Least-square estimators in linear regression models under negatively superadditive dependent random observations

Karine Bertin\(^a\), Soledad Torres\(^a\) and Lauri Viitasaar\(^b\)

\(^a\)CIMFAV-INGEMAT, Universidad de Valparaíso, Valparaíso, Chile; \(^b\)Department of Mathematics, Uppsala University, Uppsala, Sweden

ABSTRACT

In this article, we study the asymptotic behaviour of the least-square estimator in a linear regression model based on random observation instances. We provide mild assumptions on the moments and dependence structure on the randomly spaced observations and the residuals under which the estimator is strongly consistent. In particular, we consider observation instances that are negatively superadditive dependent within each other, while for the residuals we merely assume that they are generated by some continuous function. We complement our findings with a simulation study providing insights on finite sample properties.

1. Introduction

In this paper, we study the strong consistency of the least-square estimator in a simple regression model with a linear trend and continuous noise where the observation measurements are made at random times \(\tau_i\). More precisely, consider the simple regression model

\[ Y_{\tau_i} = a_{\tau_i} + \varepsilon_{\tau_i}, \quad 1 \leq i \leq N(T), \]

whose unknown parameter \(a\) must be estimated. Here \(\tau := \{\tau_i\}_{i \in \mathbb{N}}\) is a random increasing sequence of positive random variables and \(N(T) = \sum_{j \geq 1} 1_{\tau_j \leq T}\) determines the number of observations in \([0, T]\) with \(T > 0\). Random time observation models arise in many domains. In medicine, Lange et al. [1] study the disease process according to a latent continuous-time Markov chain with observation rates that depend on the individual’s underlying disease status. In finance, Aït-Sahalia and Mykland [2] consider random times in the case of continuous-time diffusions. In biomedical studies, the event of interest can occur more than once in a participant. These recurrent events can be thought as correlated random periods. The effect of the random sampling compared to discrete ones have been considered in [3]. They study the estimation of the mean and autocorrelation parameter of a stationary Gaussian process under two sampling schemes: sampling the continuous-time process at fixed equally spaced times and sampling at random times based on a renewal process. In non-parametric statistics, Masry [4] studies the problem of estimating an unknown...
probability density function, based on $n$ independent observations sampled at random times. Vilar [5] and Vilar and Vilar [6] investigate non-parametric regression estimation with randomly spaced observations associated with renewal processes. Strong consistency of the least-square estimator of the trend in model (1) is proved in [7,8]. They consider i.i.d. random times based either on jittered sampling or renewal processes, and increments of fractional Brownian motion as the noise.

In this paper, we extend the results of Araya et al. [7,8] by considering more general assumptions on the noise $\varepsilon$ and dependent observation measurements $\tau_i$ characterized by negative superadditive dependent (NSD) positive random variables. Additionally, only observations until a time $T > 0$ are considered, assuming, without loss of generality, $T = 1$.

NSD dependence for random variables have been first introduced by Hu [9] who extends the concept of negatively associated (NA) dependence (see [10]). NSD allows one to get many important inequalities and convergence results. Eghbal et al. [11] derive convergence results of sums of quadratic forms of NSD random variables under the assumption of the existence of moment of order $r > 1$. Shen et al. [12] study almost sure convergence and strong stability for weighted sums of NSD random variables. Shen et al. [13] investigate strong convergence for NSD random variables and present some moment inequalities. Wang et al. [14] study complete convergence for arrays of rowwise NSD random variables with applications to nonparametric regression.

In this work, we prove the strong consistency of the least-square estimator of the trend $a$ and derive the rate of convergence. For the noise $\varepsilon$, we merely assume that it arises from increments of a continuous function. We do not pose any requirements on the dependence structure between $\varepsilon$ and the random times $\tau_i$. On top of minimal requirements on $\varepsilon$, this fact provides additional flexibility to our model class. We illustrate the performances of our estimator in a simulation study where random times are given by the sum of NSD log-normal variables.

The paper is organized as follows. Section 2 presents the model, the assumptions on the random times, and our main results. Section 3 is devoted to the simulation study and Section 4 to the proofs.

### 2. Linear regression model with random time observations

We begin by introducing the underlying linear regression model and assumptions on the random observations required for the proofs of the main results, Theorems 2.6 and 2.7.

We assume that observations are made on random time instances $(\tau_i)_{i \in \mathbb{N}}$ determined by an increasing sequence of random variables $\tau_i$. In order to force observation instances $\tau_i$ into a fixed time interval $[0, T]$, we denote by $N(T)$ the number of observations needed to reach the fixed time $T$. That is, we set

$$N(T) = \arg \max_{i \in \mathbb{N}} \{ \tau_i : \tau_i \leq T \} = \sum_{j \geq 1} 1_{\tau_j \leq T}. \quad (2)$$

Without loss of generality and for the sake of simplicity, we let $T = 1$ and consider $N(1)$ as the number of observation time instances before time $T = 1$. The observation times $\tau_i$
are given by
\[ \tau_i = \sum_{j=1}^{i} t_j, \quad i \geq 1, \tag{3} \]
where now \( \{t_j\}_{j \in \mathbb{N}} \) is a sequence of positive random variables. With this notation, we consider the linear regression model
\[ Y_{\tau_i} = a\tau_i + \varepsilon_{\tau_i}, \quad i = 1, \ldots, N(1). \tag{4} \]
Here \( a \in \mathbb{R} \) is the unknown parameter to be estimated, and the residuals \( \varepsilon_{\tau_i} \) are given by
\[ \varepsilon_{\tau_i} = W_{\tau_i} - W_{\tau_{i-1}}, \]
where \( W \) is an arbitrary (almost surely) continuous function and we fix \( \tau_0 = 0 \). We stress that the only assumption on \( W \) we require is the continuity. Consequently, \( W \) can be chosen to be a path of any almost surely continuous stochastic process or even a deterministic continuous function.

We estimate the model parameter \( a \) by the classical least-square estimator (LSE) with a random number of observations \( N(1) \), given by
\[ \hat{a}_{N(1)} = \frac{\sum_{i=1}^{N(1)} Y_{\tau_i} \tau_i}{\sum_{i=1}^{N(1)} \tau_i^2}. \]
Clearly, we have
\[ \hat{a}_{N(1)} - a = \frac{\sum_{i=1}^{N(1)} \tau_i \left( W_{\tau_i} - W_{\tau_{i-1}} \right)}{\sum_{i=1}^{N(1)} \tau_i^2}. \tag{5} \]
In order to obtain strong consistency and the rate of convergence for the estimator (cf. Theorem 2.6), we need to pose assumptions on time instances \( t_i \) that ensure a sufficient amount of observations in the interval \([0, 1]\). We fix \( N \in \mathbb{N} \). As our first assumptions, we pose the following moment conditions.

\( (H_1) \quad \mathbb{E}(t_i) = \frac{1}{N} \) for all \( i \) (and \( N \)).

\( (H_2) \quad \) There exists \( r > 0 \) and a constant \( \tilde{C} > 0 \) such that, for all \( i \) and \( N \), we have \( \mathbb{E}(t_i^{2+r}) \leq \frac{\tilde{C}}{N^{2+r}} \).

Note that the number \( N \) describing the means of random variables \( t_i \) can be roughly understood as the sampling rate (average number of observations in \([0, 1]\)). In fact, in Lemma 4.3, we show that the number of observations \( N(1) \) on \([0, 1]\) satisfies \( N(1)/N \to 1 \) as \( N \) tends to \( \infty \). Note also that the observation times \( \tau_i \) as well as the generating random variables \( t_i \) depend on \( N \). However, for the sake of readability, we omit this dependence in the notation.

Moreover, we stress that assumptions \( (H_1) \) and \( (H_2) \) are rather mild. Condition \( (H_1) \) ensures, as we said previously, that \( N(1) \) is roughly proportional to \( N \) (see Lemma 4.3). Condition \( (H_2) \) requires slightly better than square integrability, and thus, one can consider even random times with relatively heavy tails. Moreover, the bound of \( (H_2) \) ensures that
random times $t_i$ are sufficiently concentrated around their mean $1/N$. This in turn implies that consecutive observation times $t_i$ and $t_{i+1}$ are never too far from each other.

In addition to the moment conditions $(H_1)$ and $(H_2)$ above, our third hypothesis is related to the dependence structure within the sequence $\{t_i\}_{i \in \mathbb{N}}$. For this, we need some preliminary definitions.

**Definition 2.1 ([15]):** A function $\phi : \mathbb{R}^m \mapsto \mathbb{R}$ is called superadditive, if for all $x, y \in \mathbb{R}^m$, we have

$$\phi(x \lor y) + \phi(x \land y) \geq \phi(x) + \phi(y),$$

where $\lor$ stands for componentwise maximum and $\land$ for componentwise minimum.

A characterization of smooth superadditive functions is given in the following lemma.

**Lemma 2.1 ([15]):** If $\phi$ has continuous second partial derivatives, then the superadditivity of $\phi$ is equivalent to $\frac{\partial^2 \phi}{\partial x_i \partial x_j} \geq 0$, $1 \leq i \neq j \leq m$.

Negatively superadditive dependent (NSD) variables are defined in the following way.

**Definition 2.2 ([9]):** A random vector $(X_1, X_2, \ldots, X_m)$ is said to be negatively superadditive dependent, if for any superadditive function $\phi$ such that the expectation $E\phi(X_1, X_2, \ldots, X_m)$ exists, we have

$$E\phi(X_1, X_2, \ldots, X_m) \leq E\phi(X_1^*, X_2^*, \ldots, X_m^*),$$

where $X_1^*, X_2^*, \ldots, X_m^*$ are independent and, for all $1 \leq i \leq m$, $X_i$ and $X_i^*$ are equally distributed.

Any vector consisting of independent random variables $X_i$ is trivially NSD. Many non-trivial examples of NSD variables with values in $\mathbb{R}$ or $\mathbb{R}^m$ are given in the literature. In particular, some classes of elliptical NSD variables are given in [9,16]. More specifically, variables with a density given by

$$g(x) = |\Sigma|^{-1/2} g \left(x^t \Sigma x\right), \quad x \in \mathbb{R}^m,$$

are NSD, provided that $\Sigma = (\sigma_{ij})$ satisfies $\sigma_{ij} < 0$ for $i \neq j$ and $\int_0^\infty g(t)t^{m/2-1} \, dt < \infty$.

As a consequence, a random vector that follows a multivariate normal distribution with covariance matrix $\Sigma = (\sigma_{ij})$ satisfying $\sigma_{ij} < 0$ for $i \neq j$ is NSD. Our third and final assumption on the random times $t_i$ is the following.

$(H_3)$ For all $m \in \mathbb{N}$ (and $N$), the vector $(t_1, \ldots, t_m)$ is NSD.

**Remark 2.1:** Note that $(H_1)$ and $(H_3)$ imply that, for any $i \neq j$,

$$E(t_it_j) \leq \frac{1}{N^2}.$$
Hölder inequality gives us
\[ \mathbb{E}(t_i^2) \leq \tilde{C}^{2/(2+r)} \frac{n^2}{N^2}. \]

**Remark 2.2:** Many non-trivial examples of NSD vectors presented in the literature take values in \( \mathbb{R}^m \) while for our purposes the random times \( t_i \) are assumed to be non-negative. A simple approach to construct non-negative NSD vectors, suitable for our purposes, is to choose an \( \mathbb{R}^m \)-valued NSD vector \((X_1, \ldots, X_m)\) and a non-decreasing function \( \phi : \mathbb{R} \to \mathbb{R}_+ \). Then by a routine exercise, one can show that \((\phi(X_1), \ldots, \phi(X_m))\) is non-negative and NSD. Indeed, with Lemma 2.2, one can show that if \( \phi : \mathbb{R}^m \to \mathbb{R} \) is superadditive and \( \phi : \mathbb{R} \to \mathbb{R}_+ \) is non-decreasing, then the function \( \tilde{\phi} : \mathbb{R}^m \to \mathbb{R} \) defined by

\[ \tilde{\phi}(x_1, \ldots, x_m) = \phi(\phi(x_1), \ldots, \phi(x_m)) \]  

is superadditive. As a particular example, a vector \((Y_1, \ldots, Y_m)\) of multivariate log-normal variables \( Y_i \) with parameter \( \Sigma = (\sigma_{ij}) \) satisfying \( \sigma_{ij} < 0 \) for \( i \neq j \) is NSD. Indeed, in this case, we have \( Y_i = \exp(X_i) \) for all \( 1 \leq i \leq m \), where \((X_1, \ldots, X_m)\) is a multivariate normal NSD vector. This example is used in our simulation study, provided in Section 3.

We are now ready to state the main results of this paper.

**Theorem 2.1:** Suppose that the sequence of random times \( \{t_j\}_{j \in \mathbb{N}} \) satisfies \((H_1)\) to \((H_3)\) and that \( W \) is continuous almost surely. Then we have, almost surely as \( N \to \infty \),

\[ N(1) \left( \hat{a}_{N(1)} - a \right) \to 3 \int_0^1 (W_1 - W_s) \, ds, \]

where \( N(1) \) is defined in (2).

If for some particular reason the observation window does not have to be restricted to the time interval \([0, 1]\), then one can use all observations in the estimation and consider an estimator

\[ \hat{a}_N = \frac{\sum_{i=1}^N Y_{t_i} \tau_i}{\sum_{i=1}^N \tau_i^2}. \]

In this case, the result can be formulated similarly as in Theorem 2.6.

**Theorem 2.2:** Suppose that the sequence of random times \( \{t_j\}_{j \in \mathbb{N}} \) satisfies hypothesis \((H_1)\) to \((H_3)\) and that \( W \) is continuous almost surely. Then, we have, almost surely as \( N \to \infty \),

\[ N \left( \hat{a}_N - a \right) \to 3 \int_0^1 (W_1 - W_s) \, ds. \]

**Remark 2.3:** Note that the classical fixed design model is included in our framework since in this case the variables \( t_i \) are constant variables equal to \( \frac{1}{N} \) and they satisfy hypotheses \((H_1)\), \((H_2)\) and \((H_3)\). Another type of random time sampling, called jittered sampling, is studied in [5,8], among others. Jittered sampling corresponds to the case where the sampling is nearly regular in the sense that the fluctuations from the mean are small compared
to the sampling interval. More precisely, in the jittered case, the random times \( \{\tau_j\}_{j \in \mathbb{N}} \) are given by
\[
\tau_j = \frac{j}{N} + \frac{X_j}{N}, \quad j = 1, \ldots, N - 1,
\]
where for each \( j = 1, 2, \ldots, N - 1 \) the random variable \( X_j \) is supported on \((-\frac{1}{2}, \frac{1}{2})\) and has zero expectation. In order to extend our results to cover this case as well, a careful examination of our proofs below reveals that the statements of Theorems 2.6 and 2.7 are valid, for any continuous \( W \), provided that, as \( N \to \infty \), we have

- \( \max_{1 \leq k \leq 2N} |\tau_k - \tau_{k-1}| \to 0 \),
- \( \tau_N \to 1 \) and \( \tau_{N(1)} \to 1 \),
- \( \frac{N(1)}{N} \to 1 \), and
- \( \frac{1}{N} \sum_{i=1}^{N} \tau_i^2 \to 1/3 \).

It is straightforward to check that all these conditions are satisfied in the jittered case as well regardless of the dependence structure within the random variables \( (X_j)_{j \in \mathbb{N}} \), complementing our work by covering the jittered case as well.

**Remark 2.4:** Note that, by following the techniques used in this paper, our results could be extended to cover several generalizations of our model. For example, one could add an intercept parameter to (4) leading to a model
\[
Y_{\tau_i} = \alpha + a \tau_i + \epsilon_{\tau_i}, \quad i = 1, \ldots, N(1).
\]

Another interesting generalizations would be time series models such as an AR(1) model
\[
Y_{\tau_{i+1}} = \theta Y_{\tau_i} + \epsilon_{\tau_i}, \quad i = 1, \ldots, N(1),
\]

or non-parametric regression model given by
\[
Y_{\tau_i} = r(x) + \epsilon_{\tau_i}, \quad i = 1, \ldots, N(1).
\]

In addition, one could easily extend our results to cover multivariate models with suitable correlation structures within the components as well.

### 3. Simulation

In this section, we illustrate the performance of the estimator \( \hat{a}_{N(1)} \) defined in (5). For process \( W \), we take the fractional Brownian motion and the variables \( t_j \) follow an NSD log-normal distribution.

**The fractional Brownian motion as the noise:** We take \( W = B^H \), where \( B^H \) is a fractional Brownian motion with Hurst parameter \( H \in \{0.1, 0.5, 0.9\} \). This process is one of the most popular Gaussian stochastic processes with memory and the case \( H = 0.5 \) corresponds to independent residuals. We recall the main properties of the stochastic process \( B^H \):  
- \( B^H \) is a Gaussian self-similar stochastic process with \( B^H_0 = 0 \).
• The covariance of $B^H$ is given by
  
  $R_H(t,s) = \mathbb{E}(B^H_t B^H_s) = \frac{\sigma^2}{2}(t^{2H} + s^{2H} - |t - s|^{2H})$.

In this case, the variable $\int_0^1 (W_1 - W_s) ds$ that appears in Theorems 2.6 and 2.7 is normally distributed with mean 0 and variance $\sigma_H = 9/(2H + 2)$.

**Distribution of the random times:** In the following, we specify how we obtain $t_1, \ldots, t_N$ and $\tau_1, \ldots, \tau_N$. Since $N(1)$ is obtained by (2), we have to generate a number $N_0 \in \mathbb{N}$ of variables $t_j$ and $\tau_j$ large enough to ensure $N(1)$ observations. In fact, in our simulation study $N(1)$ will be obtained by

$$N(1) = \sum_{j=1}^{N_0} 1_{\tau_j \leq 1}.$$

We simulate multivariate normal variables $(X_1, \ldots, X_{N_0})$ with mean $\mu = (\mu_1, \ldots, \mu_{N_0})$ and covariance matrix $\Sigma = (\sigma_{i,j})_{i,j=1,\ldots,N_0}$. The mean and covariance are given, for $i, j \in \{1, \ldots, N_0\}$, by $\mu_i = -\log(N) - 1/2$, $\sigma_{i,i} = 1$, and $\sigma_{i,j} = -1/4$, $j \neq i$. For $i = 1, \ldots, N_0$, we define the variables $t_i = \exp X_i$. These variables are log-normal with mean $1/N$ (see $(H_1)$). It can be easily checked also that the variables $t_i$ satisfy assumptions $(H_2)$ and $(H_3)$. The random times $\tau_1, \ldots, \tau_{N(1)}$ are then obtained by (2) and (3).

Finally, for the model parameter, we set $a = 1$. Realizations of the data $Y_\tau$ for $N = 1000$ and $H \in \{0.1, 0.5, 0.9\}$ are plotted in Figure 1. Due to the Hölder regularity of the fractional Brownian motion $B^H$, the data are more noisy for lower values of $H$.

We consider three possible values for $N$: $N = 100$, $N = 1000$, and $N = 5000$. In order to assess the finite sample properties of the estimator, we conduct a Monte Carlo simulation study with a number of replications $M = 1000$.

The absolute mean bias over the $M = 1000$ replications, defined for the $k$:th replication by

$$\left| \frac{1}{M} \sum_{k=1}^{M} \hat{a}_{N(1)}^{(k)} - a \right|,$$

for different values of $N$ and $H$ is presented in Table 1. As expected, the bias decreases as $N$ increases for all values of $H$.

**Figure 1.** Regression data $Y_\tau$ for $N = 1000$ and $H \in \{0.1, 0.5, 0.9\}$.
Table 1. Absolute mean bias of the estimator for different values of $N$ and $H$.

| $H$   | $N = 100$ | $N = 1000$ | $N = 5000$ |
|-------|-----------|------------|------------|
| 0.1   | $6.92 \cdot 10^{-4}$ | $7.44 \cdot 10^{-4}$ | $1.69 \cdot 10^{-4}$ |
| 0.5   | $6.15 \cdot 10^{-5}$ | $4.70 \cdot 10^{-6}$ | $8.31 \cdot 10^{-6}$ |
| 0.9   | $9.66 \cdot 10^{-6}$ | $3.97 \cdot 10^{-6}$ | $3.11 \cdot 10^{-6}$ |

Table 2. $L_1$ risk of the estimator for different values of $N$ and $H$.

| $H$   | $N = 100$ | $N = 1000$ | $N = 5000$ |
|-------|-----------|------------|------------|
| 0.1   | $1.65 \cdot 10^{-2}$ | $1.47 \cdot 10^{-2}$ | $1.25 \cdot 10^{-2}$ |
| 0.5   | $1.68 \cdot 10^{-3}$ | $1.39 \cdot 10^{-3}$ | $1.25 \cdot 10^{-3}$ |
| 0.9   | $3.22 \cdot 10^{-4}$ | $2.85 \cdot 10^{-4}$ | $2.39 \cdot 10^{-4}$ |

Table 3. Asymptotic variance of $N(1)(\hat{a}_{N(1)} - a)$ for different values of $N$ and $H$ over the $M = 1000$ replications.

| $H$ | Theoretical | $N = 100$ | $N = 1000$ | $N = 5000$ |
|-----|-------------|----------|-----------|-----------|
| 0.1 | 4.09        | 4.221    | 4.359     | 4.09      |
| 0.5 | 3           | 3.250    | 3.027     | 3.13      |
| 0.9 | 2.37        | 2.352    | 2.372     | 2.19      |

The $L_1$ risk of the estimator over the $M = 1000$ replications, defined for the $k$:th replication by

$$\frac{1}{M} \sum_{k=1}^{M} \left| \hat{a}_{N(1)}^{(k)} - a \right|,$$

is presented in Table 2. The estimated variance of $N(1)(\hat{a}_{N(1)} - a)$, computed in a standard way by

$$\frac{1}{M} \sum_{k=1}^{M} \left( N(1)^{(k)} \left[ \hat{a}_{N(1)}^{(k)} - a \right] \right)^2,$$

where $N(1)^{(k)}$ is the number of observations needed to reach the time 1 in the $k$:th replication, is presented in Table 3. It can be seen that, as expected, the estimated variance is close to the theoretical variance $9/(2H + 2)$.

Figure 2 shows the plots of the histograms for $N(1)(\hat{a}_{N(1)} - a)$ with values $N \in \{100, 1000, 5000\}$ and $H = 0.1$, jointly with the density of the normal distribution with mean 0 and variance $9/(2H + 2)$ corresponding to the asymptotic distribution (cf. Theorem 2.6). By plotting histograms for other values of $H$, one obtains similar figures and the asymptotic normality with mean 0 and variance $9/(2H + 2)$. Finally, Figure 3 shows the ratio between the numbers $N(1)$ and $N$. It is clearly visible in Figure 3 that as $N$ increases, the ratio $N(1)/N$ is more concentrated around 1. This is in line with Theorems 2.6 and 2.7 (see also Lemma 4.3). Again we have only plotted the case $H = 0.1$, while similar phenomena can be observed for other values of $H$ as well.
4. Proofs

This section is devoted to the proofs of our main results, Theorems 2.6 and 2.7. We begin by presenting some auxiliary results, while the proofs of the main results are presented in Section 4.2. Throughout the proofs, we denote by $C$ a generic unimportant constant which may vary from line to line.

4.1. Auxiliary results

Lemma 4.1: Suppose that the sequence $\{t_j\}_{j \in \mathbb{N}}$ satisfies hypothesis (H2). Then we have $\max_{1 \leq j \leq N} t_j \to 0$ almost surely as $N \to \infty$.

Proof: Let $\delta > 0$ be fixed and denote

$$S_N(\delta) = \sum_{j=1}^{N} 1_{t_j \geq \delta}.$$
Then, we have
\[
P\left( \max_{1 \leq j \leq N} t_j > \delta \right) \leq P(S_N(\delta) \geq 1)
\]
\[
\leq \mathbb{E}(S_N(\delta)) = \sum_{j=1}^{N} P(t_j \geq \delta)
\]
\[
\leq \delta^{-2-r} \sum_{j=1}^{N} \mathbb{E}(t_j^{2+r}).
\]
Hence, by \((H_2)\), we obtain
\[
P\left( \max_{1 \leq j \leq N} t_j > \delta \right) \leq \delta^{-2-r} \tilde{C} N^{-1-r}
\]
and consequently,
\[
\sum_{N=1}^{\infty} P\left( \max_{1 \leq j \leq N} t_j > \delta \right) < \infty.
\]
The claim follows from the Borel–Cantelli lemma.

Note that if the sequence \(\{t_j\}_{j \in \mathbb{N}}\) is NSD, then random series based on \(t_j\) behave essentially as random series based on the independent \(t_j^*\)'s that satisfy Definition 2.3. In particular, taking \(X_j = t_j - \mathbb{E}(t_j)\), we have the Rosenthal inequality (see [13]) for any \(p > 2\)
\[
\mathbb{E}\left( \left( \max_{1 \leq k \leq n} \left| \sum_{j=1}^{k} X_j \right| \right)^p \right) \leq 2 \left( \frac{15p}{\log p} \right)^p \left[ \sum_{j=1}^{n} \mathbb{E}|X_j|^p + \left( \sum_{j=1}^{n} \mathbb{E}(X_j^2) \right)^{p/2} \right],
\]
provided that \(p > 2\) and \(\mathbb{E}(t_j^p) < \infty\). Moreover, for \(p = 2\), we have
\[
\mathbb{E}\left( \left( \sum_{j=1}^{n} X_j \right)^2 \right) \leq \sum_{j=1}^{n} \mathbb{E}(X_j^2).
\]
This follows directly from the fact that
\[
(x_1, \ldots, x_n) \mapsto \left[ \sum_{j=1}^{n} \left[ x_j - \mathbb{E}t_j \right]^2 \right]
\]
is superadditive, which can be seen by taking partial derivatives and using Lemma 2.2. We also use the following proposition in the proofs of our main results. While the result could be obtained by verifying several conditions of Theorem 3.2 in [13], here we present a more concise proof for the reader’s convenience.
Proposition 4.1: Suppose that the sequence \( \{t_j\}_{j \in \mathbb{N}} \) satisfies \((H_1) - (H_3)\) and that \( g : \mathbb{N} \rightarrow \mathbb{N} \) is a function such that, as \( N \rightarrow \infty \), we have

\[
\frac{g(N)}{N} \rightarrow L
\]

for some \( L > 0 \). Then \( \sum_{j=1}^{g(N)} [t_j - \mathbb{E}t_j] \) converges almost surely to 0 as \( N \rightarrow \infty \) and consequently,

\[
\sum_{j=1}^{g(N)} t_j \rightarrow L
\]

almost surely.

Proof: We have

\[
\mathbb{E}((t_j - \mathbb{E}t_j)^2) \leq \frac{\hat{C}^2/(2+r)}{N^2}
\]

and

\[
\mathbb{E}((t_j - \mathbb{E}t_j)^{2+r}) \leq 2^{-1} (\mathbb{E}(t_j)^{2+r} + (\mathbb{E}(t_j))^{2+r}) \leq CN^{-2-r}.
\]

Using (7) with \( p = 2 + r \), we have

\[
\mathbb{E}\left(\left|\sum_{j=1}^{g(N)} t_j - \mathbb{E}t_j\right|^{2+r}\right) \leq C \left( \frac{g(N)N^{-2-r} + \left(\frac{g(N)N^{-2}}{N}\right)^{1+r/2}}{N^{-1-r/2}} \right). \tag{9}
\]

Using \( g(N)/N \rightarrow L \), we obtain that, for all \( \varepsilon > 0 \),

\[
\sum_{N=1}^{\infty} \mathbb{P}\left(\left|\sum_{j=1}^{g(N)} t_j - \mathbb{E}t_j\right| \geq \varepsilon \right) \leq \frac{1}{\varepsilon^{2+r}} \sum_{N=1}^{\infty} \left( \frac{g(N)}{N} \right)^{N^{-r-1}} \left( \frac{g(N)}{N} \right)^{1+r/2} N^{-r-1-r/2} < \infty. \tag{10}
\]

This together with the Borel-Cantelli lemma implies that \( \sum_{j=1}^{g(N)} (t_j - \mathbb{E}t_j) \) converges to 0 almost surely. Finally, \( \sum_{j=1}^{g(N)} t_j \rightarrow L \) follows from the fact \( \sum_{j=1}^{g(N)} \mathbb{E}t_j = g(N)/N \). \( \blacksquare \)

With Proposition 4.2, we can deduce the following two lemmas.

Lemma 4.2: Suppose that the sequence \( \{t_j\}_{j \in \mathbb{N}} \) satisfies hypotheses \((H_1) - (H_3)\). Then, as \( N \rightarrow \infty \), \( \tau_N \rightarrow 1 \) almost surely.

Proof: By the very definition, we have

\[
\tau_N = \sum_{j=1}^{N} t_j
\]

so it remains to apply Proposition 4.2 with \( g(N) = N \). \( \blacksquare \)
**Lemma 4.3:** Suppose that the sequence \( \{ t_j \}_{j \in \mathbb{N}} \) satisfies \((H_1)-(H_3)\) and let \( N(1) \) be given by (2). Then, as \( N \to \infty \), we have \( \frac{N(1)}{N} \to 1 \) and \( |\tau_{N(1)} - \tau_N| \to 0 \) almost surely.

**Proof:** For a fixed \( \varepsilon > 0 \), define sets \( A_+(\varepsilon) \) and \( A_-(\varepsilon) \) by

\[
A_+(\varepsilon) = \left\{ \omega : \limsup_{N \to \infty} \frac{N(1)}{N} \geq 1 + \varepsilon \right\}
\]

and

\[
A_-(\varepsilon) = \left\{ \omega : \liminf_{N \to \infty} \frac{N(1)}{N} \leq 1 - \varepsilon \right\}.
\]

If \( \omega \in A_+(\varepsilon) \), then we can find arbitrary large numbers \( N \) such that \( N(1) \geq (1 + \varepsilon)N \) leading to

\[
\tau_{N(1)} = \sum_{j=1}^{N(1)} t_j \geq \sum_{j=1}^{\lfloor (1+\varepsilon)N \rfloor} t_j.
\]

On the other hand, we obtain by Proposition 4.2 as \( N \to \infty \) that

\[
\sum_{j=1}^{\lfloor (1+\varepsilon)N \rfloor} t_j \to 1 + \varepsilon.
\]

Since \( \tau_{N(1)} \leq 1 \) by the very definition of \( N(1) = \max_{j \in \mathbb{N}} \{ t_j : t_j \leq 1 \} \), it follows that \( \mathbb{P}(A_+(\varepsilon)) = 0 \). Similarly, for \( \omega \in A_-(\varepsilon) \), we can find arbitrary large \( N \) such that \( N(1) \leq (1 - \varepsilon)N \) and get

\[
\tau_{N(1)+1} \leq \sum_{j=1}^{\lfloor (1-\varepsilon)N \rfloor + 1} t_j.
\]

We observe again by Proposition 4.2 as \( N \to \infty \),

\[
\sum_{j=1}^{\lfloor (1-\varepsilon)N \rfloor + 1} t_j \to 1 - \varepsilon.
\]

On the other hand, \( \tau_{N(1)+1} \geq 1 \) and thus \( \mathbb{P}(A_-(\varepsilon)) = 0 \) as well. It follows that for any \( \varepsilon > 0 \) we have, almost surely,

\[
1 - \varepsilon < \liminf_{N \to \infty} \frac{N(1)}{N} \leq \limsup_{N \to \infty} \frac{N(1)}{N} < 1 + \varepsilon.
\]

Since \( \varepsilon > 0 \) is arbitrary, it follows that \( \lim_{N \to \infty} \frac{N(1)}{N} = 1 \) almost surely. For the second claim, by the very definition of \( N(1) \) we get

\[
0 \leq 1 - \tau_{N(1)} \leq \tau_{N(1)+1} - \tau_N = \tau_{N(1)+1} - \tau_{N(1)}
\]

leading to

\[
0 \leq 1 - \tau_{N(1)} \leq \tau_{N(1)+1}.
\]
Now since \( \frac{N(1)+1}{N} \to 1 \) almost surely, it follows that there exists \( N_0 = N_0(\omega) \) such that \( N(1) + 1 \leq 2N \) for all \( N \geq N_0 \). This gives us
\[
t_{N(1)+1} \leq \max_{1 \leq j \leq 2N} t_j
\]
which converges to zero by the arguments in the proof of Lemma 4.1. Thus
\[
|1 - \tau_{N(1)}| \to 0
\]
from where \( |\tau - \tau_{N(1)}| \to 0 \) follows by Lemma 4.2. This completes the proof. \( \blacksquare \)

We end this subsection with the following proposition on the denominator in (5).

**Proposition 4.2:** Suppose \( \{t_j\}_{j \in \mathbb{N}} \) satisfies \((H_1)\) to \((H_3)\). Then, almost surely as \( N \to \infty \),
\[
\frac{1}{N} \sum_{k=1}^{N} \tau_k^2 \to \frac{1}{3}.
\]

**Proof:** In order to prove the claim, we first observe that it suffices to show that
\[
\frac{1}{N} \sum_{k=1}^{N} (\tau_k - E\tau_k)^2 \to 0,
\]
almost surely. Indeed, we may write
\[
\frac{1}{N} \sum_{k=1}^{N} \tau_k^2 = \frac{1}{N} \sum_{k=1}^{N} (\tau_k - E\tau_k)^2 + 2\frac{1}{N} \sum_{k=1}^{N} (\tau_k - E\tau_k) E\tau_k + \frac{1}{N} \sum_{k=1}^{N} (E\tau_k)^2.
\]
Here
\[
\frac{1}{N} \sum_{k=1}^{N} (E\tau_k)^2 = \frac{1}{N} \sum_{k=1}^{N} \binom{k}{N}^2 = \frac{N(N+1)(2N+1)}{6N^3} \to \frac{1}{3}
\]
as \( N \) tends to infinity and, by Cauchy–Schwartz inequality,
\[
\left| \frac{1}{N} \sum_{k=1}^{N} (\tau_k - E\tau_k) E\tau_k \right| \leq \sqrt{\frac{1}{N} \sum_{k=1}^{N} (\tau_k - E\tau_k)^2} \sqrt{\frac{1}{N} \sum_{k=1}^{N} (E\tau_k)^2} \to 0
\]
that converges to zero once (11) is proved. Denote \( \hat{\tau}_j = \tau_j - \mathbb{E}\tau_j \) and \( \hat{t}_j = t_j - \mathbb{E}t_j \), and let \( \epsilon > 0 \) be fixed. Markov inequality gives us

\[
\mathbb{P}\left( \frac{1}{N} \sum_{k=1}^{N} \hat{\tau}_k^2 > \epsilon \right) \leq \frac{\mathbb{E}\left( \sum_{k=1}^{N} \hat{\tau}_k^2 \right)^{\frac{2+r}{2}}}{(\epsilon N)^{\frac{2+r}{2}}}. \tag{12}
\]

Furthermore, by Minkowski’s integral inequality, we obtain

\[
\mathbb{E}\left( \sum_{k=1}^{N} \hat{\tau}_k^{2+r} \right)^{\frac{2}{2+r}} \leq \left[ \sum_{k=1}^{N} \left( \mathbb{E}|\hat{\tau}_k|^{2+r} \right)^{\frac{2}{2+r}} \right]^{\frac{2+r}{2}}. \tag{13}
\]

By applying \((H_2)\) and Rosenthal inequality (7), we can deduce

\[
\mathbb{E}|\hat{\tau}_k|^{2+r} \leq C \sum_{j=1}^{N} \mathbb{E}|\hat{\tau}_j|^{2+r} + C \left( \sum_{j=1}^{N} \mathbb{E}\hat{\tau}_j^2 \right)^{\frac{2+r}{2}} \leq CN^{-1-r} \leq CN^{-\frac{2+r}{2}} \leq CN^{-1-r/2}.
\]

Plugging into (13) leads to

\[
\mathbb{E}\left( \sum_{k=1}^{N} \hat{\tau}_k^{2} \right)^{\frac{2+r}{2}} \leq CN^{-1-r/2}N^{\frac{2+r}{2}} \leq C.
\]

In view of (12), we get

\[
\sum_{N=1}^{\infty} \mathbb{P}\left( \frac{1}{N} \sum_{k=1}^{N} \hat{\tau}_k^2 > \epsilon \right) \leq C \sum_{N=1}^{\infty} (\epsilon N)^{-1-r/2} < \infty,
\]

and hence (11) follows from the Borel–Cantelli lemma. This completes the proof. \[\Box\]

### 4.2. Proofs of Theorems 2.6 and 2.7

**Proof:** First, we will prove that

\[
\frac{1}{N(1)} \sum_{j=1}^{N(1)} \tau_j^2 \rightarrow \frac{1}{3}. \tag{14}
\]

With the convention \( \sum_{j=1}^{0} a_j = 0 \), we can write

\[
\sum_{j=1}^{N(1)} \tau_j^2 = \sum_{j=1}^{N} \tau_j^2 + \sum_{j=N+1}^{N(1)} \tau_j^2
\]
for \( N(1) \geq N \) and
\[
\sum_{j=1}^{N(1)} \tau_j^2 = \sum_{j=1}^{N} \tau_j^2 - \sum_{j=N(1)+1}^{N} \tau_j^2
\]
for \( N(1) < N \). Since \( \tau_{N(1)} \leq 1 \), here
\[
\frac{1}{N(1)} \sum_{j=N(1)+1}^{N} \tau_j^2 \leq \frac{|N - N(1)|}{N(1)} \to 0
\]
by Lemma 4.3. Similarly,
\[
\frac{1}{N(1)} \sum_{j=1}^{N} \tau_j^2 \leq \frac{\tau_N}{N} \frac{|N - N(1)|}{N(1)} \to 0
\]
by Lemmas 4.2 and 4.3. Finally, we may apply Lemma 4.3 and Proposition 4.5 to obtain
\[
\frac{1}{N(1)} \sum_{j=1}^{N(1)} \tau_j^2 = \frac{N}{N(1)} \frac{1}{N} \sum_{j=1}^{N} \tau_j^2 \to \frac{1}{3}.
\] (15)
Second, we will prove that
\[
\sum_{j=1}^{N(1)} \tau_j \left( W_{\tau_j} - W_{\tau_{j-1}} \right) \to \int_{0}^{1} (W_1 - W_t) \, dt.
\] (16)
Using summation by parts, we obtain
\[
\sum_{j=1}^{N(1)} \tau_j \left( W_{\tau_j} - W_{\tau_{j-1}} \right) = \tau_{N(1)} W_{\tau_{N(1)}} - \sum_{j=1}^{N(1)} W_{\tau_{j-1}} (\tau_j - \tau_{j-1}).
\]
Since \( W \) is almost surely continuous, Lemmas 4.2 and 4.3 implies that \( \tau_{N(1)} W_{\tau_{N(1)}} \to W_1 = \int_{0}^{1} W_1 \, dt \). We have
\[
\sum_{j=1}^{N(1)} W_{\tau_{j-1}} (\tau_j - \tau_{j-1}) = \sum_{j=1}^{N(1)+1} W_{s_{j-1}} (s_j - s_{j-1}) - W_{s_{N(1)}} (1 - \tau_{N(1)}).
\]
Since \( W \) is bounded almost surely as a continuous function, \( W_{s_{N(1)}} (1 - \tau_{N(1)}) \) converges almost surely to 0.

Let us introduce the sequence \( s_j \) by \( s_j = \tau_j \) for \( j = 1, \ldots, N(1) \) and \( s_{N(1)+1} = 1 \). This sequence satisfies
\[
\max_{1 \leq j \leq N(1)+1} (s_j - s_{j-1}) \leq |1 - \tau_{N(1)}| + \max_{1 \leq j \leq N(1)} t_j
\]
Now, since \( \frac{N(1)}{N} \to 1 \) we obtain that, in particular, for every \( \omega \) there exists \( N_0(\omega) \) such that for all \( N \geq N_0(\omega) \) we have \( \frac{N}{2} \leq N(1) \leq 2N \). Following the same lines as Lemma 4.1, we
deduce that
\[ \max_{1 \leq j \leq 2N} t_j \to 0 \]
which implies that
\[ \max_{1 \leq j \leq N(1) + 1} (s_j - s_{j-1}) \to 0. \]

Finally, by continuity of \( W \) the Riemann integral \( \int_0^1 W_t \, dt \) exists, and we have
\[
\sum_{j=1}^{N(1)+1} W_{s_j - s_{j-1}} \to \int_0^1 W_t \, dt
\]
almost surely. Hence, we obtain (16) and the theorem is a consequence of (14) and (16).

**Proof:** By Proposition 4.5, it suffices to show
\[
\sum_{j=1}^N \tau_j \left( W_{\tau_j} - W_{\tau_{j-1}} \right) \to \int_0^1 (W_1 - W_t) \, dt
\]
almost surely as \( N \to \infty \). Two cases have to be considered: \( \tau_N < 1 \) and \( \tau_N \geq 1 \). When \( \tau_N < 1 \), we can mimic the proof of Theorem 2.6 in order to derive (17).

When \( \tau_N \geq 1 \), we write
\[
\sum_{j=1}^N W_{\tau_j - \tau_{j-1}} = \sum_{j=1}^{N(1)} W_{\tau_j - \tau_{j-1}} + \sum_{j=N(1)+1}^N W_{\tau_j - \tau_{j-1}}.
\]

We prove in the proof of Theorem 2.6 that \( \sum_{j=1}^{N(1)} W_{\tau_j - \tau_{j-1}} \) converges almost surely to \( \int_0^1 (W_1 - W_t) \, dt \). For the second sum, we obtain
\[
\left| \sum_{j=N(1)+1}^N W_{\tau_j - \tau_{j-1}} \right| \leq \sup_{0 \leq s \leq \tau_N} |W_s| |\tau_N - \tau_{N(1)}|
\]
which converges to zero by boundedness of \( W \) and Lemma 4.3. This allows us to deduce (17).

**Disclosure statement**

No potential conflict of interest was reported by the author(s).

**Funding**

Karine Bertin and Soledad Torres have been supported by FONDECYT [grants 1171335 and 1190801] and Mathamsud-ANID [MATH 190002].
References

[1] Lange JM, Hubbard RA, Inoue LY T, et al. A joint model for multistate disease processes and random informative observation times, with applications to electronic medical records data. Biometrics. 2015;71(1):90–101.

[2] Aït-Sahalia Y, Mykland PA. The effects of random and discrete sampling when estimating continuous–time diffusions. Econometrica. 2003;71(2):483–549.

[3] McDunnough P, Wolfson DB. On some sampling schemes for estimating the parameters of a continuous time series. Ann Inst Stat Math. 1979;31(1):487–497.

[4] Masry E. Probability density estimation from sampled data. IEEE Trans Inform Theory. 1983;29(5):696–709.

[5] Vilar JA. Kernel estimation of the regression function with random sampling times. Test. 1995;4(1):137–178.

[6] Vilar JA, Vilar JM. Finite sample performance of density estimators from unequally spaced data. Stat Probab Lett. 2000;50(1):63–73.

[7] Araya H, Bahamonde N, Fermín L, et al. On the consistency of the least squares estimator in models sampled at random times driven by long memory noise: the renewal case. Statist Sinica. 2023;33(1).

[8] Araya H, Bahamonde N, Fermín L, et al. On the consistency of least squares estimator in models sampled at random times driven by long memory noise. The jittered case. Statist Sinica. 2023;33(2).

[9] Hu T. Negatively superadditive dependence of random variables with applications. Chinese J Appl Probab Statist. 2000;16(2):133–144.

[10] Khursheed A, Lal Saxena KM. Positive dependence in multivariate distributions. Commun Stat Theory Methods. 1981;10(12):1183–1196.

[11] Eghbal N, Amini M, Bozorgnia A. Some maximal inequalities for quadratic forms of negative superadditive dependence random variables. Stat Probab Lett. 2010;80(7–8):587–591.

[12] Shen Y, Wang XJ, Yang WZ, et al. Almost sure convergence theorem and strong stability for weighted sums of nsd random variables. Acta Math Sin Engl Ser. 2013;29(4):743–756.

[13] Shen Y, Wang XJ, Hu SH. On the strong convergence and some inequalities for negatively superadditive dependent sequences. J Inequal Appl. 2013;2013(1):448.

[14] Wang XJ, Li X, Hu SH, et al. Strong limit theorems for weighted sums of negatively associated random variables. Stoch Anal Appl. 2010;29(1):1–14.

[15] Kemperman JHB. On the fkg-inequality for measures on a partially ordered space. Indag Math. (Proc) 1977;80:313–331.

[16] Block HW, Sampson AR. Conditionally ordered distributions. J Multivar Anal. 1988;27(1):91–104.