JOURNAL OF MULTIVARIATE ANALYSIS 48, 43–69 (1994)

Limit Theorems for Change in Linear Regression

EDIT GOMBAY*

University of Alberta, Edmonton, Alberta, Canada

AND

LAJOS HORVÁTH

University of Utah

We consider some tests to detect a change-point in a multiple linear regression model. The tests are based on the maxima of the weighted cumulative sums processes. The limit distributions may be double exponential or maxima of Gaussian processes depending on the set where the maximum of the weighted cumulative sums of residuals is taken. The design-points can be fixed or random. We also give a few applications of our results.

1. INTRODUCTION

Let $y_1, y_2, ..., y_n$ be independent random variables. We assume that under H_0 they satisfy the linear model

$$y_i = \boldsymbol{\beta}\mathbf{x}_i^T + \varepsilon_i, \qquad 1 \leqslant i \leqslant n, \tag{1.1}$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_d)$ is an unknown vector and $\mathbf{x}_i = (1, x_{2,i}, ..., x_{d,i})$ are known design-points. We assume throughout this paper that the errors $\varepsilon_1, \varepsilon_2, ..., \varepsilon_n$ are independent identically distributed random variables with

$$E\varepsilon_i = 0, \quad 0 < \sigma^2 = \operatorname{var} \varepsilon_i < \infty \qquad \text{and} \qquad E|\varepsilon_i|^v < \infty \qquad \text{for some} \quad v > 2. \tag{1.2}$$

Received November 29, 1991; revised April 20, 1992.

1980 AMS subject classifications: primary 62G10, secondary 62J05.

Key words and phrases: linear model, change-point, strong approximation, Wiener process, Ornstein–Uhlenbeck process.

* Research supported in part by a NSERC Canada operating grant.

0047-259X/94 $6.00

Under H_1 there is a change in the linear model at the k^*th observation, i.e.,

$$y_i = \begin{cases} \boldsymbol{\beta} \mathbf{x}_i^T + \varepsilon_i, & 1 \leqslant i \leqslant k^* \\ \boldsymbol{\gamma} \mathbf{x}_i^T + \varepsilon_i, & k^* < i \leqslant n, \end{cases} \tag{1.3}$$

where k^* and $\boldsymbol{\gamma} \in R^d$ are unknown parameters and $\boldsymbol{\beta} \neq \boldsymbol{\gamma}$. Tests for H_0 in the case of simple regression models ($d = 2$) discussed by Quandt (1958, 1960) became the starting point of further research for change-point in linear models. For review and historical accounts we refer to Brown, Durbin, and Evans (1975) and Kim and Sigmund (1989).

Let

$$\bar{y}_k = \frac{1}{k} \sum_{1 \leqslant i \leqslant k} y_i,$$

$$\bar{\mathbf{x}}_k = \frac{1}{k} \sum_{1 \leqslant i \leqslant k} \mathbf{x}_i,$$

$$Q_n = \sum_{1 \leqslant i \leqslant n} (\mathbf{x}_i - \bar{\mathbf{x}}_n)(\mathbf{x}_i - \bar{\mathbf{x}}_n)^T$$

and

$$X_n = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{bmatrix}, \qquad Y_n = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

The least-squares estimator of $\boldsymbol{\beta}$ is denoted by

$$\hat{\boldsymbol{\beta}}_n = (\beta_{1n}, ..., \beta_{dn}) = (X_n^T X_n)^{-1} X_n^T Y_n.$$

James, James, and Siegmund (1987) suggested that we reject H_0 for large values of

$$\max_{n_1 \leqslant k \leqslant n_2} |U_n(k)|, \tag{1.4}$$

where

$$U_n(k) = \left(\frac{k}{1 - k/n} \right)^{1/2} \frac{\bar{y}_k - \bar{y}_n - \hat{\boldsymbol{\beta}}_n (\bar{\mathbf{x}}_k - \mathbf{x}_n)^T}{(1 - k(\bar{\mathbf{x}}_k - \bar{\mathbf{x}}_n)(\bar{\mathbf{x}}_k - \bar{\mathbf{x}}_n)^T / (Q_n (1 - k/n)))^{1/2}}.$$

Brown, Durbin, and Evans (1975), among others, mention the possibility of using functionals of the process of cumulative sums of residuals

$$\sum_{1 \leqslant i \leqslant k} (y_i - \bar{y}_n - \hat{\boldsymbol{\beta}}_n (\mathbf{x}_i - \bar{\mathbf{x}}_n)^T), \qquad 1 \leqslant k \leqslant n. \tag{1.5}$$

For a discussion on residuals and their applications in statistics we refer to Cook and Weisberg (1982) and Davison and Snell (1991). It is easy to see that

$$U_n(k) = w_n(k)\, R_n(k), \tag{1.6}$$

where

$$R_n(k) = \left(\frac{n}{k(n-k)}\right)^{1/2} \sum_{1 \leqslant i \leqslant k} (y_i - \bar{y}_n - \hat{\boldsymbol{\beta}}_n (\mathbf{x}_i - \bar{\mathbf{x}}_n)^T) \tag{1.7}$$

and

$$w_n(k) = (1 - k(\bar{\mathbf{x}}_k - \bar{\mathbf{x}}_n)(\bar{\mathbf{x}}_k - \bar{\mathbf{x}}_n)^T / (Q_n(1 - k/n)))^{-1/2}. \tag{1.8}$$

Thus $U_n(k)$ can be considered as weighted normalized cumulative sums, where the weight is $w_n(k)$ of (1.8). Brown *et al.* (1975) noted that $R_n(nt)$ converges weakly to a Gaussian process in $\mathscr{D}[t_1, t_2]$, $0 < t_1 < t_2 < 1$. However, the limit process is so complicated that it is very unlikely to yield exact formulas for the distribution functions of its functionals. Kim and Siegmund (1989) made a similar remark for the limit of $U_n(nt)$ in $\mathscr{D}[t_1, t_2]$, $0 < t_1 < t_2 < 1$. Also, it is very easy to check that the limits of $U_n(nt)$ and $R_n(nt)$ must be different. If we are interested in the weak convergence of $U_n(nt)$ it is essential that we consider these processes in $\mathscr{D}[t_1, t_2]$, $0 < t_1 < t_2 < 1$. It follows from the main results in Sections 2 and 3 that $\max_{1 \leqslant k < n} |U_n(k)| \xrightarrow{P} \infty$ and $\max_{1 \leqslant k < n} |R_n(k)| \xrightarrow{P} \infty$, and therefore $U_n(nt)$ and $R_n(nt)$ cannot converge weakly in $\mathscr{D}[0, 1]$ (cf. also Maronna and Yohai, 1978).

Brown *et al.* (1975) suggested the functionals of $U_n(k)$ and $R_n(k)$ as test statistics without any motivation and specifying the alternative hypothesis. Using maximum likelihood arguments, Kim and Siegmund (1989) derived $\max_{1 \leqslant k < n} |U_n(k)|$ to detect a change only in the intercept term of regression. Quandt (1958, 1960), Maronna and Yohai (1978), and Worsley (1983) derived the maximum likelihood ratio test against H_1 assuming that the errors are normal random variables. The distributions of the statistics in Quandt (1958, 1960) and Worsley (1983) are unknown. We show (cf. Theorem 2.3 and the remark at the end of Section 3) that the statistics $\max_{1 \leqslant k < n} |U_n(k)|$ and $\max_{1 \leqslant k < n} |R_n(k)|$ suggested by Brown *et al.* (1975) are consistent against H_1.

Kim and Siegmund (1989) studied the distribution of $\max_{nt_1 \leqslant k \leqslant nt_2} |U_n(k)|/\hat{\sigma}$, $0 < t_1 < t_2 < 1$, when $d = 2$ ($\hat{\sigma}^2$ is the least-squares estimator of σ^2.) They obtained approximations for the significance levels (cf. also Hawkins, 1989). We believe that it is more natural to consider $\max_{1 \leqslant k < n} |U_n(k)|$ than $\max_{nt_1 \leqslant k \leqslant nt_2} |U_n(k)|$, $0 < t_1 < t_2 < 1$. It is not clear how to choose t_1, t_2 and our decision (rejection–acceptance) may depend on the choice of these parameters. In this paper we study the asymptotics of

$$Z_n(i, j) = \max_{i \leqslant k < j} |U_n(k)| \tag{1.9}$$

and

$$T_n(i, j) = \max_{i \leqslant k < j} |R_n(k)| \tag{1.10}$$

for various choices of i and j. It turns out that under H_0, $Z_n(1, n)$ and $T_n(1, n)$ have the same limit distribution. The "middle" part $Z_n(nt_1, nt_2)$, $0 < t_1 < t_2 < 1$, does not contribute to the distribution of $Z_n(1, n)$ and the random variables $Z_n(nt_1, nt_2)$ $(0 < t_1 < t_2 < 1)$, $Z_n(1, n)$ are asymptotically indepenent if H_0 holds. We prove in Theorem 2.3 that our statistics are consistent against H_1.

In the next section we investigate the distributions of $Z_n(i, j)$ and $T_n(i, j)$ when $\mathbf{x}_i$, $1 \leqslant i \leqslant n$, the design points are non-random. Section 3 contains similar results assuming that $\mathbf{x}_i$, $1 \leqslant i \leqslant n$, are random variables. We discuss a few applications of our results in Section 4 and the proofs are presented in the last section.

2. Non-Random Designs

In this section we assume that there is a function $\mathbf{f}(t) = (f_2(t), ..., f_d(t))$ such that

$$x_{j,i} = f_j(i/n), \qquad 2 \leqslant j \leqslant d, \quad 1 \leqslant i \leqslant n. \tag{2.1}$$

We assume a few regularity conditions,

$$\max_{2 \leqslant j \leqslant d} \sup_{0 \leqslant t \leqslant 1} |f_j'(t)| < \infty, \tag{2.2}$$

$$\text{the rank of } A = \{\delta_{i,j}, 1 \leqslant i, j \leqslant d\} \text{ is } d, \tag{2.3}$$

where

$$\delta_{1,1} = 1,$$

$$\delta_{1,j} = \delta_{j,1} = \int_0^1 f_j(t)\,dt, \qquad 2 \leqslant j \leqslant d,$$

$$\delta_{i,j} = \delta_{j,i} = \int_0^1 f_i(t)\,f_j\,dt, \qquad 2 \leqslant i, j \leqslant d,$$

and

$$\sup_{0 \leqslant t \leqslant 1} g(t) < 1, \tag{2.4}$$

where

$$g(t) = \frac{1}{Q(1-t)} \sum_{2 \leqslant j \leqslant d} \left\{ \frac{1}{t} \int_0^t f_j(s)\,ds - \int_0^1 f_j(s)\,ds \right\}^2,$$

$$Q = \sum_{2 \leqslant j \leqslant d} \left\{ \int_0^1 f_j^2(s)\,ds - \left(\int_0^1 f_j(s)\,ds \right)^2 \right\}.$$

If $d = 2$, then (2.3) means that f is not constant on $[0, 1]$. Condition (2.4) implies that $w_n(k) < \infty$ for each $1 \leqslant k < n$.

Let

$$a(x) = (2 \log x)^{1/2}$$

and

$$b(x) = 2 \log x + \tfrac{1}{2} \log \log x - \tfrac{1}{2} \log \pi.$$

THEOREM 2.1. *We assume that* (1.1), (1.2), *and* (2.1)–(2.4) *hold. Then we have*

$$\lim_{n \to \infty} P\left\{ \frac{1}{\sigma} a(\log n)\, Z_n(1, n) \leqslant x + b(\log n) \right\} = \exp(-2e^{-x}) \tag{2.5}$$

and

$$\lim_{n \to \infty} P\left\{ \frac{1}{\sigma} a(\log n)\, T_n(1, n) \leqslant x + b(\log n) \right\} = \exp(-2e^{-x}). \tag{2.6}$$

Also, if $\lambda_1(n) \to 0$, $\lambda_2(n) \to 0$, $m_1 = n\lambda_1(n) \to \infty$, *and* $m_2 = n\lambda_2(n) \to \infty$, *then we have*

$$\lim_{n\to\infty} P\left\{\frac{1}{\sigma} a\left(\frac{1}{2}\log m_1\right) Z_n(1, m_1) \leqslant x + b\left(\frac{1}{2}\log m_1\right)\right\}$$
$$= \exp(-2e^{-x}), \tag{2.7}$$

$$\lim_{n\to\infty} P\left\{\frac{1}{\sigma} a\left(\frac{1}{2}\log m_1\right) T_n(1, m_1) \leqslant x + b\left(\frac{1}{2}\log m_1\right)\right\}$$
$$= \exp(-2e^{-x}), \tag{2.8}$$

$$\lim_{n\to\infty} P\left\{\frac{1}{\sigma} a\left(\frac{1}{2}\log m_2\right) Z_n(n-m_2, n) \leqslant x + b\left(\frac{1}{2}\log m_2\right)\right\}$$
$$= \exp(-2e^{-x}), \tag{2.9}$$

$$\lim_{n\to\infty} P\left\{\frac{1}{\sigma} a\left(\frac{1}{2}\log m_2\right) T_n(n-m_2, n) \leqslant x + b\left(\frac{1}{2}\log m_2\right)\right\}$$
$$= \exp(-2e^{-x}). \tag{2.10}$$

We note that Huškavá (1993) also proved (2.6). Theorem 2.1 shows that $Z(1, n)$ and $T(1, n)$ are determined by the first and the last elements of the maxima. We use large weights on the tails and therefore tests based on (2.5) and (2.6) are more powerful on the tails (k^* is too small or large) and less powerful in the middle. James *et al.* (1987) also pointed out that the restriction of the maximum to $[nt_1, nt_2]$ increases the power to detect a change occurring near $n/2$ and without restriction the power attains the minimum when change occurs near the middle. However, it is very easy to combine Theorem 1 of Kim and Siegmund (1989) with our Theorem 2.1 and we can obtain some asymptotic tests which are powerful on the tails as well as in the middle. Also, it is an interesting observation that $Z_n(1, n)$ and $Z_n(nt_1, nt_2)$ are asymptotically independent. We have the same result for $T_n(1, n)$ and $T_n(nt_1, nt_2)$.

THEOREM 2.2. *We assume that* (1.1), (1.2), (2.1)–(2.4) *hold and* $0 < t_1 < t_2 < 1$. *Then we have*

$$\lim_{n\to\infty} P\left\{\frac{1}{\sigma} a(\log n)\, Z_n(1, n) \leqslant x + b(\log n), \frac{1}{\sigma} Z_n(nt_1, nt_2) \leqslant y\right\}$$
$$= \exp(-2e^{-x}) \lim_{n\to\infty} P\left\{\frac{1}{\sigma} Z_n(nt_1, nt_2) \leqslant y\right\} \tag{2.11}$$

and

$$\lim_{n\to\infty} P\left\{\frac{1}{\sigma} a(\log n)\, T_n(1, n) \leqslant x + b(\log n), \frac{1}{\sigma} T_n(nt_1, nt_2) \leqslant y\right\}$$
$$= \exp(-2e^{-x}) \lim_{n\to\infty} P\left\{\frac{1}{\sigma} T_n(nt_1, nt_2) \leqslant y\right\}. \tag{2.12}$$

Kim and Siegmund (1989) give an approximation for the tail percentiles of the distribution functions of $(1/\sigma) Z_n(nt_1, nt_2)$ and $(1/\sigma) T_n(nt_1, nt_2)$. Hence, we can compute or at least approximate the limit distributions in Theorem 2.2. We also would like to point out that the limit distributions of $Z_n(1, n)$, $T_n(1, n)$ and also $Z_n(1, m_1)$, $Z_n(n-m_2, n)$, $T_n(1, m_1)$, $T_n(n-m_2, n)$ do not depend on the design-points, i.e., on the unknown **f**. However, the limits of $Z_n(nt_1, nt_2)$ and $T_n(nt_1, nt_2)$ *do* depend on **f**; therefore we must specify **f**, if our test is based on $Z_n(nt_1, nt_2)$ or $T_n(nt_1, nt_2)$. We can use Theorem 1 of Kim and Siegmund (1989) and Theorem 2.2 to obtain a test with good properties against changes on the tails or in the middle. We reject H_0 if $(1/\sigma)\, a(\log n)\, Z_n(1, n) - b(\log n)$ or $(1/\sigma)\, Z_n(nt_1, nt_2)$ are large. Using the independence in Theorem 2.2 and the approximate critical values in Kim and Siegmund (1989) we can construct the rejection region.

Next we discuss the consistency of the tests based on Theorems 2.1 and 2.2. For the sake of simplicity we assume that $d=2$. Thus under the alternative we have

$$y_i = \begin{cases} \beta_1 + \beta_2 x_i + \varepsilon_i, & 1 \leqslant i \leqslant [n\tau] \\ \gamma_1 + \gamma_2 x_i + \varepsilon_i, & [n\tau]+1 \leqslant i \leqslant n, \end{cases} \tag{2.13}$$

where $(\beta_1, \beta_2) \neq (\gamma_1, \gamma_2)$ and

$$x_i = f(i/n), \qquad 1 \leqslant i \leqslant n. \tag{2.14}$$

THEOREM 2.3. *We assume that* (1.2), (2.13), (2.14) *hold and f is a continuous and increasing function on* $[0, 1]$. *Then there is a $\lambda_0 \in (0, 1)$ such that*

$$\liminf_{n \to \infty} |ER_n(n\lambda_0)|/n^{1/2} > 0 \tag{2.15}$$

and

$$\liminf_{n \to \infty} |EU_n(n\lambda_0)|/n^{1/2} > 0. \tag{2.16}$$

It is clear that consistency follows from Theorem 2.3.

3. Randomly Distributed Designs

We assume in this section that the design-points $\{x_i, 1 \leqslant i \leqslant n\}$ are random. For the sake of simplicity we assume $d=2$, i.e., we consider the simple linear model

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i, \qquad 1 \leqslant i \leqslant n, \tag{3.1}$$

where $\varepsilon_1, \varepsilon_2, ..., \varepsilon_n$ are independent identically distributed random errors satisfying (1.2). The least-squares estimators of β_1 and β_2 are denoted by $\hat{\beta}_{1n}$ and $\hat{\beta}_{2n}$.

First, we consider the case when x_i is given by

$$x_i = f(\xi_i), \tag{3.2}$$

where

$$\xi, \xi_1, \xi_2, \ldots \text{ are independent, identically distributed random variables, } 0 < \xi < 1 \text{ and } E\xi = \mu > 0. \tag{3.3}$$

We also assume that

$$\{\varepsilon_i,\ 1 \leqslant i \leqslant n\} \text{ and } \{\xi_i,\ 1 \leqslant i \leqslant n\} \text{ are independent.} \tag{3.4}$$

The design function f of (3.2) satisfies the regularity conditions

$$\sup_{0 \leqslant t \leqslant 1} |f'(t)| < \infty \tag{3.5}$$

and

$$0 < \operatorname{var} f(\xi) < \infty. \tag{3.6}$$

The next result shows that Theorem 2.1 remains true when the design-points are random.

THEOREM 3.1. *We assume that* (1.2) *and* (3.1)–(3.6) *hold. Then we have* (2.5)–(2.10).

We can also prove an analogue of Theorem 2.2. Let $\{B(t),\ 0 \leqslant t \leqslant 1\}$ be a Brownian bridge.

THEOREM 3.2. *We assume that* (1.2), (3.1)–(3.6) *hold and* $0 < t_1 < t_2 < 1$. *Then we have*

$$\lim_{n \to \infty} P\left\{\frac{1}{\sigma} a(\log n)\, Z_n(1, n) \leqslant x + b(\log n),\ \frac{1}{\sigma} Z_n(nt_1, nt_2) \leqslant y\right\}$$
$$= \exp(-2e^{-x})\, P\{\sup_{t_1 \leqslant t \leqslant t_2} |B(t)|/(t(1-t))^{1/2} \leqslant y\} \tag{3.7}$$

and

$$\lim_{n \to \infty} P\left\{\frac{1}{\sigma} a(\log n)\, T_n(1, n) \leqslant x + b(\log n),\ \frac{1}{\sigma} T_n(nt_1, nt_2) \leqslant y\right\}$$
$$= \exp(-2e^{-x})\, P\{\sup_{t_1 \leqslant t \leqslant t_2} |B(t)|/(t(1-t))^{1/2} \leqslant y\}. \tag{3.8}$$

A minor modification gives that in Theorem 3.2 the weighted Brownian bridge can be replaced by a Brownian bridge which makes the applications easier.

Remark 3.1. If the conditions of Theorem 3.1 hold, then we have

$$\lim_{n\to\infty} P\left\{\frac{1}{\sigma} a(\log n)\, Z_n(1,n) \leqslant x + b(\log n),\right.$$
$$\left.\frac{1}{\sigma}\max_{1\leqslant k<n} n^{-1/2}\left|\sum_{1\leqslant i\leqslant k}(y_i-\bar{y}_n-\hat{\beta}_{2n}(x_i-\bar{x}_n))\right|\leqslant y\right\}$$
$$=\exp(-2e^{-x})\, P\{\sup_{0\leqslant t\leqslant 1}|B(t)\leqslant y\} \tag{3.9}$$

and

$$\lim_{n\to\infty} P\left\{\frac{1}{\sigma} a(\log n)\, T_n(1,n) \leqslant x + b(\log n),\right.$$
$$\left.\frac{1}{\sigma}\max_{1\leqslant k\leqslant n} n^{-1/2}\left|\sum_{1\leqslant i\leqslant k}(y_i-\bar{y}_n-\hat{\beta}_{2n}(x_i-\bar{x}_n))\right|\leqslant y\right\}$$
$$=\exp(-2e^{-x})\, P\{\sup_{0\leqslant t\leqslant 1}|B(t)|\leqslant y\}. \tag{3.10}$$

Next we consider a model in which the design-points x_i, $1\leqslant i\leqslant n$, are random and may be an increasing function of i. We assume that

$$x_i=f(\eta_i/n), \qquad \text{where} \quad \eta_i=\xi_1+\xi_2+\cdots+\xi_i. \tag{3.11}$$

We also need some regularity conditions on f,

$$\sup_{0\leqslant t\leqslant 1}|f''(t)|<\infty, \tag{3.12}$$

$$\frac{1}{\mu}\int_0^{\mu} f^2(u)\,du > \left(\frac{1}{\mu}\int_0^{\mu} f(u)\,du\right)^2 \tag{3.13}$$

and

$$\sup_{0\leqslant t\leqslant 1} r(t)<1, \tag{3.14}$$

where

$$r(t)=\frac{t((1/t)\int_0^t f(\mu s)\,ds-\int_0^1 f(\mu s)\,ds)^2}{(1-t)(\int_0^1 f^2(\mu s)\,ds-(\int_0^1 f(\mu s)\,ds)^2)}.$$

The following theorem shows that it does not make any difference whether x_i, $1\leqslant i\leqslant n$, are defined by (2.1), (3.2), or (3.11).

THEOREM 3.3. *We assume that* (1.2), (3.1), (3.3)–(3.5), *and* (3.11)–(3.14) *hold. Then we have* (2.5)–(2.10).

Conditions of Theorem 3.3 are very similar to the conditions in the non-random case. This is not surprising since in this case the random points in (3.11) can be replaced by non-random points and still have the same limit theorems. Let

$$\hat{f}(t) = f(t\mu) \tag{3.15}$$

and

$$\hat{x}_i = \hat{f}(i/n) = f(i\mu/n). \tag{3.16}$$

We again consider (3.1) but x_i is replaced by $\hat{x}_i$,

$$y_i = \beta_1 + \beta_2 \hat{x}_i + \varepsilon_i, \qquad 1 \leqslant i \leqslant n. \tag{3.17}$$

We compute the statistics in (1.9) and (1.10) using the non-random points in (3.16). The corresponding statistics are denoted by $\hat{Z}_n(i, j)$ and $\hat{T}_n(i, j)$. It turns out, for example, that $\hat{Z}_n(1, n)$ and $Z_n(1, n)$ are asymptotically the same and similarly $\hat{Z}_n(nt_1, nt_2)$ and $Z_n(nt_1, nt_2)$ have the same limit distribution.

THEOREM 3.4. *We assume that* (1.2), (3.1), (3.3)–(3.5), (3.11)–(3.14) *hold and* $0 < t_1 < t_2 < 1$. *Then we have*

$$\lim_{n \to \infty} P\left\{\frac{1}{\sigma} a(\log n)\, \hat{Z}_n(1, n) \leqslant x + b(\log n), \frac{1}{\sigma} \hat{Z}_n(nt_1, nt_2) \leqslant y\right\}$$
$$= \exp(-2e^{-x}) \lim_{n \to \infty} P\left\{\frac{1}{\sigma} \hat{Z}_n(nt_1, nt_2) \leqslant y\right\} \tag{3.18}$$

and

$$\lim_{n \to \infty} P\left\{\frac{1}{\sigma} a(\log n)\, \hat{T}_n(1, n) \leqslant x + b(\log n), \frac{1}{\sigma} \hat{T}_n(nt_1, nt_2) \leqslant y\right\}$$
$$= \exp(-2e^{-x}) \lim_{n \to \infty} P\left\{\frac{1}{\sigma} \hat{T}_n(nt_1, nt_2) \leqslant y\right\}. \tag{3.19}$$

Kim and Siegmund (1989) can be used again to obtain selected values of $P\{(1/\sigma)\, \hat{T}_n(nt_1, nt_2) \leqslant y\}$ and $P\{(1/\sigma) \hat{Z}_n(nt_1, nt_2) \leqslant y\}$.

Comparing the results of Sections 2 and 3 we can observe that the limit distributions of $Z_n(1, n)$ and $T_n(1, n)$ do not depend on the assumptions on the desing-points. We have the same limits in the non-random as in the random cases. However, we obtained completely different limit distributions for $Z_n(nt_1, nt_2)$ and $T_n(nt_1, nt_2)$. It is well known that

$$\sup_{t_1 \leqslant t \leqslant t_2} |B(t)|/(t(1-t))^{1/2} \overset{\mathscr{D}}{=} \sup_{0 \leqslant t \leqslant (\log(((1-t_1)/t_1)(t_2/(1-t_2))))/2} |V(t)|,$$

where $\{V(t), -\infty < t < \infty\}$ is an Ornstein–Uhlenbeck process with parameter 1 (i.e., $EV(t)\,V(s) = \exp(-|t-s|)$). Thus the tables in DeLong (1981) can be used to obtain the numerical values of the limit distributions (3.7) and (3.8).

We note hat Theorem 2.3 remains valid in the case of randomly distributed design-points.

4. Applications

We discuss a few applications of our results. The limit theorems in Sections 2 and 3 contain the unknown variance σ^2. The mean-square estimator $\hat{\sigma}_n^2$ of σ^2 is an unbiased estimator of σ^2 under H_0. Assuming that $v \geqslant 4$ in (1.2), one can easily show that our results remain true when σ^2 is replaced by $\hat{\sigma}_n^2$. However, under H_1, $\hat{\sigma}_n^2$ grossly overestimates σ^2, which may cause a failure to detect that H_0 is false. Under H_1 we have only one change-point, so we can use the following estimator of σ^2. We split the data into two subsets $\{(y_i, \mathbf{x}_i), 1 \leqslant i < n/2\}$ and $\{(y_i, \mathbf{x}_i)\ n/2 \leqslant i \leqslant n\}$. Let $\hat{\sigma}_{n,1}^2$ be the mean-square estimator of σ^2 from $\{(y_i, \mathbf{x}_i), 1 \leqslant i < n/2\}$ and $\hat{\sigma}_{n,2}^2$ be the mean-square estimator of σ^2 from the rest of the data. We define $\sigma_n^{*2} = \min(\hat{\sigma}_{n,1}^2, \hat{\sigma}_{n,2}^2)$. An elementary exercise shows that $|\sigma_n^{*2} - \sigma^2| = O_P(n^{-1/2})$ under H_0 as well as H_1, if $v \geqslant 4$ in (1.2). Hence our results remain true if σ is replaced by σ_n^*. (If we have at most k change points, the data must be split into $k+1$ subsets, and the estimator is the smallest mean-square estimator from the subsets.)

Our first example is the Old Faithful geyser. A geyser is a hot spring that occasionally becomes unstable and erupts hot water and steam into the air. One of th most well known geysers is Old Faithful in Yellowstone National Park (U.S.A.). Old Faithful erupts at an interval of about 40–100 min, with eruptions lasting from 1 to 6 min. Cook and Weisberg (1982) contains a scatter plot of $y =$ interval to the next eruption versus $x =$ duration of current eruption for 237 eruptions of Old Faithful in October 1980. The scatter plot suggests that the duration can be divided into two categories, short and long durations. We asked whether we should fit separate lines to the short and long durations. We used $Z_{237}(1, 237)/\sigma_{237}^*$ and the observed value of the properly scaled and shifted $Z(1, n)$ was 2.37, giving a p-value 0.17, so we can use a linear model to predict the next eruption of the geyser.

The next example uses multiple linear regression. Moore and McCabe (1989) report data which were collected on 224 freshmen computer science majors at a large university. Of particular interest was the cumulative grade point average after three semesters. The average high school in mathematics (x_2), science (x_3), and English (x_4) were recorded, and therefore our design vector is $\mathbf{x} = (1, x_2, x_3, x_4)$. Moore and McCabe (1989) suggest that multiple linear regression can be fitted to these data. First, we ordered these data according to the scores in mathematics. The rejection of H_0 should mean that we must use different multiple linear regressions for low and high scores in mathematics. We computed $Z_{224}(1, 224)/\sigma^*_{224} = 3.11$ and the corresponding p-value, calculated from the extreme value limit theorem, is 0.13. Then we ordered the data according to the scores in science and English and obtained 0.15 and 0.25 as p-values, respectively. This means that the same multiple linear regression can be used for all scores.

Finally, we considered the model of an index of gross domestic product in the United States as depending on labor-input index (x_2) and capital-input index (x_3) between 1929 and 1967. Maddala (1977) considered the log-linear model. His F-test was not significant, contrary to his expectations. Our asymptotic test gave p-value 0.003, clearly indicating that the coefficients are not stable.

We also performed a small scale Monte-Carlo simulation. We assumed that the errors follow a triangular distribution with density $1 + x$ if $-1 < x \leqslant 0$ and $1 - x$ if $0 < x < 1$. We considered the case of non-stochastic designs as well as random x_i's. The sample sizes were in the range 40–200, and the simulations were run 50 times in each case. We considered the simple regression model ($d = 1$) and multiple regressions ($d = 2, 3$). The outcome of the simulations did not depend on d. Under H_0 we could not reject H_0 at the 0.1 significance level at 95% of the simulations. The behavior of the rest was very good under H_1. We always rejected H_0 when H_1 was true at the 0.05 significance level. It seems to us that practically we always reject H_0 when it is false. Thus these tests are conservative.

The distribution of the maximum of a Gaussian process converges slowly to the double exponential distribution. It follows from the proofs that the double exponential limit distribution comes from the result that it is the limit of the distribution function of the suitably normalized $\sup_{1/n \leqslant t \leqslant 1 - 1/n} |B(t)|/(t(1-t))^{1/2}$, where $\{B(t), 0 \leqslant t \leqslant 1\}$ is a Brownian bridge. In the case of small sample sizes ($50 \leqslant n \leqslant 200$) the discretized version of $|B(t)|/(t(1-t))^{1/2}$, as in Yao and Davis (1986), provides a better approximation for the critical values.

5. Proofs

We start with a well-known result.

Lemma 5.1. *We assume that* (1.1), (1.2), *and* (2.1)–(2.4) *hold. Then, as* $n \to \infty$, $n^{1/2}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ *goes in distribution to a d-dimensional normal random variable with mean* $\mathbf{0}$ *and covariance matrix* $\sigma^2 A^{-1}$.

We need only a very simple consequence of Lemma 5.1. Let $\boldsymbol{\beta}_n^* = (\beta_{2n}, ..., \beta_{dn})$ and $\boldsymbol{\beta}^* = (\beta_2, ..., \beta_d)$. By Lemma 5.1 we have

$$n^{1/2}(\boldsymbol{\beta}_n^* - \boldsymbol{\beta}^*) = O_P(1). \tag{5.1}$$

Using the definitions of R_n and $\boldsymbol{\beta}_n^*$ we can write

$$R_n(k) = V_n(k) - A_n(k), \tag{5.2}$$

where

$$V_n(k) = \left(\frac{n}{k(n-k)}\right)^{1/2} \left(\sum_{1 \leqslant i \leqslant k} \varepsilon_i - \frac{k}{n} \sum_{1 \leqslant i \leqslant n} \varepsilon_i\right),$$

$$A_n(k) = \left(\frac{n}{k(n-k)}\right)^{1/2} k(\boldsymbol{\beta}_n^* - \boldsymbol{\beta}^*)(\bar{\mathbf{x}}_k^* - \bar{\mathbf{x}}_n^*)^T,$$

and $\mathbf{x}_i^* = (x_{2,i}, ..., x_{d,i})$, $\bar{\mathbf{x}}_k^* = (1/k) \sum_{1 \leqslant i \leqslant k} \mathbf{x}_i^*$. The distribution of $\max |V_n(k)|$ is well known (cf., for example, Csörgő and Horváth (1988)) and it is summarized in the following lemma.

Lemma 5.2. *We assume that* (1.2) *holds. Then we have*

$$\lim_{n \to \infty} P\left\{\frac{1}{\sigma} a(\log n) \max_{1 \leqslant k < n} |V_n(k)| \leqslant x + b(\log n)\right\} = \exp(-2e^{-x}). \tag{5.3}$$

If $\lambda_1(n) \to 0$, $\lambda_2(n) \to 0$, $m_1 = n\lambda_1(n) \to \infty$, *and* $m_2 = n\lambda_2(n) \to \infty$, *then we have*

$$\lim_{n \to \infty} P\left\{\frac{1}{\sigma} a\left(\frac{1}{2}\log m_1\right) \max_{1 \leqslant k \leqslant m_1} |V_n(k)| \leqslant x + b\left(\frac{1}{2}\log m_1\right)\right\}$$
$$= \exp(-2e^{-x}) \tag{5.4}$$

and

$$\lim_{n \to \infty} P\left\{\frac{1}{\sigma} a\left(\frac{1}{2}\log m_2\right) \max_{n - m_2 \leqslant k < n} |V_n(k)| \leqslant x + b\left(\frac{1}{2}\log m_2\right)\right\}$$
$$= \exp(-2e^{-x}). \tag{5.5}$$

If

$$\gamma(n)=\frac{c_2(n)}{n-c_2(n)}\frac{n-c_1(n)}{c_1(n)}\to\infty,$$

then we have

$$\lim_{n\to\infty} P\left\{\frac{1}{\sigma}a\left(\frac{1}{2}\log\gamma_n\right)\max_{c_1\leqslant k<c_2}|V_n(k)|\leqslant x+b\left(\tfrac{1}{2}\log\gamma_n\right)\right\}$$
$$=\exp(-2e^{-x}). \tag{5.6}$$

The limit theorems for $\max|R_n(k)|$ will follows from Lemma 5.2 if we can show that $A_n(k)$ is negligible. We also prove that $\max|R_n(k)|$ and $\max|U_n(k)|$ have the same limit distributions. These proofs are based on the following technical lemma.

LEMMA 5.3. *We assume that* (2.1)–(2.4) *hold. Then we have*

$$\max_{1\leqslant k<n}\left(\frac{n}{n-k}\right)^2(\bar{\mathbf{x}}_k-\bar{\mathbf{x}}_n)(\bar{\mathbf{x}}_k-\bar{\mathbf{x}}_n)^T\leqslant C_1, \tag{5.7}$$

$$\max_{1\leqslant k<n}|w_n(k)-(1-g(k/n))^{-1/2}|\leqslant C_2/n \tag{5.8}$$

and

$$\max_{1\leqslant k<n}\frac{n^2}{k(n-k)}|1-(1-g(k/n))^{-1/2}|\leqslant C_3, \tag{5.9}$$

where C_1, C_2, *and* C_3 *are constants.*

Proof. It follows from (2.1) and (2.2) that

$$\begin{aligned}|\bar{x}_{j,k}-\bar{x}_{j,n}|&=\left|\frac{1}{k}\sum_{1\leqslant i\leqslant k}f_j(i/n)-\frac{1}{n}\sum_{1\leqslant i\leqslant n}f_j(i/n)\right|\\&\leqslant\left(\frac{1}{k}-\frac{1}{n}\right)\left|\sum_{1\leqslant i\leqslant k}f_j(i/n)\right|+\frac{1}{n}\left|\sum_{k<i\leqslant n}f_j(i/n)\right|\\&\leqslant C_4(n-k)/n,\end{aligned} \tag{5.10}$$

which implies (5.7). Similar arguments give

$$\left|(\bar{\mathbf{x}}_k-\bar{\mathbf{x}}_n)(\bar{\mathbf{x}}_k-\bar{\mathbf{x}}_n)^T-\sum_{2\leqslant j\leqslant d}\left\{\frac{n}{k}\int_0^{k/n}f_j(t)\,dt-\int_0^1 f_j(t)\,dt\right\}^2\right|$$
$$=\left|(\bar{\mathbf{x}}_k-\bar{\mathbf{x}}_n)(\bar{\mathbf{x}}_k-\bar{\mathbf{x}}_n)^T-\sum_{2\leqslant j\leqslant d}\left\{\frac{n-k}{n}\int_0^{k/n}f_j(t)\,dt-\int_{k/n}^1 f_j(t)\,dt\right\}^2\right|$$
$$\leqslant C_5(n-k)/n^2. \tag{5.11}$$

It is easy to see that

$$Q_n/n = Q + O(1/n). \tag{5.12}$$

Hence, from (5.11) and (5.12) we obtain

$$\left|\frac{k(\bar{\mathbf{x}}_k - \bar{\mathbf{x}}_n)(\bar{\mathbf{x}}_k - \bar{\mathbf{x}}_n)^T}{(1+k/n)\,Q_n} - g(k/n)\right| \leqslant C_6/n, \tag{5.13}$$

which implies (5.8). One can verify

$$\lim_{t \to 0} g(t)/t < \infty, \tag{5.14}$$

$$\lim_{t \to 1} g(t)/(1-t) < \infty, \tag{5.15}$$

and therefore (5.9) follows from (5.8).

Now we are ready to proceed with the proof of Theorem 2.1.

Proof of Theorem 2.1. First we prove (2.6). Let $a_1 = n/\log n$. We see from (5.1) and (5.10) that

$$\max_{1 \leqslant k \leqslant a_1} |A_n(k)| = O_P((\log n)^{-1/2}), \tag{5.16}$$

$$\max_{a_1 \leqslant k \leqslant n-a_1} |A_n(k)| = O_P(1) \tag{5.17}$$

and

$$\max_{n-a_1 \leqslant k < n} |A_n(k)| = O_P((\log n)^{-1/2}). \tag{5.18}$$

Using (5.6) we obtain

$$\max_{a_1 \leqslant k \leqslant n-a_1} |V_n(k)| = O_P((\log\log\log n)^{1/2}), \tag{5.19}$$

which immediately implies

$$a(\log n) \max_{a_1 \leqslant k \leqslant n-a_1} |V_n(k)| - \sigma(x + b(\log n)) \xrightarrow{P} -\infty. \tag{5.20}$$

Similarly, by (5.17), (5.19), and (5.2) we have

$$a(\log n) \max_{a_1 \leqslant k \leqslant n-a_1} |R_n(k)| - \sigma(x + b(\log n)) \xrightarrow{P} -\infty. \tag{5.21}$$

From (5.16), (5.17), (5.18), (5.20), and (5.21) we obtain

$$\lim_{n\to\infty} P\left\{\frac{1}{\sigma} a(\log n) \max_{1\leqslant k<n} |V_n(k)-A_n(k)| \leqslant x+b(\log n)\right\}$$
$$= \lim_{n\to\infty} P\left\{\frac{1}{\sigma} a(\log n) \max(\max_{1\leqslant k\leqslant a_1} |V_n(k)-A_n(k)|, \right.$$
$$\left. \max_{n-a_1\leqslant k<n} |V_n(k)-A_n(k)|) \leqslant x+b(\log n)\right\}$$
$$= \lim_{n\to\infty} P\left\{\frac{1}{\sigma} a(\log n) \max(\max_{1\leqslant k\leqslant a_1} |V_n(k)|, \right.$$
$$\left. \max_{n-a_1\leqslant k<n} |V_n(k)|) \leqslant x+b(\log n)\right\}. \tag{5.22}$$

Using (5.20) and (5.4) we have

$$\lim_{n\to\infty} P\left\{\frac{1}{\sigma} a(\log n) \max(\max_{1\leqslant k\leqslant a_1} |V_n(k)|, \right.$$
$$\left. \max_{n-a_1\leqslant k<n} |V_n(k)|) \leqslant x+b(\log n)\right\}$$
$$= \lim_{n\to\infty} P\left\{\frac{1}{\sigma} a(\log n) \max_{1\leqslant k<n} |V_n(k)| \leqslant x+b(\log n)\right\}$$
$$=\exp(-2e^{-x}), \tag{5.23}$$

which completes the proof of (2.6).

Next we prove (2.8). Let $a_2 = m_1/\log m_1$. Arguing similarly to (5.16)–(5.21) we obtain

$$\max_{1\leqslant k\leqslant a_2} |A_n(k)| = O_P((\log m_1)^{-1/2}), \tag{5.24}$$
$$\max_{a_2\leqslant k\leqslant m_1} |A_n(k)| = O_P((m_1/n)^{1/2}) = o_P(1), \tag{5.25}$$
$$\max_{a_2\leqslant k\leqslant m_1} |V_n(k)| = O_P((\log\log\log m_1)^{1/2}). \tag{5.26}$$

Hence we have

$$a(\tfrac{1}{2}\log m_1) \max_{a_2\leqslant k\leqslant m_1} |R_n(k)| - \sigma(x+b(\tfrac{1}{2}\log m_1)) \xrightarrow{P} -\infty \tag{5.27}$$

and

$$a(\tfrac{1}{2}\log m_1) \max_{a_2 \leqslant k \leqslant m_1} |V_n(k)| - \sigma(x + b(\tfrac{1}{2}\log m_1)) \xrightarrow{P} -\infty. \tag{5.28}$$

We just proved that $\max_{1 \leqslant k \leqslant m_1} |R_n(k)|$, $\max_{1 \leqslant k \leqslant a_2} |R_n(k)|$, $\max_{1 \leqslant k \leqslant a_2} |V_n(k)|$, and $\max_{1 \leqslant k \leqslant m_1} |V_n(k)|$ must have the same limit distribution. Hence (2.8) follows from (5.4).

The proof of (2.10) is similar to that of (2.8) and therefore is omitted.

Now we show that (2.6), (2.8), and (2.10) imply (2.5), (2.7), and (2.9). Using (5.8) we find that $\max_{1 \leqslant k < n} |R_n(k)\, w_n(k)|$ and $\max_{1 \leqslant k < n} |R_n(k)(1 - g(k/n))^{-1/2}|$ must have the same limit. We apply (5.9), (2.8) and obtain

$$\max_{1 \leqslant k \leqslant a_1} |R_n(k)(1 - g(k/n))^{-1/2}| = O_P((\log\log n)^{1/2}/\log n) \tag{5.29}$$

and (5.9), (2.10) give

$$\max_{n - a_1 \leqslant k < n} |R_n(k)(1 - g(k/n))^{-1/2}| = O_P((\log\log n)^{1/2}/\log n). \tag{5.30}$$

Similarly to (5.21) we have

$$a(\log n) \max_{a_1 \leqslant k \leqslant n - a_1} |R_n(k)(1 - g(k/n))^{-1/2}| - \sigma(x + b(\log n)) \xrightarrow{P} -\infty \tag{5.31}$$

Hence (2.5) follows from (5.21) and (2.6).

By (5.8) it is enough to consider $\max_{1 \leqslant k \leqslant m_1} |R_n(k)(1 - g(k/n))^{-1/2}|$. Using (5.9) and (2.8) we have

$$\max_{1 \leqslant k \leqslant a_2} |R_n(k)((1 - g(k/n))^{-1/2} - 1)| = O_P\left(\frac{m_1}{n} (\log\log m_1)^{1/2}/\log m_1\right). \tag{5.32}$$

Similarly to (5.27) we have

$$a(\tfrac{1}{2}\log m_1) \max_{a_2 \leqslant k \leqslant m_1} |R_n(k)(1 - g(k/n))^{-1/2}| - \sigma(x + b(\tfrac{1}{2}\log m_1)) \xrightarrow{P} -\infty. \tag{5.33}$$

Hence (2.7) follows from (5.27), (5.33), (5.32), and (2.8).

The proof of (2.9) is very similar to that of (2.7) and hence is omitted.

Proof of Theorem 2.2. We show the asymptotic independence in (2.12) only. Using Lemma 5.3, (2.11) follows from (2.12). In the proof of Theorem 2.1 we showed that

$$\lim_{n\to\infty} P\left\{\frac{1}{\sigma} a(\log n)\, T_n(1, n) \leqslant x + b(\log n)\right\}$$
$$= \lim_{n\to\infty} P\left\{\frac{1}{\sigma} a(\log n) \max\left(\max_{1\leqslant k\leqslant a_1} |V_n(k)|, \max_{n-a_1\leqslant k<n} |V_n(k)|\right) \leqslant x + b(\log n)\right\}. \tag{5.34}$$

By the central limit theorem and the definition of a_1 we have

$$\max_{1\leqslant k\leqslant a_1}\left(\frac{n}{k(n-k)}\right)^{1/2}\frac{k}{n}\left|\sum_{1\leqslant i\leqslant n}\varepsilon_i\right| = O_P(1/(\log n)^{1/2}) \tag{5.35}$$

and

$$\max_{n-a_1\leqslant k<n}\left(\frac{n}{k(n-k)}\right)^{1/2}\frac{n-k}{n}\left|\sum_{1\leqslant i\leqslant n}\varepsilon_i\right| = O_P(1/(\log n)^{1/2}). \tag{5.36}$$

Putting together (5.34)–(5.36) we obtain

$$\lim_{n\to\infty} P\left\{\frac{1}{\sigma} a(\log n)\, T_n(1, n) \leqslant x + b(\log n)\right\}$$
$$= \lim_{n\to\infty} P\left\{\frac{1}{\sigma} a(\log n) \max\left(\max_{1\leqslant k\leqslant a_1}\left(\frac{n}{k(n-k)}\right)^{1/2}\left|\sum_{1\leqslant i\leqslant k}\varepsilon_i\right|,\right.\right.$$
$$\left.\left.\max_{n-a_1\leqslant k<n}\left(\frac{n}{n(n-k)}\right)^{1/2}\left|\sum_{k\leqslant i\leqslant n}\varepsilon_i\right|\right) \leqslant x + b(\log n)\right\}. \tag{5.37}$$

Let $\tilde{\boldsymbol\beta}_n^*$ denote the least-squares estimator of $\boldsymbol\beta^*$ based on $\{(y_i, \mathbf{x}_i), a_1 < i < n - a_1\}$. It is easy to check that under (1.2)

$$|\boldsymbol\beta_n^* - \tilde{\boldsymbol\beta}_n^*| = o_P(n^{-1/2}). \tag{5.38}$$

We define

$$\tilde{A}_n(k) = \left(\frac{n}{k(n-k)}\right)^{1/2} k(\tilde{\boldsymbol\beta}_n^* - \boldsymbol\beta^*)(\bar{\mathbf{x}}_k^* - \bar{\mathbf{x}}_n^*).$$

Lemma 5.3 and (5.38) imply

$$\sup_{nt_1\leqslant k\leqslant nt_2} |A_n(k) - \tilde{A}_n(k)| = o_P(1). \tag{5.39}$$

Next we introduce

$$\tilde{V}_n(k)=\left(\frac{n}{k(n-k)}\right)^{1/2}\left(\sum_{a_1<i\leqslant k}\varepsilon_i-\frac{k}{n}\sum_{a_1<i<n-a_1}\varepsilon_i\right),\qquad nt_1\leqslant k\leqslant nt_2.$$

The central limit theorem gives

$$\max_{nt_1\leqslant k\leqslant nt_2}|V_n(k)-\tilde{V}_n(k)|=O_P(1/(\log n)^{1/2}). \tag{5.40}$$

Thus we have

$$\max_{nt_1\leqslant k\leqslant nt_2}|R_n(k)-\tilde{R}_n(k)|=o_P(1), \tag{5.41}$$

where $\tilde{R}_n(k)=\tilde{V}_n(k)-\tilde{A}_n(k)$. It is clear that $\max_{nt_1\leqslant k\leqslant nt_2}|\tilde{R}_n(k)|$ and

$$\max\left(\max_{1\leqslant k\leqslant a_1}\left(\frac{n}{n(n-k)}\right)^{1/2}\left|\sum_{1\leqslant i\leqslant k}\varepsilon_i\right|,\max_{n-a_1\leqslant k<n}\left(\frac{n}{n(n-k)}\right)^{1/2}\left|\sum_{k\leqslant i\leqslant n}\varepsilon_i\right|\right)$$

are independent, and therefore (2.12) follows from (5.37) and (5.41).

Proof of Theorem 2.3. The estimator of β_2 is denoted by $\hat{\beta}_{2n}$. It is easy to see that

$$\lim_{n\to\infty}\hat{\beta}_{2n}=\beta^*=\gamma_2+(\beta_1-\gamma_1)\frac{\int_0^\tau f(t)\,dt-\tau\int_0^1 f(t)\,dt}{\int_0^1 f^2(t)\,dt-(\int_0^1 f(t)\,dt)^2}$$
$$+(\beta_2-\gamma_2)\frac{\int_0^\tau f^2(t)\,dt-\int_0^\tau f(t)\,dt\int_0^1 f(s)\,ds}{\int_0^1 f^2(t)\,dt-(\int_0^1 f(t)\,dt)^2}.$$

Thus we have

$$\lim_{n\to\infty}\frac{1}{n}E\sum_{1\leqslant i\leqslant n\lambda}(y_i-\bar{y}_n-\hat{\beta}_{n2}(x_i-\bar{x}_n))=\begin{cases}f_1(\lambda), & \text{if}\quad 0<\lambda\leqslant\tau\\ f_2(\lambda), & \text{if}\quad \tau\leqslant\lambda<1,\end{cases}$$

where

$$f_1(\lambda)=\lambda(1-\tau)(\beta_1-\gamma_1)+(\beta_2-\beta^*)\left\{\int_0^\lambda f(t)\,dt-\lambda\int_0^1 f(t)\,dt\right\}$$
$$+\lambda(\beta_2-\gamma_2)\int_\tau^1 f(t)\,dt$$

and

$$f_2(\lambda)=\tau(\beta_1-\gamma_1)+\lambda\gamma_1+\beta_2\int_0^\tau f(t)\,dt+\gamma_2\int_\tau^\lambda f(t)\,dt$$
$$-\lambda\left\{\tau\beta_1+\beta_2\int_0^\tau f(t)\,dt+(1-\tau)\gamma_1+\gamma_2\int_\tau^1 f(t)\,dt\right\}$$
$$-\beta^*\left\{\int_0^\lambda f(t)\,dt-\lambda\int_0^1 f(t)\,dt\right\}.$$

If $f_1(\lambda)=0$ for all $\lambda\in(0,\tau]$, then we must have $f_1'(\lambda)<0$, which implies $\beta_2=\beta^*$. Similarly, if $f_2(\lambda)=0$ for all $\lambda\in[\tau,1)$ we must have $\gamma_2=\beta^*$. Hence $\beta_2=\gamma_2=\beta^*$ and therefore $f_1(\lambda)=0$ implies $\beta_1=\gamma_1$. Thus we have

$$\lim_{n\to\infty}\frac{1}{n}E\sum_{1\leqslant i\leqslant n\lambda}(y_i-\bar{y}_n-\hat{\beta}_{n2}(x_i-\bar{x}_n))=0$$

for all $\lambda\in(0,1)$ if and only if $\beta_1=\gamma_1$ and $\beta_2=\gamma_2$, which completes the proof of (2.15).

It is clear that (2.15) implies (2.16).

The proofs of Theorem 3.1–3.3 require a generalization of Lemmas 5.1 and 5.3 for random design points.

LEMMA 5.4. *We assume that* (1.2), (3.1)–(3.6) *hold. Then, as* $n\to\infty$, *we have*

$$n^{1/2}(\hat{\beta}_{2n}-\beta_2)=O_P(1) \tag{5.42}$$

and

$$\max_{1\leqslant k<n}\left(\frac{nk}{n-k}\right)^{1/2}|\bar{x}_k-\bar{x}_n|=O_P((\log\log n)^{1/2}), \tag{5.43}$$

$$\max_{1\leqslant k<n}|w_n(k)-1|=O_P((\log\log n)/n). \tag{5.44}$$

Proof. It is well known that

$$\hat{\beta}_{2n}-\beta_2=\frac{n\sum_{1\leqslant i\leqslant n}x_i\varepsilon_i-\sum_{1\leqslant i\leqslant n}x_i\sum_{1\leqslant i\leqslant n}\varepsilon_i}{n\sum_{1\leqslant i\leqslant n}x_i^2-(\sum_{1\leqslant i\leqslant n}x_i)^2}. \tag{5.45}$$

Using (3.3) we have immediately

$$\frac{1}{n}\sum_{1\leqslant i\leqslant n}x_i^2-\left(\frac{1}{n}\sum_{1\leqslant i\leqslant n}x_i\right)^2=\operatorname{var}f(\xi)+O_P(n^{-1/2}). \tag{5.46}$$

Thus it is enough to show that

$$n^{-1/2} \sum_{1 \leqslant i \leqslant n} x_i \varepsilon_i = O_P(1) \tag{5.47}$$

and

$$\frac{1}{n} \sum_{1 \leqslant i \leqslant n} x_i n^{-1/2} \sum_{1 \leqslant i \leqslant n} \varepsilon_i = O_P(1). \tag{5.48}$$

The central limit theorem and the weak law of large numbers imply (5.47) and (5.48), which also completes the proof of (5.42).

We apply the central limit theorem and obtain

$$Q_n/n = \operatorname{var} f(\xi) + O_P(n^{-1/2}). \tag{5.49}$$

Let $z_i = x_i - Ex_i$. Thus we can write $\bar{x}_k - \bar{x}_n = (1/k)\{\sum_{1 \leqslant i \leqslant k} z_i - (k/n) \sum_{1 \leqslant i \leqslant n} z_i\}$, and therefore Lemma 5.2 implies (5.43).

Putting together (5.43) and (5.49) we obtain

$$\max_{1 \leqslant k < n} \frac{k(\bar{x}_k - \bar{x}_n)^2}{Q_n(1-k/n)} = O_P((\log \log n)/n),$$

which gives (5.44).

Proof of Theorem 3.1. We follow the proof of Theorem 2.1. Using (5.42) instead of (5.1) we can check that the contribution of $A_n(k)$ to the maximum of $R_n(k)$ is negligible, and therefore Lemma 5.2 implies (2.6), (2.8), and (2.10).

By (2.6) and (5.44) we have

$$\max_{1 \leqslant k < n} |R_n(k)(w_n(k) - 1)| = O_P((\log \log n)^{3/2}/n),$$

and therefore (2.6), (2.8), and (2.10) imply (2.5), (2.7), and (2.9).

Proof of Theorem 3.2. It is enough to prove (3.8) because by (5.44) it implies (3.7). As in the proof of Lemma 5.4, let $z_i = x_i - Ex_i$. Donsker's invariance principle yields

$$n^{-1/2}\left(\sum_{1 \leqslant i \leqslant nt} z_i - t \sum_{1 \leqslant i \leqslant n} z_i\right) \xrightarrow{\mathscr{D}[0,1]} (\operatorname{var} \xi)^{1/2} B(t),$$

where $\{B(t), 0 \leqslant t \leqslant 1\}$ is a Brownian bridge. Hence (5.42) gives

$$\max_{nt_1 \leqslant k \leqslant nt_2} |A_n(k)| = O_P(n^{-1/2}). \tag{5.50}$$

In the proof of Theorem 2.2 we showed that $\max_{1 \leqslant k < n} |V_k|$ and

$$T_n^* = \max\left(\max_{1 \leqslant k \leqslant a_1} \left(\frac{n}{k(n-k)}\right)^{1/2} \left| \sum_{1 \leqslant i \leqslant k} \varepsilon_i \right|, \max_{n-a_1 \leqslant k < n} \left(\frac{n}{k(n-k)}\right)^{1/2} \left| \sum_{k \leqslant i \leqslant n} \varepsilon_i \right| \right)$$

are asymptotically equivalent, where $a_1 = n/\log n$. By (5.40) and (5.50), $\max_{nt_1 \leqslant k \leqslant nt_2} |R_n(k)|$ and $\max_{nt_1 \leqslant k \leqslant nt_2} |\tilde{V}_n(k)|$ must have the same limit distribution. The random variables T_n^* and $\max_{nt_1 \leqslant k \leqslant nt_2} |\tilde{V}_n(k)|$ are independent for each n, which establishes the asymptotic independence in (3.8). The asymptotic distribution of $T_n(1, n)$ was computed in Theorem 3.1 and therefore it is enough to show that

$$\frac{1}{\sigma} \max_{nt_1 \leqslant k \leqslant nt_2} |\tilde{V}_n(k)| \xrightarrow{\mathscr{D}} \sup_{t_1 \leqslant t \leqslant t_2} |B(t)|/(t(1-t))^{1/2}. \tag{5.51}$$

Donsker's invariance principle gives

$$\frac{1}{\sigma} V_n(nt) \xrightarrow{\mathscr{D}[t_1, t_2]} B(t)/(t(1-t))^{1/2},$$

and therefore by (5.40) we have immediately (5.51).

Proof of Remark 3.1. It follows from the proof of Theorem 3.2. We showed that it is enough to consider T_n^* and $\max_{1 \leqslant k \leqslant n}((k(n-k))^{1/2}/n)\ |V_n(k)|$. Applying (5.40) we obtain

$$\max_{1 \leqslant k \leqslant n} \frac{(k(n-k))^{1/2}}{n} |V_n(k) - \tilde{V}_n(k)| = o_P(1),$$

which implies (3.10). Using (5.44) one can easily check that (3.10) yields (3.9).

The following lemma is needed in the proofs of Theorem 3.3 and 3.4. We compare $\hat{\beta}_{2n} - \beta_2$ of (5.45) and

$$\hat{\beta}_{2n} - \beta_2 = \frac{n \sum_{1 \leqslant i \leqslant n} \varepsilon_i \hat{x}_i - \sum_{1 \leqslant i \leqslant n} \hat{x}_i \sum_{1 \leqslant i \leqslant n} \varepsilon_i}{n \sum_{1 \leqslant i \leqslant n} \hat{x}_i^2 - (\sum_{1 \leqslant i \leqslant n} \hat{x}_i)^2}.$$

We also define

$$\hat{A}_n(k) = \left(\frac{n}{k(n-k)}\right)^{1/2} k(\hat{\beta}_{2n} - \beta_2) \left(\frac{1}{k} \sum_{1 \leqslant i \leqslant k} \hat{x}_i - \frac{1}{n} \sum_{1 \leqslant i \leqslant n} \hat{x}_i\right).$$

LEMMA 5.5. *We assume that* (1.2), (3.1), (3.3)–(3.5), *and* (3.11)–(3.14) *hold. Then, as* $n \to \infty$, *we have*

$$n^{1/2}(\beta_2 - \hat{\beta}_{2n}) = O_P(1), \tag{5.52}$$

$$\max_{1 \leqslant k < n} |A_n(k) - \hat{A}_n(k)| = O_P(n^{-1/2} \log n), \tag{5.53}$$

$$\max_{1 \leqslant k < n} |R_n(k) - \hat{R}_n(k)| = O_P(n^{-1/2} \log n) \tag{5.54}$$

and

$$\max_{1 \leqslant k < n} |w_n(k) - (1 - r(k/n))^{-1/2}| = O_P(n^{-1/2} \log n). \tag{5.55}$$

Proof. Komlós, Major, and Tusnády (1975, 1976) constructed a Wiener process $\{W(t), 0 \leqslant t < \infty\}$ such that

$$|\eta_k - \mu k - \gamma W(k)| \overset{\text{a.s.}}{=} O(\log k) \tag{5.56}$$

and

$$E\,|\eta_k - \mu k - \gamma W(k)|^{\tau} = O(\log k) \tag{5.57}$$

for all $\tau \geqslant 1$, where $\gamma^2 = \operatorname{var} \xi$. A two-term Taylor expansion and (3.11) give

$$|x_i - \hat{x}_i - f'(i\mu/n)(\eta_i - i\mu)/n| \leqslant C_7(\eta_i - i\mu)^2/n^2. \tag{5.58}$$

Thus we have

$$E\,|(x_i - \hat{x}_i)/i^{1/2}| = O(1/n). \tag{5.59}$$

Similar arguments give

$$E\,|(x_i^2 - \hat{x}_i^2)/i^{1/2}| = O(1/n). \tag{5.60}$$

It follows from (5.59) and (5.60) that

$$\frac{1}{n} \sum_{1 \leqslant i \leqslant n} \hat{x}_i^2 - \left(\frac{1}{n} \sum_{1 \leqslant i \leqslant n} \hat{x}_i\right)^2 - \left(\frac{1}{n} \sum_{1 \leqslant i \leqslant n} x_i^2 - \left(\frac{1}{n} \sum_{1 \leqslant i \leqslant n} x_i\right)^2\right) = O_P(n^{-1/2}). \tag{5.61}$$

Next we show

$$\left|\sum_{1 \leqslant i \leqslant n} \varepsilon_i(\hat{x}_i - x_i)\right| = O_P(1) \tag{5.62}$$

and

$$\sum_{1 \leqslant i \leqslant n} (\hat{x}_i^2 - x_i^2) \sum_{1 \leqslant i \leqslant n} \varepsilon_i = O_P(n). \tag{5.63}$$

We note that

$$E \sum_{1 \leqslant i \leqslant n} \varepsilon_i (\hat{x}_i - x_i) = 0, \tag{5.64}$$

and by (5.57) and (5.58) we have

$$\begin{aligned} E\left(\sum_{1 \leqslant i \leqslant n} \varepsilon_i(\hat{x}_i - x_i)\right)^2 &= \sigma^2 \sum_{1 \leqslant i \leqslant n} E(\hat{x}_i - x_i)^2 \\ &= O\left(\sum_{1 \leqslant i \leqslant n} (f'(i\mu/n))^2 \frac{i}{n^2} + \sum_{1 \leqslant i \leqslant n} \frac{i^2}{n^4}\right) \\ &= O(1). \end{aligned} \tag{5.65}$$

Now (5.62) follows from (5.64) and (5.65). The central limit theorem implies that

$$n^{-1/2} \sum_{1 \leqslant i \leqslant n} \varepsilon_i = O_P(1). \tag{5.66}$$

By (5.58) we have that $n^{-1/2} \sum_{1 \leqslant i \leqslant n} (\hat{x}_i^2 - x_i^2)$ is asymptotically normal, which completes the proof of (5.63). By (5.61)–(5.63) we have immediately (5.52).

Next we use (5.58) and obtain

$$\begin{aligned} &\left|\bar{x}_k - \bar{x}_n - \left(\frac{1}{k}\sum_{1 \leqslant i \leqslant k} \hat{x}_i - \frac{1}{n}\sum_{1 \leqslant i \leqslant n} \hat{x}_i\right) - \left\{\frac{1}{nk}\sum_{1 \leqslant i \leqslant k} f'\left(\frac{i\mu}{n}\right)(\eta_i - i\mu)\right.\right. \\ &\qquad \left.\left. - \frac{1}{n^2}\sum_{1 \leqslant i \leqslant n} f'\left(\frac{i\mu}{n}\right)(\eta_i - i\mu)\right\}\right| \leqslant C_7 \max_{1 \leqslant i \leqslant n} |\eta_i - i\mu|/n^2 \\ &\quad = O_P(n^{-3/2}). \end{aligned} \tag{5.67}$$

We apply (5.56) and obtain

$$\max_{1 \leqslant k \leqslant n} \left|\frac{1}{nk} \sum_{1 \leqslant i \leqslant k} f'\left(\frac{i\mu}{n}\right)(\eta_i - i\mu - \gamma W(i))\right| \overset{\text{a.s.}}{=} O((\log n)/n). \tag{5.68}$$

The scale transformation of the Wiener process gives

$$\left\{\frac{\gamma}{nk}\sum_{1\leqslant i\leqslant k} f'\left(\frac{i\mu}{n}\right) W(i),\ 1\leqslant k\leqslant n\right\}$$
$$\overset{\mathscr{D}}{=}\left\{\frac{\gamma}{kn^{1/2}}\sum_{1\leqslant i\leqslant k} f'\left(\frac{i\mu}{n}\right) W\left(\frac{i}{n}\right),\ 1\leqslant k\leqslant n\right\}. \tag{5.69}$$

By Garsia (1970) and (3.12) we have

$$\max_{1\leqslant k\leqslant n}\left|\frac{1}{kn^{1/2}}\sum_{1\leqslant i\leqslant k} f'\left(\frac{i\mu}{n}\right) W\left(\frac{i}{n}\right)-\frac{n^{1/2}}{k}\int_0^{k/n} f'(\mu t)\, W(t)\, dt\right|$$
$$= O_P((\log n)/n). \tag{5.70}$$

It is easy to see that

$$\int_0^1 f'(t)\, W(t)\, dt = O_P(1). \tag{5.71}$$

Integration by parts gives

$$\int_0^u f'(\mu t)\, W(t)\, dt = \frac{1}{\mu}\int_0^u W(t)\, df(\mu t)$$
$$= \frac{1}{\mu} f(\mu u)\, W(u) - \frac{1}{\mu}\int_0^u f(\mu t)\, dW(t),$$

and therefore Darling and Erdős (1956) imply

$$\max_{1\leqslant k<n}\left|\left(\frac{n}{k}\right)^{1/2}\int_0^{k/n} f'(t\mu)\, W(t)\, dt\right| = O_P((\log\log n)^{1/2}). \tag{5.72}$$

Hence we have

$$\max_{1\leqslant k\leqslant n/2} |A_n(k)-\hat{A}_n(k)|$$
$$\leqslant \max_{1\leqslant k\leqslant n/2}\left|\left(\frac{n}{k(n-k)}\right)^{1/2} k(\hat{\beta}_{2n}-\beta_2)\right.$$
$$\left.\times\left\{\bar{x}_k-\bar{x}_n-\left(\frac{1}{k}\sum_{1\leqslant i\leqslant k}\hat{x}_i-\frac{1}{n}\sum_{1\leqslant i\leqslant n}\hat{x}_i\right)\right\}\right|$$
$$+\max_{1\leqslant k\leqslant n/2}\left|\left(\frac{n}{k(n-k)}\right)^{1/2} k(\beta_{2n}-\hat{\beta}_{2n})(\bar{x}_k-\bar{x}_n)\right|$$
$$= D_{1n}+D_{2n}. \tag{5.73}$$

Using (5.1), (5.52), and (5.67)–(5.72) we have

$$D_{1n} = O_P(n^{-1/2}(\log\log n)^{1/2}). \tag{5.74}$$

Similarly, (5.52) and (5.10) give

$$D_{2n} = O_P(n^{-1/2}). \tag{5.75}$$

Observing that

$$n^{1/2}\left\{\frac{1}{k}\int_0^{k/n} f'(\mu t)\, W(t)\, dt - \frac{1}{n}\int_0^1 f'(\mu t)\, W(t)\, dt\right\}$$
$$= n^{1/2}\left\{\left(\frac{1}{k}-\frac{1}{n}\right)\int_0^{k/n} f'(\mu t)\, W(t)\, dt - \frac{1}{n}\int_{k/n}^1 f'(\mu t)\, W(t)\right\},$$

Erdős and Darling (1956) imply

$$\max_{n/2 \leqslant k < n} n^{1/2}\left|\frac{1}{k}\int_0^{k/n} f'(\mu t)\, W(t)\, dt - \frac{1}{n}\int_0^1 f'(\mu t)\, W(t)\, dt\right| \Big/ ((n-k)^{1/2}/n)$$
$$= O_P((\log\log n)^{1/2}). \tag{5.76}$$

Thus similarly to (5.74) and (5.75) we have

$$\max_{n/2 \leqslant k < n} |A_n(k) - \hat{A}_n(k)| = O_P(n^{-1/2}\log n),$$

which completes the proof of (5.53).

By (5.2), (5.53) implies (5.54).

The proof of (5.55) is similar to that of (5.53) and therefore it is omitted.

Proof of Theorems 3.3 *and* 3.4. It follows from Lemma 5.5 that $Z_n(i,j)$ ($T_n(i,j)$) and $\hat{Z}_n(i,j)$ ($\hat{T}_n(i,j)$) must have the same limit distribution. Hence these results follow from Theorems 2.1 and 2.2.

Acknowledgments

The authors are grateful to Professor R. D. Cook for providing the Old Faithful data and to a referee for the careful reading of our paper and for useful remarks.

References

[1] Andrews, D. F., and Herzberg, A. M. (1985). *Data.* Springer-Verlag, New York.

[2] Brown, R. L., Durbin, J., and Evans, J. M. (1975). Techniques for testing the constancy of regression relationships over time (with discussions). *J. Roy. Statist. Soc. Ser. B* **37** 149–192.

[3] Cook, R. D., and Weisberg, S. (1982). *Residuals and Influence in Regression*. Chapman & Hall, New York.

[4] Csörgő, and Horváth, L. (1988). Nonparametric methods for changepoint problems. In *Handbook of Statistics*, Vol. 7 (P. R. Krishnaiah and C. R Rao, Eds.), pp. 403–425. North-Holland, Amsterdam.

[5] Darling, D. A., and Erdős, P. (1956). A limit theorem for a maximum of normalized sums of independent random variables. *Duke Math. J.* **23** 143–155.

[6] Davison, A. C., and Snell, E. J. (1991). Residuals and diagnostics. In *Statistics, Theory and Modelling* (D. V. Hinkley, N. Reid, and E. J. Snell, Eds.), pp. 83–106. Chapman & Hall, New York.

[7] De Long, D. M. (1981). Crossing probabilities for a square root boundary by a Bessel process. *Comm. Statist. Theory Methods Ser. A* **10** 2197–2213.

[8] Garsia, A. M. (1970) Continuity properties of Gaussian processes with multidimensional time parameter. In *Proceedings 6th Berkeley Sympos. Math. Statist. Probab.*, Vol. 2, pp. 369–374. University of Carolina, Berkeley.

[9] Hawkins, D. L. (1989). A *U–I* approach to retrospective testing for shifting parameters in a linear model. *Comm. Statist. Theory Methods Ser. A* **18** 3117–3134.

[10] Hušková, M. (1993). Nonrecursive procedures for detecting changes in simple linear regression. In *Proceedings of IMSIBAC 4*, to appear.

[11] James, B., James, K. L., and Siegmund, D. (1987). Tests for a change-point. *Biometrika* **74** 71–84.

[12] Kim, H. J., and Siegmund, D. (1989). The likelihood ratio test for a change-point in simple linear regression. *Biometrika* **76** 409–423.

[13] Komlós, J., Major, P., and Tusnády, G. (1975). An approximation of partial sums of independent R. V.'s and the sample DF, I. *Z. Wahrsch. verw. Gebiete* **32** 111–131.

[14] Komlós, J., Major, P., and Tusnády, G. (1976). An approximation of partial sums of independent R. V.'s and the sample of DF, II. *Z. Wahrsch. verw. Gebiete* **34** 33–58.

[15] Maddala, G. S. (1977). *Econometrics*. McGraw–Hill, New York.

[16] Maronna, R., and Yohai, V. J. (1978). A bivariate test for the detection of a systematic change in mean. *J. Amer. Statist. Assoc.* **73** 640–645.

[17] Moore, D. S., and Mccabe, G. P. (1989). *Introduction to the Practice of Statistics*. Freeman, New York.

[18] Quandt, R. E. (1958). The estimation of the parameter of a linear regression system obeying two separate regimes. *J. Amer. Statist. Assoc.* **53** 873–880.

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

[20] Worsley, K. J. (1986). Testing a two-phase multiple regression. *Technometrics* **26** 35–42.

[21] Yao, Y.-C., and Davis, R. A. (1986). The asymptotic behavior of the likeliood ratio statistic for testing a shift in mean in a sequence of independent normal variates. *Sankhyā Ser. A* **48** 339–353.