JOURNAL OF MULTIVARIATE ANALYSIS 31, 59–68 (1989)

Order of Convergence of Regression Parameter Estimates in Models with Infinite Variance

A. LE BRETON AND M. MUSIELA*

University of Grenoble, Grenoble, France; and
University of New South Wales, Kensington, New South Wales, Australia

Communicated by C. R. Rao

A semimartingale driven continuous time linear regression model is studied. Assumptions concerning errors allow us to consider also models with infinite variance. The order of the almost sure convergence of a class of estimates which includes least squares estimates is given. In the presence of errors with heavy tails a modification of least squares estimates is suggested and shown to be better than the latter.

1. INTRODUCTION AND THE MAIN RESULT

The aim of this paper is to study the strong consistency of regression parameter estimates in a multiple linear regression model of the form

$$y_t = \int_{[0,t]} \theta^* x_s \, dV_s + e_t, \qquad t \geq 0. \tag{1.1}$$

Here $\theta = (\theta^1, ..., \theta^d)^*$ is an unknown parameter in $\mathbf{R}^d$ and $x = (x_t; t \geq 0)$ is a nonrandom Borel measurable locally bounded $\mathbf{R}^d$-valued function standing for a given design. The deterministic increasing $\mathbf{R}^+$-valued function $V = (V_t; t \geq 0)$ represents a given time scale; it is assumed to be right continuous and have limits from the left (cadlag.). The process of errors $e = (e_t; t \geq 0)$ is a cadlag. $\mathbf{R}$-valued semimartingale. Finally, $y = (y_t; t \geq 0)$ stands for a response process. Elementary facts and notations concerning semimartingales and stochastic integration which are used here can be found, e.g., in Jacod [3].

Received January 29, 1988; revised January 1, 1989.

AMS 1980 subject classifications: Primary 62J05; Secondary 62M10, 60F15.

Key words and phrases: multiple regression, strong consistency, semimartingale, stochastic integration.

* Research supported by the Australian Research Grants Scheme.

0047-259X/89 $3.00

For a nonrandom Borel measurable locally bounded function $c = (c_t; t \geq 0)$ and a cadlag. real semimartingale $S = (S_t; t \geq 0)$ we use $c \cdot S = (c \cdot S_t; t \geq 0)$ to denote the cadlag. semimartingale defined by

$$c \cdot S_t = \int_{[0,t]} c_s \, dS_s; \qquad t \geq 0.$$

$\mathbf{L}(S)$ (resp. $\mathbf{L}^p(V)$) stands for the set of all such real functions c for which $c \cdot S_{\infty-} = \lim_{t \to \infty} c \cdot S_t$ exists in $\mathbf{R}$ a.s. (resp. $|c|^p \cdot V_{\infty-} < \infty$).

Let $w\colon \mathbf{R}^+ \to \mathbf{R}^+$ be a Borel measurable locally bounded function and let the matrix valued function $\Lambda\colon \mathbf{R}^+ \to \mathbf{R}^d \otimes \mathbf{R}^d$ be given by

$$\Lambda_t = xx^* w \cdot V_t; \qquad t \geq 0. \tag{1.2}$$

Note that $\Lambda = (\Lambda_t; t \geq 0)$ is increasing symmetric and positive (≥ 0). When the matrix Λ_t is positive definite (> 0) we define an estimate of θ, based on the observation $(y_s; 0 \leq s \leq t)$ of the response process on the interval $[0, t]$, by the formula

$$\theta_t = \Lambda_t^{-1}(xw \cdot y_t). \tag{1.3}$$

The above estimate reduces to the least squares estimate when $w = 1$.

In recent years several authors attempted to prove the strong consistency of θ_t in the case of $w = 1$ under minimal assumptions on the design and errors. It seems that for discrete time regression models the most general results were obtained by Lai, Robbins, and Wei [5] and Chen, Lai, and Wei [2]. Le Breton and Musiela [6] have extended these results to the case of continuous time models of form (1.1) proving that $\theta_{\infty-} = \theta$ a.s. if $\Lambda_{\infty-}^{-1} = 0$ and $\mathbf{L}^2(V) \subset \mathbf{L}(e)$ and finding the order of convergence. In an attempt to cover a larger class of errors they have also obtained [7] the order of convergence of θ_t to θ assuming $\Lambda_{\infty-}^{-1} = 0$ and $\mathbf{L}^2(V) \subset \mathbf{L}(g \cdot e)$, where $g\colon \mathbf{R}^+ \to \mathbf{R}$ is a nonvanishing Borel measurable real function such that $|g|$ is decreasing. Unfortunately important classes of errors which appear, for instance, in financial economics (cf., e.g., Mandelbrot [8]) fail to satisfy either of these conditions. However, stable errors of order p or more generally Pareto-like errors of order p with $p < 2$ which fail to satisfy $\mathbf{L}^2(V) \subset \mathbf{L}(e)$ fulfil $\mathbf{L}^p(V) \subset \mathbf{L}(e)$. It seems to us that this is one of the reasons for modifying the least squares estimate which performs well under $\mathbf{L}^2(V) \subset \mathbf{L}(e)$ but is not robust in the presence of errors with heavy tails, the latter corresponding to $\mathbf{L}^p(V) \subset \mathbf{L}(e)$.

To illustrate the above let us consider the following:

EXAMPLE 1. Let $\varepsilon = (\varepsilon_n; n \geq 1)$ be a sequence of i.i.d. p-stable random variables with characteristic function $\exp(-|t|^p)$, $1 < p < 2$, and let the response sequence $y = (y_n; n \geq 1)$ be given by

$$\Delta y_n = y_n - y_{n-1} = \theta n^{(1-p)/p} + \varepsilon_n, \qquad n \geq 1,\ y_0 = 0,$$

where $\theta \in \mathbf{R}$. Then the least squares estimate of θ, based on the observation $(y_k; 1 \leqslant k \leqslant n)$, is of the form

$$\theta_n = \left(\sum_{k=1}^{n} k^{2(1-p)/p} \right)^{-1} \sum_{k=1}^{n} k^{(1-p)/p} \Delta y_k.$$

It is easy to see that $E \exp(it(\theta_n - \theta)) = \exp(-a_n |t|^p)$, where

$$a_n = \left(\sum_{k=1}^{n} k^{2(1-p)/p} \right)^{-p} \sum_{k=1}^{n} k^{1-p}$$

and therefore since $a_{\infty -} = (2-p)^{p-1} p^{-p}$, $\theta_n - \theta$ converges in distribution to a p-stable random variable with characteristic function $\exp(-(2-p)^{p-1} p^{-p} |t|^p)$. Note that the first condition of convergence, namely $\Lambda_{\infty -}^{-1} = 0$, holds since $\sum_{k=1}^{\infty} k^{2(1-p)/p} = \infty$, but the second, that is, $l^2 = \{c = (c_n; n \geqslant 1);\ \sum_{n=1}^{\infty} |c_n|^2 < \infty\} \subset \mathbf{L}(e)$, fails to be fulfilled because $\mathbf{L}(e) = l^p = \{c = (c_n; n \geqslant 1); \sum_{n=1}^{\infty} |c_n|^p < \infty\}$, where $e_n = \sum_{k=1}^{n} \varepsilon_k$ (cf., e.g., Cambanis, Rosiński, and Woyczyński [1]). On the other hand, if we use the estimate

$$\theta_n^{\#} = \left(\sum_{k=1}^{n} k^{-1} \right)^{-1} \sum_{k=1}^{n} k^{(-1)/p} \Delta y_k,$$

then we have $\theta_n^{\#} - \theta = (\sum_{k=1}^{n} k^{-1})^{-1} \sum_{k=1}^{n} k^{(-1)/p} \varepsilon_k$. But, from the inequality $\sum_{k=1}^{n} k^{-1} (\sum_{j=1}^{k} j^{-1})^{-p} \leqslant p/(p-1)$, we deduce that $\sum_{k=1}^{n} k^{(-1)/p} (\sum_{j=1}^{k} j^{-1})^{-1} \varepsilon_k$ converges in $\mathbf{R}$ a.s., which in turn, using Kronecker's lemma, implies that $\theta_{\infty -}^{\#} = \theta$ a.s.

Summarizing, the least squares estimate is not consistent and one can construct a linear estimate of θ which is strongly consistent. However, even if it persuades some of the readers that it is necessary to modify the classical estimate it still does not answer the question how to do it. We do not know how to solve this problem in general but we shall now try to justify our choice by investigating an extension of Example 1.

EXAMPLE 2. Let ε be as in Example 1 and let

$$\Delta y_n = y_n - y_{n-1} = \theta x_n + \varepsilon_n; \qquad n \geqslant 1,\ y_0 = 0,$$

where $\theta \in \mathbf{R}$ and $(x_n; n \geqslant 1)$ is some sequence of real numbers. Consider a linear unbiased estimate of θ based on $(y_k; 1 \leqslant k \leqslant n)$, i.e., $\theta_n = \sum_{k=1}^{n} \lambda_k \Delta y_k$. Since $E\theta_n = (\sum_{k=1}^{n} \lambda_k x_k)\theta$, one must have $\sum_{k=1}^{n} \lambda_k x_k = 1$. Therefore $\theta_n - \theta$ is a p-stable random variable with characteristic function $\exp\{-(\sum_{k=1}^{n} |\lambda_k|^p) |t|^p\}$. We argue that one should choose an estimate which minimizes the "dispersion" $\sum_{k=1}^{n} |\lambda_k|^p$ under the constraint $\sum_{k=1}^{n} \lambda_k x_k = 1$. From

$$1 = \left| \sum_{k=1}^{n} \lambda_k x_k \right| \leqslant \left(\sum_{k=1}^{n} |\lambda_k|^p \right)^{1/p} \left(\sum_{k=1}^{n} |x_k|^{p/(p-1)} \right)^{(p-1)/p}$$

we have

$$\sum_{k=1}^{n} |\lambda_k|^p \geqslant \left(\sum_{k=1}^{n} |x_k|^{p/(p-1)} \right)^{1-p}$$

and therefore the minimum is attained when

$$\lambda_k = |x_k|^{(2-p)/(p-1)} x_k \left(\sum_{k=1}^{n} |x_k|^{p/(p-1)} \right)^{-1}.$$

So, in the proposed sense, a "best" linear unbiased estimate of θ is given by

$$\theta_n^{\#} = \left(\sum_{k=1}^{n} w_k x_k^2 \right)^{-1} \sum_{k=1}^{n} w_k x_k \Delta y_k,$$

where $w_k = |x_k|^{(2-p)/(p-1)}$. Note that in Example 1, $\theta_n^{\#}$ coincides with that defined here.

Taking into account the above considerations, we suggest, in addition, that in case (1.1), when $\mathbf{L}^p(V) \subset \mathbf{L}(e)$ and the matrix

$${}_{\#}\Lambda_t = xx^* \, |x|^{(2-p)/(p-1)} \cdot V_t$$

is positive definite, the estimate given by

$$\theta_t^{\#} = {}_{\#}\Lambda_t^{-1} (x \, |x|^{(2-p)/(p-1)} \cdot y_t) \tag{1.4}$$

be used.

The aim of this paper is to establish the order of the almost sure convergence of estimates of form (1.3) and show that the estimate $\theta_t^{\#}$ given in (1.4) corresponds to an optimal choice of w in the sense that condition (1.5) below holds for any design x. For that we shall prove the following theorem in Section 2.

THEOREM. *Let V and e be as in (1.1) and let $\mathbf{L}^p(V) \subset \mathbf{L}(g \cdot e)$, where $g: \mathbf{R}^+ \to \mathbf{R}$ is a nonvanishing Borel measurable function such that $|g|$ is* *decreasing while* $1 < p \leqslant 2$. *Let the matrix* A_{t_0-}, *defined in* (1.2), *be positive definite and for all* $t \geqslant t_0 > 0$, *let* $v_t^i = 1/\mu_t^{ii}$, *where* $A_t^{-1} = ((\mu_t^{ij}))$. *Assume that there exists a constant* $C > 0$ *such that*

$$w_t \leqslant C\,|x|_t^{(2-p)/(p-1)} \qquad \text{for all} \quad t \geqslant 0,\ (0^0 = 1). \tag{1.5}$$

If

$$\lim_{t\to\infty} v_t^i = \infty, \qquad \lim_{t\to\infty}\left(\int_{]t_0,t]} (f\circ v^i)_s^{-1}\, d\operatorname{tr} A_s\right)^{1-p/2} < \infty, \tag{1.6}$$

where $f\colon \mathbf{R}^+ \to \mathbf{R}^+$ *is an increasing function such that* $\int_0^\infty f_t^{-1}\,dt < \infty$, *then the* i*th component* θ_t^i *of estimate* (1.3) *satisfies*

$$\theta_t^i - \theta^i = o((f\circ v^i)^{1/p}\,|v^i g|_t^{-1}) \qquad \text{a.s..} \tag{1.7}$$

Remarks. 1. Note, in particular, that the theorem asserts that if $\mathbf{L}^p(V) \subset \mathbf{L}(e)$ is satisfied then the estimate $\theta_t^\#$ is strongly consistent when ${}_\# A_{\infty-}^{-1} = 0$ and $\lim_{t\to\infty} (\int_{]t_0,t]} (\operatorname{tr}{}_\# A_s^{-1})^p\, d\operatorname{tr}{}_\# A_s)^{1-(p/2)} < \infty$. Moreover, when $p = 2$, $\theta_t^\#$ is the least squares estimate and the last condition is trivially satisfied so that the statement above is nothing but Theorem 1 of Le Breton and Musiela [6].

2. Note that if $\mathbf{L}^2(V) \subset \mathbf{L}(g\cdot e)$ and $w \equiv 1$, the statement in the theorem reduces to the result of Le Breton and Musiela [7], where the readers may find examples (see also Chen, Lai, and Wei [2]). Discussed in Section 3 are some examples where $\mathbf{L}^p(V) \subset \mathbf{L}(g\cdot e)$ for $1 < p < 2$.

3. The analytic description of the space $\mathbf{L}(g\cdot e)$, in the case when e is a process with independent increments, was given by Urbanik and Woyczyński [10] and Urbanik [9]. It turns out that $\mathbf{L}(g\cdot e)$ is a Musielak–Orlicz $\mathbf{L}_\phi$ space. Later, a number of authors studied the space $\mathbf{L}(g\cdot e)$ under various restrictions on e. It seems that the most general result was recently obtained by Kwapień and Woyczyński [4]. In the case when e is a left quasi-continuous semimartingale they proved that $\mathbf{L}(g\cdot e)$ is a randomized Musielak–Orlicz space $\mathbf{L}_{\phi(\omega)}$, where ϕ is explicitly expressed in terms of the Grigelionis characteristics of $g\cdot e$.

2. Proof of the Theorem

First we consider the case of $d = 1$. Note that since $A_t = x^2 w\cdot V_t$ and

$$\int_{[0,\infty[} (f\circ A)_t^{-1}\,|xw|_t^p\, dV_t \leqslant C^{p-1}\int_{[0,\infty[} (f\circ A)_t^{-1}\, dA_t$$

$$\leqslant C^{p-1}\int_0^\infty f_t^{-1}\,dt < \infty,$$

$(f\circ \Lambda)^{(-1)/p}\, xw \operatorname{sign}(g)\in \mathbf{L}(g\cdot e)$ and hence $(f\circ\Lambda)^{(-1)/p}|g|\in \mathbf{L}(xw\cdot e)$. Moreover, because $(f\circ\Lambda)^{1/p}\,|g|^{-1}$ is increasing and $\Lambda_{\infty -}=\infty$ we also have $(f\circ\Lambda)^{1/p}\,|g|^{-1}_{\infty -}=\infty$, which, using the Kronecker lemma, leads to

$$xw\cdot e_t = o((f\circ\Lambda)^{1/p}\,|g|_t^{-1}) \qquad \text{a.s.} \tag{2.1}$$

But $\theta_t-\theta=\Lambda_t^{-1}(xw\cdot e_t)$, and (2.1) finishes the proof for $d=1$.

Next take $d>1$ and note that it is sufficient to prove the theorem for $i=1$. Setting $\tilde{e}_t=w\cdot e_t$ and $\tilde{V}_t=w\cdot V_t$ we can write $\Lambda=xx^*\cdot\tilde{V}$ and for $t\geqslant t_0$, $\theta_t-\theta=\Lambda_t^{-1}(x\cdot\tilde{e}_t)$. Moreover, if we partition x and Λ as

$$x_t=\begin{pmatrix} x_t^1\\ T_t\end{pmatrix},\qquad \Lambda_t=\begin{pmatrix}(x^1)^2\cdot\tilde{V}_t & K_t\\ K_t^* & H_t\end{pmatrix},$$

where $H=TT^*\cdot\tilde{V}$ while $K=x^1T^*\cdot\tilde{V}$, then using Lemma 2 in [6], we can also write that for $t\geqslant t_0$, $\theta_t^1-\theta^1=u_t/v_t$, where u and $v=v^1$ satisfy

$$\begin{aligned} u_t&=u_{t_0}+\int_{]t_0,t]}\delta_s\,d\tilde{e}_s-\int_{]t_0,t]}\delta T^*H_-^{-1}(T\cdot\tilde{e}_-)_s\,d\tilde{V}_s,\\ v_t&=v_{t_0}+\int_{]t_0,t]}\delta^2\gamma_s\,d\tilde{V}_s\end{aligned}\tag{2.2}$$

with $\delta=x^1-KH^{-1}T$ and $\gamma=1+T^*H_-^{-1}T\,\Delta\tilde{V}$.

Now observe that to get conclusion (1.7) it is sufficient to show

$$u_t=o((f\circ v)^{1/p}\,|g|_t^{-1}) \qquad \text{a.s.} \tag{2.3}$$

But

$$\begin{aligned}&\int_{]t_0,t]}(f\circ v)_s^{-1}\,|\delta w|_s^p\,dV_s\\ &\quad\leqslant C\int_{]t_0,t]}|\delta|_s^p\,|\gamma|_s^{p/2}\,|x|_s^{2-p}\,(f\circ v)_s^{-1}\,d\tilde{V}_s\\ &\quad\leqslant C\left(\int_{]t_0,t]}\delta_s^2\gamma_s(f\circ v)_s^{-1}\,d\tilde{V}_s\right)^{p/2}\left(\int_{]t_0,t]}|x|_s^2\,(f\circ v)_s^{-1}\,d\tilde{V}_s\right)^{1-p/2}\\ &\quad\leqslant C\left(\int_{]t_0,t]}(f\circ v)_s^{-1}\,dv_s\right)^{p/2}\left(\int_{]t_0,t]}(f\circ v)_s^{-1}\,d\operatorname{tr}\Lambda_s\right)^{1-p/2},\end{aligned}$$

which implies that $(f\circ v)^{(-1)/p}\,\delta w\operatorname{sign}(g)\in\mathbf{L}^p(V)$. The same arguments as those used to establish (2.1) lead to

$$\int_{]t_0,t]}\delta w_s\,de_s=o((f\circ v)^{1/p}\,|g|_t^{-1}) \qquad \text{a.s.} \tag{2.4}$$

Moreover, since the second integral in (2.2) is equal to $\delta \cdot \eta_t$, where $\eta_t = 0$ on $[0, t_0]$ and

$$\eta_t = \int_{]t_0, t]} T^* H_-^{-1} (T \cdot \tilde{e}_-)_s \, d\tilde{V}_s, \qquad t > t_0, \tag{2.5}$$

the lemma below with x, A, c, and ϕ replaced by T, H, $(f \circ v)^{(-1)/p} |g| \delta$, and $(f \circ v)^{1/p}$, respectively, concludes that

$$\delta \cdot \eta_t = o((f \circ v)^{1/p} |g|_t^{-1}) \qquad \text{a.s.},$$

which together with (2.4) proves (2.3) and hence the statement.

LEMMA. *Let the notation and assumptions be as in the theorem and let c: $\mathbf{R}^+ \to \mathbf{R}$ be a Borel measurable locally bounded function. If there exists a strictly positive increasing function ϕ: $\mathbf{R}^+ \to \mathbf{R}^+$ such that*

$$\int_{]t_0, \infty[} c^2 g^{-2} \phi^{2-p} (1 + \operatorname{tr} A_-^{-1} \Delta A) w_t \, dV_t < \infty$$

and

$$\lim_{t \to \infty} \left(\int_{]t_0, t]} \phi_s^{-p} d \operatorname{tr} A_s \right)^{1 - p/2} < \infty,$$

then $c \in \mathbf{L}(\xi)$, where $\xi_t = 0$ on $[0, t_0]$ and

$$\xi_t = \int_{]t_0, t]} x^* A_-^{-1} (xw \cdot e)_- \, w_s \, dV_s, \qquad t > t_0. \tag{2.6}$$

Proof. Assume $d = 1$. It is easy to verify that the function h: $\mathbf{R}^+ \to \mathbf{R}^+$, where $h = 0$ on $[0, t_0[$ and $h_t = \int_{]t, \infty[} cg^{-1} x A_-^{-1} w_s \, dV_s$ on $[t_0, \infty[$ is well defined. We shall prove first that $hxw \in \mathbf{L}^p(V)$.

Since $|hxw|^p \cdot V_t \leqslant C^{p-1} (h^2 \phi^{2-p} \cdot A_t)^{p/2} (1_{[t_0, \infty[} \phi^{-p} \cdot A_t)^{1-p/2}$ it suffices to prove that $\phi^{2-p} \cdot (h^2 \cdot A)_{\infty -} < \infty$. But $h^2 \cdot A_t = h^2 A_t - A_- \cdot h_t^2 = h^2 A_t - 2A_- h \cdot h_t + A_- \Delta h \cdot h_t = h^2 A_t + 2hcg^{-1} xw \cdot V_t + 1_{[t_0, \infty[} c^2 g^{-2} x^2 A_-^{-1} w^2 \Delta V \cdot V_t$; therefore,

$$\begin{aligned}
\phi^{2-p} \cdot (h^2 \cdot A)_t &= \phi^{2-p} h^2 A_t - h^2 A_- \cdot \phi_t^{2-p} \\
&\quad + 2\phi^{2-p} hcg^{-1} xw \cdot V_t + \phi^{2-p} c^2 g^{-2} x^2 A_-^{-1} w \, \Delta V \cdot V_t \\
&\leqslant a_t + 2(1_{[t_0, \infty[} \phi^{2-p} c^2 g^{-2} w \cdot V_t)^{1/2} \\
&\quad \times (\phi^{2-p} h^2 \cdot A_t)^{1/2} + 1_{[t_0, \infty[} c^2 g^{-2} x^2 \\
&\quad \times A_-^{-1} A w \cdot V_t, \qquad \text{where} \quad a_t = \phi^{2-p} h^2 A_t - h^2 A_- \cdot \phi_t^{2-p}.
\end{aligned}$$

Moreover, since for all $t \geqslant t_0$, $h^2 \Lambda_t \leqslant \int_{]t,\infty[} c^2 g^{-2} \Lambda \Lambda_-^{-1} w_s\, dV_s$, we also have $\sup |a_t| = |a|^* < \infty$, which together with the above leads to the inequality $\phi^{2-p} \cdot (h^2 \cdot \Lambda)_t \leqslant |a|^* + 2I^{1/2}(\phi^{2-p} \cdot (h^2 \cdot \Lambda_t))^{1/2} + I$, where $I = \int_{]t_0,\infty[} c^2 g^{-2} x^2 \Lambda_-^{-1} \Lambda w_s\, dV_s < \infty$. This implies that $\phi^{2-p} \cdot (h^2 \cdot \Lambda)_t \leqslant I^{1/2} + (|a|^* + 2I)^{1/2}$ for all $t \geqslant t_0$ and hence also that $hxw \in \mathbf{L}^p(V)$.

Now consider h to be a function of c, i.e., $h = h(c)$. It is clear that if $h(c)\, xw \in \mathbf{L}^p(V)$ then $h(|c| \operatorname{sign}(g^{-1}x))\, xw \in \mathbf{L}^p(V)$ and consequently $g^{-1}h(cg)\, xw$ and $g^{-1}h(|c|\, g \operatorname{sign} x)$ also belong to $\mathbf{L}^p(V)$. But $\mathbf{L}^p(V) \subset \mathbf{L}(g \cdot e)$ and therefore the functions $h(cg)$ and $h(|c|\, g \operatorname{sign} x)$ belong to $\mathbf{L}(xw \cdot e)$. Moreover, integration by parts on the $[0, t]$ interval leads to

$$c \cdot \xi_t = h(cg) \cdot (xw \cdot e)_t - h(cg)_t\, (xw \cdot e_t)$$

and hence to prove that $c \in \mathbf{L}(\xi)$ it is enough to show that the second term converges. But $|h(cg)_t\, (xw \cdot e_t)| \leqslant h(|c|\, g \operatorname{sign} x)_t\, |xw \cdot e_t|$ and the right-hand side converges to zero by Kronecker's lemma.

Suppose now that the assertion holds for $1 \leqslant \dim x \leqslant d-1$. We shall prove that it then holds for $\dim x = d$. Direct computations or Lemma 2 in [6] implies that for $t \geqslant t_0$,

$$c \cdot \xi_t = c \cdot \eta_t + \int_{]t_0,t]} c\gamma\, \delta u_-\, v_-^{-1} w_s\, dV_s, \tag{2.7}$$

where η and ξ are given in (2.5) and (2.6), respectively. Moreover, since $\dim T = d-1$ and $T^* H_-^{-1} T \leqslant x^* \Lambda_-^{-1} x$ we obtain that $c \in \mathbf{L}(\eta)$ by the induction hypothesis. Therefore it remains to show that the second term in the right-hand side of (2.7) converges in $\mathbf{R}$ a.s. as $t \to \infty$. But setting $\hat{c} = c\gamma^{1/2}$ and $\hat{x} = \delta\gamma^{1/2}$ we get

$$\begin{aligned} \int_{]t_0,t]} & c\gamma\, \delta u_-\, v_-^{-1} w_s\, dV_s \\ &= u_{t_0} \int_{]t_0,t]} g\phi^{(p-2)/2} v_-^{-1} \hat{x}\hat{c} g^{-1} \phi^{(2-p)/2} w_s\, dV_s \\ &\quad + \int_{]t_0,t]} \hat{c}\hat{x} v_-^{-1} (1_{[t_0,\infty[}\, \gamma^{(-1)/2} \hat{x} \cdot e)_-\, w_s\, dV_s \\ &\quad + \int_{]t_0,t]} \hat{c}\hat{x} v_-^{-1} (1_{[t_0,\infty[}\, \gamma^{(-1)/2} \hat{x} \cdot \eta)_-\, w_s\, dV_s. \end{aligned} \tag{2.8}$$

Now using the Cauchy–Schwarz inequality we prove that the first integral converges in $\mathbf{R}$. Moreover, since

$$\begin{aligned} \int_{]t_0,\infty[} & \hat{c}^2 g^{-2} \phi^{2-p} u v_-^{-1} w_s\, dV_s \\ &= \int_{]t_0,\infty[} c^2 g^{-2} \phi^{2-p} (1 + \operatorname{tr} \Lambda_-^{-1}\, \Delta\Lambda) w_s\, dV_s < \infty, \end{aligned}$$

repeating arguments used in the first part of the proof we show that $\hat{h}\hat{x}w$ and hence $\hat{h}\,\delta w$ belongs to $\mathbf{L}^p(V)$. Here $\hat{h}=0$ on $[0, t_0[$ and $\hat{h}_t = \int_{]t,\infty[} \hat{c}g^{-1}\hat{x}v_-^{-1} w_s \, dV_s$ on $[t_0, \infty[$. This is sufficient to assure the a.s. convergence in $\mathbf{R}$ of the second and the third integrals in (2.8).

3. Linear Regression Models in Discrete Time

Consider the multiple regression model

$$z_j = \theta_1 x_{1j} + \cdots + \theta_d x_{dj} + \varepsilon_j; \qquad j = 1, 2, ..., \tag{3.1}$$

where x_{ij} are design constants and ε_j are random errors. Note that if we set $y_0 = 0$, $y_n = y_{n-1} + z_n$, $n \geqslant 1$, and $e_0 = 0$, $e_n = e_{n-1} + \varepsilon_n$, $n \geqslant 1$, then the above model can be written in the form

$$y_n = \theta^*(x_1 + \cdots + x_n) + e_n; \qquad n = 1, 2, ..., \tag{3.2}$$

where $x_n = (x_{1n}, ..., x_{dn})^*$. Furthermore, if we define $x_0 = 0$, $y_t = y_n$, $e_t = e_n$, $x_t = x_n$ for $n \leqslant t < n+1$, $n = 0, 1, ...$, then we can represent (3.2) in continuous time as (1.1), with V equal to the cadlag. distribution function of the measure $\sum_{n=1}^{\infty} \delta_n$, where δ_n is the Dirac measure at point n. It is also clear that the estimate (1.3) is given by

$$\theta_n = \left(\sum_{k=1}^{n} x_k x_k^* w_k \right)^{-1} \sum_{k=1}^{n} x_k w_k \, \Delta y_k, \qquad n \leqslant t < n+1, \tag{3.3}$$

and hence defining $X_n = [x_1, ..., x_n]$, $W_n = \operatorname{diag}(w_1, ..., w_n)$, and $Z_n = (z_1, ..., z_n)^*$, we can write that

$$\theta_n = (XWX^*)_n^{-1} XWZ_n, \qquad n \leqslant t < n+1. \tag{3.4}$$

Note that choosing $w_k = |x_k|^{(2-p)/(p-1)}$ in (3.3) or (3.4) we obtain $\theta_n^{\#}$. Now assume that $(\varepsilon_n : n \geqslant 1)$ is a martingale difference sequence such that for some $1 < p \leqslant 2$ and all $n = 1, 2, ...$, $E\,|\varepsilon_n|^p < \infty$ and define $g_n = (\max_{1 \leqslant k \leqslant n} E\,|\varepsilon_k|^p)^{(-1)/p}$. Since $E \sum_{k=1}^{n} |c_k g_k \varepsilon_k|^p \leqslant \sum_{k=1}^{n} |c_k|^p$ we deduce, using standard arguments, that $\mathbf{l}^p \subset \mathbf{L}(g \cdot e)$.

Finally, let $(\varepsilon_n, n \geqslant 1)$ be a sequence of independent random variables with characteristic functions $\exp(-\lambda_n^p |t|^p)$, $n \geqslant 1$. Note that in this case $E\,|\varepsilon_n|^p = \infty$, $n \geqslant 1$. Nevertheless, $\mathbf{l}^p \subset \mathbf{L}(g \cdot e)$, where $g_n = (\max_{1 \leqslant k \leqslant n} \lambda_k)^{-1}$ because

$$\begin{aligned} \mathbf{L}(g \cdot e) &= \mathbf{L}(g\lambda \cdot (\lambda^{-1} \cdot e)) \\ &= \left\{ c = (c_n : n \geqslant 1) : \sum_{n=1}^{\infty} |cg\lambda|_n^p < \infty \right\}. \end{aligned}$$

References

[1] Cambanis, S., Rosiński, J., and Woyczyński, W. A. (1985). Convergence of quadratic forms in p-stable variables and θ_p-radonifying operators. *Ann. Probab.* **13** 885–897.

[2] Chen, G. J., Lai, T. L., and Wei, C. Z. (1981). Convergence systems and strong consistency of least squares estimates in regression models. *J. Multivariate Anal.* **11** 319–333.

[3] Jacod, J. (1979). *Calcul Stochastique et Problèmes de Martingales.* Lecture Notes in Mathematics, Vol. 714, Springer-Verlag, New York/Berlin.

[4] Kwapień, S., and Woyczyński, W. A. (1986). *Semimartingale Integrals via Decoupling Inequalities and Tangent Processes.* Preprint 86-56. Department of Mathematics and Statistics, Case Western Reserve University.

[5] Lai, T. L., Robbins, H., and Wei, C. Z. (1979). Strong consistency of least squares estimates in multiple regression, II. *J. Multivariate Anal.* **9** 343–361.

[6] Le Breton, A., and Musiela, M. (1987). Strong consistency of least squares estimates in linear regression models driven by semimartingales. *J. Multivariate Anal.* **23** 77–92.

[7] Le Breton, A., and Musiela, M. (1986). Ordre de convergence d'estimateurs des moindres carrés en régression linéaire multiple. *C.R. Acad. Sci. Paris Ser. I Math.* **302**, No. 8 333–336.

[8] Mandelbrot, B. (1963). The variation of certain speculative prices. *J. Business* **36** 394–419.

[9] Urbanik, K. (1968). Random measures and harmonizable sequences. *Studia Math.* **31** 61–88.

[10] Urbanik, K., and Woyczyński, W. A. (1967). A random integral and Orlicz spaces. *Bull. Acad. Polon. Sci.* **15** 161–169.