\big| + \sup_{t\in[a,b]} \big|x(F_n(t)) - x(F_0(t))\big| \\ &\quad \leq \sup_{z\in[0,1]} \big|x_n(z) - x(z)\big| + \sup_{t\in[a,b]} \big|x(F_n(t)) - x(F_0(t))\big| \to 0 \end{aligned}$$

because x is uniformly continuous. Thus the assertion is proved by Theorem 5.5 of Billingsley (1968). □

REFERENCES

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

BISCHOFF, W. (1996). Properties of certain change-point test statistics and sample path behaviors of residual partial sums processes. Preprint.

BISCHOFF, W. and MILLER, F. (1998). Optimal tests and optimal designs for change-point problems in regression models. Preprint.

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.

CHERNOFF, H. and ZACKS, S. (1964). Estimating the current mean of a normal distribution which is subjected to changes in time. *Ann. Math. Statist.* **35** 999–1018.

GARDNER, L. A. (1969). On detecting changes in the mean of normal variates. *Ann. Math. Statist.* **40** 116–126.

HERZBERG, A. M. and TRAVES, W. N. (1994). An optimal experimental design for the Haar regression model. *Canad. J. Statist.* **22** 357–364.

JANDHYALA, V. K. (1993). A property of partial sums of least squares regression residuals and its applications. *J. Statist. Plann. Inference* **37** 317–326.

JANDHYALA, V. K. and MACNEILL, I. B. (1989). Residual partial sum limit process for regression models with applications to detecting parameter changes at unknown times. *Stochastic Process. Appl.* **33** 309–323.

JANDHYALA, V. K. and MACNEILL, I. B. (1991). Tests for parameter changes at unknown times in linear regressiom models. *J. Statist. Plann. Inference* **27** 291–316.

JANDHYALA, V. K. and MACNEILL, I. B. (1992). On testing for the constancy of regression coefficients under random walk and change-point alternatives. *Econometric Theory* **8** 501–517.

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

MACNEILL, I. B. (1978b). Limit processes for sequences of partial sums of regression residuals. *Ann. Probab.* **6** 695–698.

SEN, A. and SRIVASTAVA, M. S. (1975). On tests for detecting change in mean. *Ann. Statist.* **3** 98–108.

STROOCK, D. W. (1994). *A Concise Introduction to the Theory of Integration*. Birkhäuser, Boston.

TANG, S. M. and MACNEILL, I. B. (1993). The effect of serial correlation on tests for parameter change at unknown time. *Ann. Statist.* **21** 552–575.

WATSON, G. S. (1995). Detecting a change in the intercept in multiple regression. *Statist. Probab. Lett.* **23** 69–72.

INSTITUTE OF MATHEMATICAL STOCHASTICS
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF KARLSRUHE
D-76128 KARLSRUHE
GERMANY
E-MAIL: wolfgang.bischoff@math.uni-karlsruhe.de