

## THE THEORY OF FUNCTIONAL LEAST SQUARES

SÁNDOR CSÖRGÓ

(Received 19 August 1981; revised 1 May 1982)

Communicated by R. L. Tweedie

### Abstract

The functional least squares procedure of Chambers and Heathcote for estimating the slope parameter in a linear regression model is analysed. Strong uniform consistency for the family of these estimators is proved together with a necessary and sufficient condition for weak convergence in the space of continuous vector valued functions. These results are then used to develop the asymptotic normality of an adaptive version of the functional least squares estimator with minimum limiting variance.

1980 *Mathematics subject classification* (*Amer. Math. Soc.*): primary 62 J 05, 62 F 35; secondary 60 F 17.

### 1. Introduction

Let  $p \geq 1$  be an integer and consider the linear regression model

$$Y_j = \langle x_j, \beta \rangle + \varepsilon_j, \quad j = 1, \dots, n,$$

where the  $x_j = (x_{j1}, \dots, x_{jp})$ ,  $j = 1, 2, \dots$ , are known deterministic vectors,  $\beta = (\beta_1, \dots, \beta_p)$  is the unknown slope vector to be estimated from the observations  $Y_1, \dots, Y_n$ , and where the errors  $\varepsilon_1, \varepsilon_2, \dots$  are independent identically distributed real random variables with common distribution function  $F(x) = \Pr\{\varepsilon_j \leq x\}$ ,  $x \in \mathbf{R}$ , and characteristic function

$$(1.1) \quad C(t) = \int_{-\infty}^{\infty} e^{itx} dF(x).$$

---

Work done while the author was a Visiting Fellow of the Australian National University at the Department of Statistics, The Faculties. He expresses his deep gratitude to Professor C. R. Heathcote for his invitation.

© Copyright Australian Mathematical Society 1983

Throughout  $\langle \cdot, \cdot \rangle$  will denote inner product. Any intercept term is absorbed into the errors. It is assumed that there is a unique slope vector  $\beta_0$  to be estimated, that is, the sequence  $\{\epsilon_j\}$  of errors can in fact be written as  $\{Y_j - \langle x_j, \beta_0 \rangle\}$ . The usual least squares estimator  $\hat{\beta}_n$  of  $\beta_0$ , which dates back to Gauss, is the statistic minimising the sample version

$$\frac{1}{n} \sum_{j=1}^n (Y_j - \bar{Y} - \langle x_j - \bar{x}, \beta \rangle)^2$$

of the error variance  $\sigma_\epsilon^2 = \int_{-\infty}^{\infty} (x - E\epsilon_j)^2 dF(x)$ . It is well-known that, under some restrictions on the vectors  $x_j, j = 1, 2, \dots$ ,  $\hat{\beta}_n$  is a consistent estimator for  $\beta_0$  and  $n^{1/2}(\hat{\beta}_n - \beta_0)$  is asymptotically normally distributed provided  $\sigma_\epsilon^2 < \infty$ . On the other hand, several authors have reported non-robust properties of  $\beta_n$  if the distribution of the  $\epsilon_j$  deviates from the normal distribution.

Chambers and Heathcote (1981) observed that when the errors are normally distributed, then least squares estimation is equivalent to minimising the empirical counterpart of the function

$$L(t) = -\frac{1}{t^2} \log|C(t)|^2,$$

since in the normal case this is simply  $\sigma_\epsilon^2$  for all  $t$ . If  $\epsilon$  is not normal but  $\sigma_\epsilon^2 < \infty$ , then least square estimation is equivalent to minimising the sample version of the constant term in the power series expansion of  $L(t)$ . In this case  $L(t) = \sigma_\epsilon^2 + o(t)$  as  $t \rightarrow 0$ . Let us introduce the function

$$(1.2) \quad L(\beta; t) = -\frac{1}{t^2} \log|C(t)\phi(t(\beta_0 - \beta))|^2,$$

where  $\phi(s), s \in \mathbb{R}^p$ , is the limiting average characteristic function of the nonstochastic vectors  $x_1, \dots, x_n$ , assumed to exist by condition (2.i) below. Note that  $L(\beta_0; t) = L(t)$ . Based upon the above observations, Chambers and Heathcote (1981) proposed the fascinating idea of constructing the *functional* least squares estimator  $\beta_n^*(t)$  of  $\beta_0$  defined, for each  $t \neq 0$  in some interval around the origin, to be the random variable minimising

$$(1.3) \quad L_n(\beta; t) = -\frac{1}{t^2} \log \left| \frac{1}{n} \sum_{j=1}^n e^{it(Y_j - \langle x_j, \beta \rangle)} \right|^2,$$

the sample version of  $L(\beta; t)$ . Upon taking partial derivatives with respect to the components of  $\beta$ , they arrived at the estimating equations

$$(1.4) \quad l_n(\beta; t) = (0, \dots, 0)$$

where

$$(1.5) \quad l_n(\beta; t) = (l_{n1}(\beta; t), \dots, l_{np}(\beta; t))$$

with

(1.6)

$$l_{nk}(\beta; t) = \frac{1}{n^2} \frac{1}{t} \sum_{r=1}^n \sum_{j=1}^n x_{jk} \sin(t[Y_j - Y_r - \langle x_j - x_r, \beta \rangle]), \quad k = 1, \dots, p.$$

They also noted that in the limit  $t \rightarrow 0$  the estimating equations (1.4) become the normal equations of least squares estimation, and therefore  $\beta_n^*(0)$  can always be interpreted as  $\hat{\beta}_n$ . Under conditions (2.i)–(2.vi) of the next section they proved that at each point  $t$  the estimator  $\beta_n^*(t)$  is strongly consistent for  $\beta_0$ , and that  $n^{1/2}(\beta_n^*(t) - \beta_0)$  is asymptotically normally distributed with mean zero, covariance matrix  $\sigma^2(t)A^{-1}$ , where the  $p \times p$  symmetric matrix  $A^{-1}$  is the inverse of the matrix  $A$  of condition (2.iii) below and the scalar “variance function”  $\sigma^2(t)$  is that of (6.1) below. Chambers and Heathcote (1981) computed  $\sigma^2(t)$  for a number of error distributions and linked an integrated, or averaged version of the estimator to robust estimation theory. Heathcote (1982) further illustrated functional least squares estimation showing that it may be viewed as perturbed ordinary least squares estimation, treating efficiency problems and connecting it to the theory of angular random variables by some illuminating comments. He also constructed confidence bands for the variance function  $\sigma^2(t) = t^{-2} \sinh \sigma_\varepsilon^2 t^2$  of normally distributed errors.

Now if the reasonable criterion of minimum asymptotic variance is accepted for differentiating among possible estimates of  $\beta_0$ , then  $\hat{\beta}_n$  is appropriate only if  $\sigma^2(t)$  has a global minimum at zero. But Chambers and Heathcote (1981) have shown that in the class of distributions which lack a variance, or are normal, or are leptokurtic, the normal is the only one for which this is the case. (It would be of interest to know whether this holds in the class of all distributions.) If  $\sigma^2(t)$  achieves its global minimum elsewhere, say at  $t_0 \neq 0$ , then  $\beta_n^*(t_0)$  would be the ideal estimator. But in practice  $\sigma^2(t)$ , determined completely by the error characteristic function  $C(t)$  of (1.1), is not known. Chambers and Heathcote proposed an estimator  $\sigma_n^2(t)$  (see (6.3) below) for  $\sigma^2(t)$ , and reasoned heuristically that the random variable  $t_n$  minimising  $\sigma_n^2(t)$  is the second best choice, that is, the adaptive estimator  $\beta_n^*(t_n)$  is to be used for estimating  $\beta_0$ .

One of the main goals of this paper is to justify the above reasoning, or, rather, to see the extent to which it can be justified and beyond which it cannot be justified. This is expressed by condition (3.i) below. This condition reflects the tail behaviour of the distribution of the errors and, fortunately, is satisfied in all practical situations. For example, it is satisfied by errors with any stable distribution. The main result, presented in Section 7, lies somewhat deeper and requires for its proof (almost sure) uniform consistency of  $\beta_n^*(\cdot)$  and a weak convergence theorem for the sequence of stochastic processes  $n^{1/2}(\beta_n^*(\cdot) - \beta_0)$ , the condition

for the latter being necessary and sufficient. These results are perhaps of interest in their own right, and are presented in Sections 4 and 5, respectively, after listing the conditions for the  $x_j$  sequence in Section 2 and some preliminaries from Csörgő (1981a, b) and Marcus (1981) in Section 3. Strong uniform consistency of  $\sigma_n^2(\cdot)$  is developed in Section 6, together with the strong consistency of  $t_n$ . Some concluding remarks are found in Section 8.

### 2. Assumptions

We shall use the following conditions for the  $x_j = (x_{j1}, \dots, x_{jp})$  vectors:

(2.i) The limit  $\phi(s) = \lim_{n \rightarrow \infty} \phi_n(s)$  exists for all  $s \in \mathbf{R}^p$  for the functions

$$\phi_n(s) = \frac{1}{n} \sum_{j=1}^n e^{i\langle x_j, s \rangle}$$

and the function  $\phi(s)$  is continuous at  $s = (0, \dots, 0)$ .

(2.ii) For any  $k$  and  $m$ ,  $1 \leq k, m \leq p$ , the following limits exist:

$$\bar{x}_{km} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n x_{jk} x_{jm}, \quad \bar{x}_k = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n x_{jk}.$$

(2.iii) The  $p \times p$  matrix  $A = (a_{km})$  with elements  $a_{km} = \bar{x}_{km} - \bar{x}_k \bar{x}_m$  is nonsingular.

(2.iv) For any  $k$  and  $m$ ,  $1 \leq k, m \leq p$ , we have

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{j=1}^n x_{jk}^2 x_{jm}^2 &= 0, & \lim_{n \rightarrow \infty} \frac{\max_{1 \leq j \leq n} x_{jk}^2 x_{jm}^2}{\sum_{j=1}^n x_{jk}^2 x_{jm}^2} &= 0, \\ \lim_{n \rightarrow \infty} \frac{\max_{1 \leq j \leq n} (x_{jk} - \bar{x}_k)^2}{\sum_{j=1}^n (x_{jk} - \bar{x}_k)^2} &= 0, & \lim_{n \rightarrow \infty} \frac{\max_{1 \leq j \leq n} x_{jk}^2}{\sum_{j=1}^n x_{jk}^2} &= 0. \end{aligned}$$

Throughout we will restrict  $t$  to a compact set  $S \subset \mathbf{R}$  not containing zero, for example  $S$  can be  $[-b, -a] \cup [a, b]$  with  $0 < a < b$ . Also we restrict  $\beta$  to a compact set  $K \subset \mathbf{R}^p$  with a non-empty interior, and thus consider the functions  $L(\beta; t)$  and  $L_n(\beta; t)$  of (1.2) and (1.3) on the compact set  $K \times S \subset \mathbf{R}^{p+1}$ . Hence we define  $\beta_n^*(t)$  by the equation

$$(2.1) \quad L_n(\beta_n^*(t); t) = \inf_{\beta \in K} L_n(\beta; t), \quad t \in S.$$

As Chambers and Heathcote (1981) pointed out,  $\beta_n^*(t)$  is not necessarily uniquely defined, but any of the many possible ways of making it unique would work. We shall also use the following two further conditions:

(2.v)  $C(t) \neq 0$  for any  $t \in S$ .

(2.vi)  $\beta_0$  is an interior point of the compact set  $K$ , and  $0 < |\phi(s)| < 1 = \phi(0, \dots, 0)$  for any  $s \neq (0, \dots, 0)$  in the set

$$(2.2) \quad R = \{s = t(\beta_0 - \beta) : \beta \in K, t \in S\}.$$

Apart from that they did not specify compact sets, the above conditions are the same as those required by Chambers and Heathcote (1981), with the extra continuity of  $\phi$  at the origin in condition (2.i) and the extra requirement of the second and fourth relations in condition (2.iv). Formally they do not require the third relation of (2.iv), but they implicitly make use of it. It is, in fact, necessary. This third relation may well imply the fourth for nonzero  $\bar{x}_k$ 's as well, but I cannot prove it. Since

$$L(\beta; t) = -\frac{1}{t^2} \log|C(t)|^2 - \frac{1}{t^2} \log|\phi(t(\beta_0 - \beta))|^2,$$

condition (2.vi) is equivalent to the condition

$$(2.3) \quad L(\beta_0; t) < L(\beta; t), \quad \beta \in K, \beta \neq \beta_0, t \in S.$$

For  $n$  large enough  $\beta_n^*(t)$  is a solution of (1.4) almost surely. Applying the implicit function theorem (Rudin (1976), page 224) we see that  $\beta_n^*(t)$  is unique and is continuously differentiable (for  $n$  large enough, almost surely) on a suitably chosen  $S$ . Of course we assume that our  $S$  above is such a choice. In particular,  $\beta_n^*(\cdot)$ , for  $n$  large enough, can be considered a random element of the separable Banach space  $\mathcal{C}^p(S)$  of continuous  $p$ -dimensional vector valued functions  $f(t) = (f_1(t), \dots, f_p(t))$  endowed with the supremum norm  $\sup\{|f(t)| : t \in S\}$ . The Euclidean norm of a vector  $s = (s_1, \dots, s_p) \in \mathbf{R}^p$  will be denoted by  $|s|$ , while its maximum norm  $\max(|s_1|, \dots, |s_p|)$  by  $\|s\|$ . Of course, the space  $\mathcal{C}^2(S)$  will be identified with the corresponding space of continuous complex valued functions on  $S$ .

### 3. Preliminaries

Consider the empirical characteristic function

$$C_n(t) = \frac{1}{n} \sum_{j=1}^n e^{it\epsilon_j}$$

of the first  $n$  error variables. As a consequence of the Glivenko-Cantelli theorem and the continuity theorem of characteristic functions we have

$$(3.1) \quad \sup_{t \in S} |C_n(t) - C(t)| \xrightarrow{\text{a.s.}} 0,$$

where  $\xrightarrow{a.s.}$  denotes almost sure convergence. The symbol  $\xrightarrow{Pr}$  will mean convergence in probability, and all convergence relations are understood as  $n \rightarrow \infty$  if not specified otherwise. Consider also the empirical characteristic process

$$D_n(t) = n^{1/2}(C_n(t) - C(t)) = n^{-1/2} \sum_{j=1}^n (e^{ite_j} - C(t))$$

and let  $D(t)$  be a univariate complex Gaussian process with zero mean and covariance

$$ED(t) \overline{D(s)} = C(t - s) - C(t)C(-s).$$

As shown in Csörgő (1981a), this can be represented as the stochastic integral

$$(3.2) \quad D(t) = \int_{-\infty}^{\infty} e^{itx} dB(F(x)),$$

where  $B$  is a Brownian bridge process in  $[0, 1]$ . Now with  $U(t) = \text{Re } C(t)$  and  $\psi(t) = (1 - U(t))^{1/2}$  set

$$m(y) = \lambda\{t \in (-\frac{1}{2}, \frac{1}{2}) : \psi(t) < y\}$$

where  $\lambda$  is the Lebesgue measure. Making use of the Dudley-Fernique necessary and sufficient condition for the sample continuity of a stationary Gaussian process, Csörgő (1981a) showed that  $D$  is sample continuous if and only if the condition

$$(3.i) \quad \int_0^1 \frac{\bar{\psi}(h)}{h(\log 1/h)^{1/2}} dh < \infty$$

holds, where  $\bar{\psi}$  is the nondecreasing rearrangement of  $\psi$  defined as

$$\bar{\psi}(h) = \sup\{y : m(y) < h\}.$$

Therefore (3.i) is a necessary condition of the weak convergence of  $D_n(\cdot)$  in  $\mathcal{C}^2(S)$  to  $D(\cdot)$ . To see what this condition is all about we note that it is generally not satisfied if

$$\int_{-\infty}^{\infty} \log^+(|x|) dF(x) < \infty,$$

but it is satisfied if for a positive  $\delta$  we have

$$(3.3) \quad \int_{-\infty}^{\infty} (\log^+(|x|))^{1+\delta} dF(x) < \infty.$$

A finer analysis is found in Csörgő (1981a), who also proved that  $Pr\{\{\epsilon_j\} > x\} = O(x^{-\alpha})$ , as  $x \rightarrow \infty$ , is sufficient for weak convergence with any small positive  $\alpha$ . (In fact, strong approximation results were proved with specified rates depending on  $\alpha$ .) As far as weak convergence goes Marcus (1981) filled the gap between the latter condition and condition (3.i) by giving two nice proofs for the sufficiency part of the following result.

**THEOREM A.**  $D_n(\cdot)$  converges weakly in  $\mathcal{C}^2(S)$  to  $D(\cdot)$  if and only if condition (3.i) holds.

The proof of the weak convergence theorem for  $\beta_n^*$  in Section 5 will require the following slight extension of this result.

**THEOREM B.** Let  $b_{nj}, j = 1, \dots, n; n = 1, 2, \dots,$  be a triangular array of real numbers such that  $(1/n)\sum_{j=1}^n b_{nj}^2 = 1$  for any  $n$ , and  $(1/n)\max_{1 \leq j \leq n} b_{nj}^2 \rightarrow 0$ . Then the process

$$D_n^b(t) = n^{-1/2} \sum_{j=1}^n b_{nj}(e^{it\epsilon_j} - C(t))$$

converges weakly in  $\mathcal{C}^2(S)$  to  $D(\cdot)$  if and only if condition (3.i) holds.

**PROOF.** The covariance structure of  $D_n^b$ , for any  $n$ , is the same as that of  $D(\cdot)$ . Therefore, according to the obvious multidimensional analogue of Liapunov’s central limit theorem for row-wise independent triangular arrays of random variables (see Chung (1974), page 201, Corollary), the finite-dimensional distributions of  $D_n^b$  always converge to those of  $D$ . Hence (3.i) is necessary for weak convergence in  $\mathcal{C}^2(S)$ . Concerning sufficiency, Marcus’ longer first proof evidently goes through. In fact his much simpler second proof also goes through with trivial modifications (as in the case of the first one) if one notes that the starting point, namely Fernique’s extension of a theorem of Pisier (Fernique (1978), Théorème 1.3) also holds true for independent but non-identically distributed summands.

This proof also indicates that the multivariate extension of Theorem B is just as straightforward as that of Theorem A in Csörgő (1981b). Of course,  $S$  in (3.1) and Theorems A and B above could have been any compact set from the line, containing also the origin. Finally we note here that an annotated bibliography of the empirical characteristic function is compiled in Csörgő (1980).

#### 4. Strong uniform consistency of $\beta_n^*$

We begin with a lemma which will be useful in Section 6 as well.

**LEMMA 1.** Under condition (2.i) we have

$$\Delta_n = \sup_{(\beta, t) \in K \times S} \left| \frac{1}{n} \sum_{j=1}^n e^{it(\epsilon_j + \langle x_j, \beta_0 - \beta \rangle)} - C(t)\phi(t(\beta_0 - \beta)) \right| \xrightarrow{\text{a.s.}} 0.$$

PROOF. The claim is equivalent to the following

$$(4.1) \quad \Delta_n = \sup_{(s, t) \in R \times S} \left| \frac{1}{n} \sum_{j=1}^n e^{i(t\varepsilon_j + \langle x_j, s \rangle)} - C(t)\phi(s) \right| \xrightarrow{\text{a.s.}} 0,$$

where the set  $R$  is of (2.2). Let  $G_j(y) = G_j(y_1, \dots, y_p)$  denote the distribution function of the deterministic vector  $x_j$ , that is,  $G_j(y) = \mathbf{1}_{R(j)}(y)$ ,  $y \in \mathbf{R}^p$ , where  $R(j)$  is the infinite rectangle  $R(j) = \{s \in \mathbf{R}^p: x_j < s\}$ . Clearly,  $\phi_n(s)$  of condition (2.i) is the characteristic function of the probability distribution function

$$\tilde{F}_n(y) = \frac{1}{n} \sum_{j=1}^n G_j(y), \quad y \in \mathbf{R}^p.$$

Applying the  $p$ -variate continuity theorem, it follows, on the one hand, from condition (2.i) that  $\phi(s)$  is a proper  $p$ -variate characteristic function. On the other hand, it follows from (2.i) that

$$(4.2) \quad \tilde{F}_n(y) \rightarrow G(y)$$

at every continuity point  $y \in \mathbf{R}^p$  of  $G$ , where  $G(y)$  is the (unique) distribution function belonging to  $\phi(s)$ . It is plain (Fubini theorem) that

$$(4.3) \quad \phi(s)C(t) = \int_{\mathbf{R}^{p+1}} e^{i\langle (y_1, \dots, y_{p+1}), (s, t) \rangle} dG(y_1, \dots, y_p)F(y_{p+1}),$$

$$(s, t) \in \mathbf{R}^{p+1},$$

that is, the product  $\phi(s)C(t)$  is the characteristic function of the  $(p + 1)$ -variate distribution function  $G(y_1, \dots, y_p)F(y_{p+1})$ .

Now consider the sequence of independent (but not identically distributed)  $(p + 1)$ -dimensional random vectors  $(x_j, \varepsilon_j)$  with their distribution functions

$$F_j(y_1, \dots, y_{p+1}) = \Pr\{(x_j, \varepsilon_j) < (y_1, \dots, y_{p+1})\}$$

$$= G_j(y_1, \dots, y_p)F(y_{p+1}),$$

and let  $\hat{F}_n(y) = \hat{F}_n(y_1, \dots, y_{p+1})$  be the empirical distribution function of  $(x_1, \varepsilon_1), \dots, (x_n, \varepsilon_n)$ . Clearly

$$\hat{F}_n(y_1, \dots, y_{p+1}) = \frac{1}{n} \sum_{j=1}^n G_j(y_1, \dots, y_p) \mathbf{1}_{(\varepsilon_j, \infty)}(y_{p+1})$$

and the averages in (4.1) are the characteristic functions of the  $\hat{F}_n$ . Since characteristic functions always converge uniformly on compact sets if they converge pointwise on the whole space and since  $R \times S$  is a compact set, (4.1) will therefore follow from the  $(p + 1)$ -variate continuity theorem if we can show that

$$\hat{F}_n(y_1, \dots, y_{p+1}) \xrightarrow{\text{a.s.}} G(y_1, \dots, y_p)F(y_{p+1})$$

at every continuity point  $(y_1, \dots, y_{p+1})$  of  $GF$ . But

$$\begin{aligned} & \left| \hat{F}_n(y_1, \dots, y_{p+1}) - G(y_1, \dots, y_p)F(y_{p+1}) \right| \\ & \leq \left| \hat{F}_n(y_1, \dots, y_{p+1}) - \tilde{F}_n(y_1, \dots, y_p)F(y_{p+1}) \right| \\ & \quad + \left| \tilde{F}_n(y_1, \dots, y_p)F(y_{p+1}) - G(y_1, \dots, y_p)F(y_{p+1}) \right|, \end{aligned}$$

where the second term goes to zero by (4.2) in every continuity point of  $G$  and the first term is

$$\left| \frac{1}{n} \sum_{j=1}^n G_j(y_1, \dots, y_p) \{ \mathbf{1}_{(\epsilon_j, \infty)}(y_{p+1}) - F(y_{p+1}) \} \right| \xrightarrow{\text{a.s.}} 0$$

in every point  $y \in \mathbf{R}^{p+1}$  by Kolmogorov’s strong law of large numbers for independent non-identically distributed random variables. Hence the lemma is proved.

LEMMA 2. Under conditions (2.i), (2.v) and (2.vi) we have

$$M_n^{(1)} = \sup_{(\beta, t) \in K \times S} |L_n(\beta; t) - L(\beta; t)| \xrightarrow{\text{a.s.}} 0.$$

PROOF. Setting  $M = 2 \sup\{t^{-2}: t \in S\} < \infty$  and introducing the notation

$$\hat{\phi}_n(s; t) = \frac{1}{n} \sum_{j=1}^n e^{i(t\epsilon_j + \langle x_j, s \rangle)}$$

for the function figuring in  $\Delta_n$  of (4.1), we obtain by easy manipulation that

$$\begin{aligned} M_n^{(1)} & \leq M \sup_{(s, t) \in R \times S} \left| \log \left| \frac{\hat{\phi}_n(s; t)}{\phi(s)C(t)} \right| \right| \\ & \leq M \max \left\{ \log \left( 1 + \frac{\Delta_n}{\inf_{(s, t) \in R \times S} |\phi(s)C(t)|} \right), \log \left( 1 + \frac{\Delta_n}{\inf_{(s, t) \in R \times S} |\hat{\phi}_n(s; t)|} \right) \right\}. \end{aligned}$$

This upper bound tends to zero almost surely, since Lemma 1, or (4.1) also implies that

$$\inf_{(s, t) \in R \times S} |\hat{\phi}_n(s; t)| \xrightarrow{\text{a.s.}} \inf_{(s, t) \in R \times S} |\phi(s)C(t)|,$$

and this limit is strictly positive by conditions (2.v) and (2.vi).

We can now prove strong uniform consistency for the vector of functional least squares estimators.

THEOREM 1. Under conditions (2.i), (2.v) and (2.vi) we have

$$\sup_{t \in S} |\beta_n^*(t) - \beta_0| \xrightarrow{\text{a.s.}} 0.$$

PROOF. From Lemma 2 we infer that  $\sup_{t \in S} |\inf_{\beta \in K} L_n(\beta; t) - \inf_{\beta \in K} L(\beta; t)| \xrightarrow{\text{a.s.}} 0$ , or, what is the same, that

$$(4.4) \quad W(n) = \sup_{t \in S} |L_n(\beta_n^*(t); t) - L(\beta_0, t)| \xrightarrow{\text{a.s.}} 0.$$

It also follows from Lemma 2 that

$$(4.5) \quad W(n; \delta) = \sup_{t \in S} \left| \inf_{\beta \in K(\delta)} L_n(\beta; t) - \inf_{\beta \in K(\delta)} L(\beta; t) \right| \xrightarrow{\text{a.s.}} 0$$

for any  $\delta > 0$  for which the compact set  $K(\delta) = K \setminus \{\beta \in K: |\beta - \beta_0| < \delta\}$  is not empty.

Let  $n_k = n_k(\omega)$  be any random subsequence of the positive integers tending to infinity as  $k \rightarrow \infty$  on each elementary event  $\omega$ . Then (4.4) and (4.5) imply the following two equalities:

$$(4.6) \quad \Pr \left\{ \lim_{k \rightarrow \infty} W(n_k) = 0 \right\} = 1,$$

$$(4.7) \quad \Pr \left\{ \bigcap_{j=1}^{\infty} \left\{ \lim_{k \rightarrow \infty} W \left( n_k; \frac{1}{j} \right) = 0 \right\} \right\} = \Pr \{ B \} = 1.$$

What we have to show is that the event

$$\Omega_0 = \left\{ \limsup_{n \rightarrow \infty} \sup_{t \in S} |\beta_n^*(t) - \beta_0| > 0 \right\}$$

is of probability zero. Clearly  $\Omega_0 = \{m < \infty\}$ , where the random positive integer  $m = m(\omega)$  is defined as  $m = \inf\{j: \sup_{t \in S} |\beta_n^*(t) - \beta_0| > 1/j \text{ i.o.}\}$ ,  $\inf \emptyset = \infty$ . Let  $n_k = n_k(\omega) \rightarrow \infty$  be such that  $\sup_{t \in S} |\beta_{n_k}^*(t) - \beta_0| > 1/m$  for all  $k$  ( $n_k(\omega) = k$  if  $m(\omega) = \infty$ ). Then, by (4.6) and (4.7),

$$\begin{aligned} \Pr\{\Omega_0\} &= \Pr \left\{ \Omega_0 \cap \left\{ \lim_{k \rightarrow \infty} W(n_k) = 0 \right\} \right\} \\ &= \Pr \left\{ \Omega_0 \cap \left\{ \lim_{k \rightarrow \infty} \sup_{t \in S} \left| \inf_{\beta \in K(1/m)} L_{n_k}(\beta; t) - L(\beta_0; t) \right| = 0 \right\} \right\} \\ &= \Pr \left\{ \Omega_0 \cap \left\{ \lim_{k \rightarrow \infty} \sup_{t \in S} \left| \inf_{\beta \in K(1/m)} L_{n_k}(\beta; t) - L(\beta_0; t) \right| = 0 \right\} \cap B \right\} \\ &= 0 \end{aligned}$$

by the definition of event B in (4.7), for condition (2.vi) implies (see (2.3)) that  $L(\beta_0; t) < \inf_{\beta \in K(1/m)} L(\beta; t)$ ,  $t \in S$  for any  $m$ , and on no elementary event can the functions  $\inf_{\beta \in K(1/m)} L_n(\beta; t)$  have two different limit functions.

### 5. Weak convergence of $\beta_n^*$

When proving the weak convergence of the processes

$$(5.1) \quad B_n^*(t) = n^{1/2}(\beta_n^*(t) - \beta_0), \quad t \in S,$$

we start out from the estimating equations (1.4). Applying the one-term Taylor formula we obtain

$$l_n(\beta_n^*(t); t) = l_n(\beta_0; t) + A_n(\tilde{\beta}_n(t); t)(\beta_n^*(t) - \beta_0),$$

where  $\tilde{\beta}_n(t) = (\tilde{\beta}_{n1}(t), \dots, \tilde{\beta}_{np}(t))$  with component vectors  $\tilde{\beta}_{nk}(t)$  satisfying

$$(5.2) \quad \|\tilde{\beta}_{nk}(t) - \beta_0\| \leq \|\beta_n^*(t) - \beta_0\|, \quad t \in S, k = 1, \dots, p.$$

The matrix  $A_n(\tilde{\beta}_n(t); t)$  is obtained by evaluating the  $k$ th row of  $A_n(\beta; t)$  at  $\beta = \tilde{\beta}_{nk}(t)$ ,  $k = 1, \dots, p$ , where

$$A_n(\beta; t) = \begin{pmatrix} \frac{\partial l_{n1}(\beta; t)}{\partial \beta_1} & \frac{\partial l_{n1}(\beta; t)}{\partial \beta_2} & \dots & \frac{\partial l_{n1}(\beta; t)}{\partial \beta_p} \\ \vdots & \vdots & & \vdots \\ \frac{\partial l_{np}(\beta; t)}{\partial \beta_1} & \frac{\partial l_{np}(\beta; t)}{\partial \beta_2} & \dots & \frac{\partial l_{np}(\beta; t)}{\partial \beta_p} \end{pmatrix}$$

with  $l_{nk}(\beta; t)$  of (1.6), from which

$$(5.3)$$

$$\begin{aligned} a_{km}^{(n)}(\beta; t) &= \frac{\partial l_{nk}(\beta; t)}{\partial \beta_m} \\ &= -\frac{1}{n^2} \sum_{r=1}^n \sum_{j=1}^n x_{jk}(x_{jm} - x_{rm}) \cos(t\{Y_j - Y_r - \langle x_j - x_r, \beta \rangle\}), \end{aligned}$$

$1 \leq k, m \leq p$ . Since  $l_n(\beta_n^*(t); t) = (0, \dots, 0)$ , a.s. for  $n$  large enough, we have

$$(5.4) \quad -A_n(\tilde{\beta}_n(t); t)B_n^*(t) = n^{1/2}l_n(\beta_0; t).$$

Therefore, in order to prove weak convergence in  $\mathcal{C}^p(S)$  for  $B_n^*$  of (5.1), it will suffice to show that the matrix  $-A_n(\tilde{\beta}_n(t); t)$  converges uniformly in probability to some nonrandom matrix  $A(t)$  and that the vector process  $n^{1/2}l_n(\beta_0; t)$  converges weakly.

LEMMA 3. Under conditions (2.ii), (2.iv) and (3.i) we have for any matrix norm  $|\cdot|$

$$\sup_{t \in S} |(-A_n(\beta_0; t) - A(t))| \xrightarrow{\text{Pr}} 0,$$

where  $A(t) = |C(t)|^2 A$  with the  $p \times p$  matrix  $A$  of condition (2.iii).

PROOF. Convergence for any matrix norm will follow if we show that

$$(5.5) \quad \sup_{t \in S} |(-a_{km}^{(n)}(t)) - |C(t)|^2 a_{km}| \xrightarrow{\text{Pr}} 0, \quad 1 \leq k, m \leq p,$$

where  $a_{km}^{(n)}(t) = a_{km}^{(n)}(\beta_0; t)$ . Because

$$(5.6) \quad Y_j - Y_r - \langle x_j - x_r, \beta_0 \rangle = \varepsilon_j - \varepsilon_r$$

in expression (5.3) for the latter, we obtain, upon applying the cosine addition formula and some rearrangements,

$$\begin{aligned} -a_{km}^{(n)}(t) &= \text{Re } C_n(t) \text{Re} \left\{ \frac{1}{n} \sum_{j=1}^n x_{jk} x_{jm} e^{it\varepsilon_j} \right\} \\ &\quad + \text{Im } C_n(t) \text{Im} \left\{ \frac{1}{n} \sum_{j=1}^n x_{jk} x_{jm} e^{it\varepsilon_j} \right\} \\ &\quad - \text{Re} \left\{ \frac{1}{n} \sum_{j=1}^n x_{jk} e^{it\varepsilon_j} \right\} \text{Re} \left\{ \frac{1}{n} \sum_{j=1}^n x_{jm} e^{it\varepsilon_j} \right\} \\ &\quad - \text{Im} \left\{ \frac{1}{n} \sum_{j=1}^n x_{jk} e^{it\varepsilon_j} \right\} \text{Im} \left\{ \frac{1}{n} \sum_{j=1}^n x_{jm} e^{it\varepsilon_j} \right\}. \end{aligned}$$

Because of (3.1) and conditions (2.ii) and (2.iv) it is enough to show that if  $b_1, b_2, \dots$  are real numbers such that

$$\frac{1}{n} \sum_{j=1}^n b_j \rightarrow b, \quad \frac{1}{n^2} \sum_{j=1}^n b_j^2 \rightarrow 0$$

and

$$\frac{1}{n} \max_{1 \leq j \leq n} b_j^2 = \frac{1}{n} \max_{1 \leq j \leq n} \frac{b_j^2}{(1/n) \sum_{r=1}^n b_r^2} \rightarrow 0,$$

then

$$M_n^{(2)} = \sup_{t \in S} \left| \frac{1}{n} \sum_{j=1}^n b_j e^{it\varepsilon_j} - bC(t) \right| \xrightarrow{\text{Pr}} 0.$$

This is indeed enough, since for both choices  $b_j = x_{jk}x_{jm}$ ,  $b_j = x_{jk}$  the above three conditions are ensured by (2.ii) and (2.iv). Now

$$M_n^{(2)} \leq \left| \frac{1}{n} \sum_{j=1}^n b_j - b \right| + \sup_{t \in S} \left| \frac{1}{n} \sum_{j=1}^n b_j (e^{it\epsilon_j} - C(t)) \right|$$

$$= \left| b - \frac{1}{n} \sum_{j=1}^n b_j \right| + \left( \frac{1}{n^2} \sum_{j=1}^n b_j^2 \right) \sup_{t \in S} \left| \frac{1}{n^{1/2}} \sum_{j=1}^n b_{nj} (e^{it\epsilon_j} - C(t)) \right|,$$

and this bound goes to zero in probability, for the latter sup has a limit distribution by Theorem B.

We note that the third relation in condition (2.iv) was not used here, and so it is not needed in Lemma 4 below either. Putting now together (5.2), Theorem 1 and Lemma 3, we obtain easily the following.

LEMMA 4. Under conditions (2.i), (2.ii), (2.iv), (2.v), (2.vi) and (3.i) we have

$$\sup_{t \in S} \left| (-A_n(\tilde{\beta}_n(t); t) - |C(t)|^2 A) \right| \xrightarrow{\text{Pr}} 0.$$

Now we are able to prove the weak convergence of  $B_n^*$  of (5.1).

THEOREM 2. Assume the six conditions (2.i)–(2.vi) hold. Then  $B_n^*(\cdot)$  converges weakly in  $\mathcal{C}^p(S)$  to the Gaussian process  $G(\cdot) = (G_1(\cdot), \dots, G_p(\cdot))$  with mean vector

$$E(G_1(t), \dots, G_p(t)) = (0, \dots, 0)$$

and covariance matrix

$$E \begin{pmatrix} G_1(t)G_1(s) & \cdots & G_1(t)G_p(s) \\ \vdots & & \\ G_p(t)G_1(s) & \cdots & G_p(t)G_p(s) \end{pmatrix} = \sigma(t, s)A^{-1},$$

where, with  $C(t) = U(t) + iV(t)$ ,

$$\sigma(t, s) = \frac{h(t, s)}{ts |C(t)|^2 |C(s)|^2},$$

$$h(t, s) = \frac{1}{2} \{ U(t-s)[U(t)U(s) + V(t)V(s)] \\ + U(t+s)[V(t)V(s) - U(t)U(s)] \\ + V(t-s)[V(t)U(s) - U(t)V(s)] \\ - V(t+s)[U(t)V(s) + V(t)U(s)] \},$$

if and only if condition (3.i) holds.

PROOF. In view of (5.4), Lemma 4 and condition (2.iii) it is enough to prove that  $n^{1/2}l_n(\beta_0; t)$  converges weakly in  $\mathcal{C}^p(S)$  to a Gaussian process  $Q(t) = (Q_1(t), \dots, Q_p(t))$  with zero mean vector and covariance matrix

$$\frac{h(t, s)}{ts} A.$$

Introducing the notation

$$(5.7) \quad y_{jk} = x_{jk} - \bar{x}_k, \quad k = 1, \dots, p; j = 1, 2, \dots,$$

we start out from formula (1.6) for the  $k$ th component of the process in question. Using (5.6) and the sine addition formula and then centering the terms, that is, following exactly the lines of Chambers and Heathcote (1981), we obtain after some rearrangements that

$$\begin{aligned} n^{1/2}l_{nk}(\beta_0; t) &= \frac{1}{t} \frac{1}{n^{1/2}} \sum_{j=1}^n y_{jk} \{U(t)\sin(\varepsilon_j t) - V(t)\cos(\varepsilon_j t)\} \\ &\quad + \left( \frac{1}{n} \sum_{j=1}^n y_{jk} \right) \frac{1}{t} \frac{1}{n^{1/2}} \sum_{j=1}^n \{V(t)\cos(\varepsilon_j t) - U(t)\sin(\varepsilon_j t)\} \\ &\quad + \operatorname{Re}\{C_n(t) - C(t)\} \frac{1}{t} \frac{1}{n^{1/2}} \sum_{j=1}^n y_{jk} \{\sin(\varepsilon_j t) - V(t)\} \\ &\quad - \operatorname{Im}\{C_n(t) - C(t)\} \frac{1}{t} \frac{1}{n^{1/2}} \sum_{j=1}^n y_{jk} \{\cos(\varepsilon_j t) - U(t)\} \\ &= \frac{1}{t} \{U(t)\operatorname{Im} Z_{nk}(t) - V(t)\operatorname{Re} Z_{nk}(t)\} \\ &\quad + \left( \frac{1}{n} \sum_{j=1}^n y_{jk} \right) \frac{1}{t} \{V(t)\operatorname{Re} D_n(t) - U(t)\operatorname{Im} D_n(t)\} \\ &\quad + \operatorname{Re}\{C_n(t) - C(t)\} \frac{1}{t} \operatorname{Im} Z_{nk}(t) - \operatorname{Im}\{C_n(t) - C(t)\} \frac{1}{t} \operatorname{Re} Z_{nk}(t), \end{aligned}$$

after extra centering in the first two lines of the first equation, where  $D_n(t)$  is the process of Theorem A and

$$Z_{nk}(t) = n^{-1/2} \sum_{j=1}^n y_{jk} \{e^{it\varepsilon_j} - C(t)\}.$$

We have

$$(5.8) \quad \left( \frac{1}{n} \sum_{j=1}^n y_{jk} \right) \sup_{t \in S} \left| \frac{1}{t} \{V(t)\operatorname{Re} D_n(t) - U(t)\operatorname{Im} D_n(t)\} \right| \xrightarrow{\Pr} 0,$$

since by (5.7) and the condition (2.ii) the average of the  $y_{jk}$  converges to zero and the sup in (5.8) has a limit distribution, under condition (3.i), by Theorem A. So the problem is reduced to the weak convergence problem of the  $Z_{nk}(t)$  processes.

Consider the normalised process

$$D_n^{b(k)}(t) = \frac{Z_{nk}(t)}{\left((1/n)\sum_{j=1}^n y_{jk}^2\right)^{1/2}} = \frac{1}{n^{1/2}} \sum_{j=1}^n b_{nj}(k) \{e^{it\epsilon_j} - C(t)\}$$

with

$$b_{nj}(k) = \frac{y_{jk}}{\left((1/n)\sum_{j=1}^n y_{jk}^2\right)^{1/2}}.$$

The third relation of condition (2.iv) ensures that the array  $b_{nj}(k)$  satisfies the conditions of Theorem B for any  $k, 1 \leq k \leq p$ . Therefore, under condition (3.i),  $D_n^{b(k)}(\cdot)$  converges weakly in  $\mathcal{C}(S) = \mathcal{C}^2(S)$  to the Gaussian process  $D(\cdot)$  of Theorems A and B. Consequently,  $Z_{nk}(\cdot)$  converges weakly in  $\mathcal{C}(S)$  to the process  $(\bar{a}_{kk})^{1/2}D(\cdot), k = 1, \dots, p$ , by condition (2.ii). Hence, the second sup possessing a limit distribution,

$$(5.9) \quad \sup_{t \in S} |C_n(t) - C(t)| \sup_{t \in S} \left| \frac{1}{t} Z_{nk}(t) \right| \xrightarrow{\text{Pr}} 0$$

for any  $k, 1 \leq k \leq p$ , by (3.i). The above formula for the  $k$ th component, (5.8) and (5.9) imply that under (3.i) the weak limit in  $\mathcal{C}^p(S)$  of the processes  $n^{1/2}l_n(\beta_0; t)$  is identical to that of the processes

$$Q_n(t) = \frac{H_n(t)}{t} = \frac{1}{t} \{U(t)\text{Im } Z_n(t) - V(t)\text{Re } Z_n(t)\},$$

$Z_n(t) = (Z_{n1}(t), \dots, Z_{np}(t))$ , if the latter converges weakly. But the sequences of the individual components of  $Q_n$  converge weakly in  $\mathcal{C}(S)$ , hence they are tight. This obviously implies the tightness of the vector valued sequence  $\{Q_n(\cdot)\}$ . As to the convergence of its finite-dimensional distributions to those of  $Q(\cdot)$  it is enough to note that with  $H_n(t) = (H_{n1}(t), \dots, H_{np}(t))$  we have

$$E \begin{pmatrix} H_{n1}(t)H_{n1}(s) & \cdots & H_{n1}(t)H_{np}(s) \\ \vdots & & \vdots \\ H_{np}(t)H_{n1}(s) & \cdots & H_{np}(t)H_{np}(s) \end{pmatrix} \rightarrow h(t, s)A$$

by condition (2.ii), since the third relation of condition (2.iv) ensures this convergence via Liapunov's central limit theorem for a triangular array of row-wise independent random vectors referred to in the proof of Theorem B. Hence the sufficiency part of the theorem is proved.

The necessity of condition (3.i) is trivial, since the limit process  $G(t)$  is sample-continuous if and only if  $H(t)$ , the limit process of  $H_n(t)$ , is such. But the *individual components* of  $H(t)$  can be represented as

$$(5.10) \quad H_k(t) = (\bar{a}_{kk})^{1/2} \{U(t)\text{Im } D(t) - V(t)\text{Re } D(t)\}, \quad k = 1, \dots, p.$$

with  $D(t)$  of (3.2), and this is sample-continuous if and only if (3.i) holds.

REMARK. (5.10) is equality in distribution for the two processes for each single  $k = 1, \dots, p$ . It does not mean that the distribution of  $\bar{H}(t) = ((\bar{a}_{11})^{1/2}\{ \}, \dots, (\bar{a}_{pp})^{1/2}\{ \})$  would coincide with that of  $H(t)$  in  $\mathcal{C}^p(S)$ . In fact a representation in the form of  $\bar{H}$  for  $H$  is impossible because the covariance matrix of  $\bar{H}$  is of rank 1. I owe this remark to Peter Hall.

### 6. The variance function and its estimator

Chambers and Heathcote (1981) call

$$(6.1) \quad \begin{aligned} \sigma^2(t) &= \sigma(t, t) \\ &= \frac{U^2(t) + V^2(t) + U(2t)[V^2(t) - U^2(t)] - 2V(2t)U(t)V(t)}{2t^2[U^2(t) + V^2(t)]^2} \\ &= \frac{1}{2t^2} g(U(t), V(t), U(2t), V(2t)) \end{aligned}$$

the variance function. Consider the random complex function

$$\begin{aligned} C_n^*(t) &= C_n^*(t; \beta_n^*(t)) = U_n^*(t; \beta_n^*(t)) + iV_n^*(t; \beta_n^*(t)) \\ &= U_n^*(t) + iV_n^*(t) \end{aligned}$$

defined by

$$(6.2) \quad C_n^*(t) = \frac{1}{n} \sum_{j=1}^n e^{it[Y_j - \langle x_j, \beta_n^*(t) \rangle]},$$

and form the estimator

$$(6.3) \quad \sigma_n^2(t) = \frac{1}{2t^2} g(U_n^*(t), V_n^*(t), U_n^*(2t), V_n^*(2t))$$

where we adopted the convention  $C_n^*(2t) = C_n^*(2t; \beta_n^*(t))$ .

THEOREM 3. Under conditions (2.i), (2.v) and (2.vi) we have

$$\sup_{t \in S} |\sigma_n^2(t) - \sigma^2(t)| \xrightarrow{\text{a.s.}} 0.$$

PROOF. It is clearly enough to show that

$$(6.4) \quad M_n^{(3)} = \sup_{t \in S} |C_n^*(t) - C(t)| \xrightarrow{\text{a.s.}} 0.$$

This is easy now since

$$\begin{aligned} M_n^{(3)} &\leq \sup_{t \in S} |C_n^*(t) - C(t)\phi(t(\beta_0 - \beta_n^*(t)))| \\ &\quad + \sup_{t \in S} |C(t)\phi(t(\beta_0 - \beta_n^*(t))) - C(t)| \\ &\leq \Delta_n + \sup_{t \in S} |\phi(t(\beta_0 - \beta_n^*(t))) - 1|, \end{aligned}$$

where  $\Delta_n$  is of Lemma 1, converging to zero almost surely. The argument of  $\phi$  converges almost surely uniformly to zero by Theorem 1,  $\phi(0, \dots, 0) = 1$ , and since  $\phi$ , being a characteristic function, is uniformly continuous on the whole  $\mathbf{R}^p$ , the second term also goes to zero almost surely.

Since  $\sigma^2(t)$  is always an even function, from now on we shall use only positive  $t$ 's. Let  $t_0$  be the smallest number for which the infimum of  $\sigma^2(t)$  is attained, that is,

$$t_0 = \inf \left\{ s: \sigma^2(s) = \inf_{t \geq 0} \sigma^2(t) \right\}.$$

We assume

$$(6.i) \quad t_0 > 0.$$

This is the case (Chambers and Heathcote (1981), Theorem 2) if  $\epsilon$  lacks a variance, or if its distribution function is leptokurtic. In these cases there is a unique  $s$  for which the inside infimum is attained in the definition of  $t_0$ . Let now  $S_0 = [a_0, b_0]$  be any such interval that  $0 < a_0 < t_0 < b_0$  and  $\sigma^2(t_0) < \sigma^2(t)$  for any other  $t$  in  $S_0$ , and introduce the random variable

$$t_n = \inf \left\{ s: \sigma_n^2(s) = \inf_{t \in S_0} \sigma_n^2(t) \right\}.$$

Condition (6.i) and the choice of  $S_0$ , hence the definition of  $t_n$  itself, leads to the following simple result.

LEMMA 5.  $t_n \xrightarrow{\text{a.s.}} t_0$ .

PROOF. Let  $\delta > 0$  be any such number that the set  $S(\delta) = S_0 \setminus \{t \in S_0: |t - t_0| < \delta\}$  is non-empty. Then by Theorem 3 we have

$$\sigma_n^2(t_n) = \inf_{t \in S_0} \sigma_n^2(t) \xrightarrow{\text{a.s.}} \inf_{t \in S_0} \sigma^2(t) = \sigma^2(t_0)$$

and

$$\inf_{t \in S(\delta)} \sigma_n^2(t) \xrightarrow{\text{a.s.}} \inf_{t \in S(\delta)} \sigma^2(t) > \sigma^2(t_0).$$

Using these relations the proof can be completed along the lines of those of Theorem 1.

### 7. The adaptive least squares estimator

Throughout in this section condition (6.i) will be assumed and  $S_0$  will be the interval of the preceding section. Accordingly, conditions (2.v) and (2.vi) will be interpreted on this  $S_0$  instead of the general  $S$ . We call  $\beta_n^*(t_n)$  the adaptive least squares estimate for  $\beta_0$ . It is obvious from Theorem 1 that  $\beta_n^*(t_n) \xrightarrow{\text{a.s.}} \beta_0$ , that is, this estimator is strongly consistent provided conditions (2.i), (2.v) and (2.vi) hold. The main result of this paper is the following.

**THEOREM 4.** *If conditions (2.i)–(2.vi) and (3.i) hold, then the distribution of  $n^{1/2}(\beta_n^*(t_n) - \beta_0)$  converges to the  $p$ -variate normal distribution with zero mean vector and covariance matrix  $\sigma^2(t_0)A^{-1}$ .*

**PROOF.** Under the conditions stated,  $B_n^*(\cdot)$  of (5.1) converges weakly in  $\mathcal{C}^p(S_0)$  to  $G(\cdot)$  of Theorem 2. According to a well-known theorem of Skorohod (1956), the original sequence  $\{\varepsilon_j\}$  can be redefined, without changing its distribution, on some probability space  $(\Omega, \mathcal{A}, P)$  carrying a copy of  $G(\cdot)$  such that for the resulting  $\beta_n^*(\cdot)$  we have

$$\sup_{t \in S_0} |B_n^*(t) - G(t)| \xrightarrow{\text{a.s.}} 0.$$

We work on this probability triple and claim that

$$(7.1) \quad M_n^{(4)} = |B_n^*(t_n) - G(t_0)| \xrightarrow{\text{a.s.}} 0.$$

Indeed,

$$\begin{aligned} M_n^{(4)} &\leq |B_n^*(t_n) - G(t_n)| + |G(t_n) - G(t_0)| \\ &\leq \sup_{t \in S_0} |B_n^*(t) - G(t)| + |G(t_n) - G(t_0)| \xrightarrow{\text{a.s.}} 0. \end{aligned}$$

Here the first term goes to zero by construction, while the second by Lemma 5 and the sample-continuity of  $G$ . (7.1) implies that  $B_n^*(t_n)$  converges in distribution to  $G(t_0)$ . But this statement is independent of the choice of the underlying probability space and is therefore identical to the claim of the theorem.

## 8. Concluding remarks

Chambers and Heathcote (1981) and Heathcote (1982) propose methods for the computation of  $\beta_n^*(t)$  and  $t_n$ . If the user finds  $t_n$  very close to zero, then this indicates that ordinary least squares estimation can be used. Otherwise the following procedure may be used. If the data contains a suspected outlier, then upon eliminating it one computes a new  $t_{n-1}$  to see whether it is 'zero'. If so, then reject the eliminated value and use the ordinary least squares estimate  $\hat{\beta}_{n-1}$ . If not and there are more suspected outliers, then this procedure, proposed by Chambers and Heathcote (1981) and Heathcote (1982) who also demonstrate it on a concrete example, can be continued. If after  $k$  steps  $t_{n-k}$  is still not zero but there are no more suspected outliers, then it is perhaps best to conclude that the removed data points are not outliers, or gross errors, but represent essential features of the actual error distribution such as long tail, asymmetry, or leptokurtic behaviour. In this case the proposed estimator is  $\beta_n^*(t_n)$ .

Heathcote (1982) explains that if the sample size is not too large, then  $(n/(n-p))\sigma_n^2(t)$  (or  $(n/(n-p-1))\sigma_n^2(t)$  if an intercept term is present) is to be preferred for estimating  $\sigma^2(t)$ , and this should be minimised to find  $t_n$ . Of course, the asymptotic theory remains the same.

However large samples we are given in practice, it is always impossible of course to check all the conditions (2.i)–(2.vi). But this is in fact the case with the fewer conditions ((2.ii), (2.iii), part of (2.iv)) for the asymptotic normality of ordinary least squares. It is hard to imagine practical situations where condition (3.3) (sufficient for (3.i)) would be violated. Nevertheless this condition can be checked empirically. If in a small neighborhood of the origin the curve of the function  $1 - \operatorname{Re} C_n^*(t)$ , with  $C_n^*(t)$  of (6.2), lies under the curve of  $1/\log(1/|t|)$ , then by (6.4) and the discussion following formula (1.17) in Csörgő (1981a) we may well conclude that (3.i), or even (3.3), is satisfied.

## Acknowledgement

I wish to thank Peter Hall, Chip Heathcote and two anonymous referees for their helpful remarks.

## References

- R. L. Chambers and C. R. Heathcote (1981), 'On the estimation of slope and the identification of outliers in linear regression', *Biometrika* **68**, 21–33.
- K. L. Chung (1974), *A course in probability theory* (Academic Press, New York).
- S. Csörgő (1980), *Empirical characteristic functions* (Carleton Mathematical Lecture Notes No. 26, Carleton University, Ottawa, Canada).

- S. Csörgő (1981a), 'Limit behaviour of the empirical characteristic function', *Ann. Probability* **9**, 130–144.
- S. Csörgő (1981b), 'Multivariate empirical characteristic functions', *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **55**, 203–229.
- X. Fernique (1978), 'Continuité et théorème central limite pour les transformées de Fourier des mesures aléatoires du second ordre', *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **42**, 57–66.
- C. R. Heathcote (1982), 'Linear regression by functional least squares', *J. Appl. Probability* **19A**, *Essays in Statistical Science*, P.A.P. Moran Festschrift, edited by J. Gani and E. J. Hannan, pp.
- M. B. Marcus (1981), 'Weak convergence of the empirical characteristic function', *Ann. Probability* **9**, 194–201.
- W. Rudin (1976), *Principles of mathematical analysis* (McGraw-Hill, Tokyo).
- A. V. Skorohod (1956), 'Limit theorems for stochastic processes', *Teor. Veroyatnost. i Primenen.* **1**, 289–319. English translation: *Theory of Probability Appl.* **1**, 261–290.

Bolyai Institute  
Szeged University  
H-6720 Szeged  
Hungary