On the consistency of the least squares estimator in models sampled at random times driven by long memory noise

H. Araya, N. Bahamonde, L. Fermín, T. Roa and S. Torres

Pontificia Universidad Católica de Valparaíso and Universidad de Valparaíso, Chile

natalia.bahamonde@pucv.cl

Abstract

In this study, we prove the strong consistency of the least squares estimator in a random sampled linear regression model with long-memory noise and an independent set of random times given by renewal process sampling. Additionally, we illustrate how to work with a random number of observations up to time $T = 1$. A simulation study is provided to illustrate the behavior of the different terms, as well as the performance of the estimator under various values of the Hurst parameter H.

1. Introduction

In many applications, data are observed at random times. This situation arises from a variety of causes, such as machinery faults or the inability to observe data in certain periods. For the random modeling of observations, the renewal case represents progressive randomness and distance from periodic sampling.

This study offers an alternative approach to constructing trend regression models by taking into account long-memory behavior in the noise term. The interest in the long-memory noise model lies in the behavior of its covariance structure, which can cover a general class of noise. Parameter estimation problems in time series, represented as a trend plus long-memory noise, are well studied. In contrast, in time series models with long-memory, parameter estimation in models sampled at random times is more rare.

The concept of long-memory is very well characterized in terms of the spectral density function. However, the existence of this function is limited to stationary processes. When jointly considering a model with trend and long-memory properties, there is no stationarity property essential to defining the spectral density function in which the spectral estimate rests. In contrast, for spectral estimations, wavelet methods have been proposed for irregularly sampled real-valued data, including regression problems and long-memory estimation.

2. The least squares estimator in a simple regression model

In this study, we examine the least squares estimator (LSE) in a simple regression model, nonstationary in trend, with long-memory noise and observation measurements at random times. We also show how to deal with the number of observations needed to reach a fixed time T assuming, without loss of generality (w.l.o.g.), $T = 1$. To explain the long-memory or long-range dependence phenomenon in a model, it is common to represent it using the Hurst exponent H, which takes values in $(0, 1)$. In particular, long-range dependence can be seen when $H \in (1/2, 1)$. One of the most popular Gaussian stochastic processes with long-memory is fractional Brownian motion. We consider the following simple regression model:

$$Y_{\tau_{i+1}} = a\tau_i + \Delta B^H_{\tau_{i+1}}, \quad i = 0, \dots, N(1), \tag{2.1}$$

where $a \in \mathbb{R}$ is the unknown drift parameter of the model. Long-memory is represented by the noise, defined as $\Delta B^H_{\tau_{i+1}} = B^H_{\tau_{i+1}} - B^H_{\tau_i}$. Here, $\tau := \{\tau_i, 0 \leq i\}$ is a random increasing sequence of positive random times depending on N. However, this dependence is expressed through the distribution function, and the initial value, τ_0, is also a positive random variable; see the next section for a detailed definition. Note that $N(1) = \sum_{j\geq 1} 1_{\{\tau_j \leq 1\}}$ determines the number of events in $[0, 1]$. From the definition of τ, $N(1)$ is a discrete random variable, and N represents the expected number of observations within $[0, 1]$. The process $Y := \{Y_{\tau_{i+1}}, 0 \leq i\}$, defined in equation (2.1), is nonstationary. The long-memory or long-range dependence refers to the type of noise used. However, note that the long-memory property does not necessarily hold when working with random times.

The LSE estimator for a, the drift parameter of the random sampled regression model with long-memory noise in (2.1), is determined by $\hat{a}_{N(1)} = \sum_{i=0}^{N(1)} \tau_i Y_{\tau_{i+1}} / \sum_{i=0}^{N(1)} \tau_i^2$. Working with random times that are not upper bounded is a challenge, because both, the observation times and the number of observations within the interval, are random. Our way of dealing with this task is to divide the problem into three stages.

1. we study the almost sure convergence of $N(1)/N$ to 1
2. we define an auxiliary least squares type estimator, $\hat{a}_N = \sum_{i=0}^{N-1} \tau_i Y_{\tau_{i+1}} / \sum_{i=0}^{N-1} \tau_i^2$, considering a fixed number $N \in \mathbb{N}$, corresponding to the sampling frequency or sampling rate and, study the convergence of $\hat{a}_N \to a$.
3. we ensure the convergence of $|\hat{a}_N - \hat{a}_{N(1)}|$ to zero.

In practice, our estimator is based on $N(1)$ observations, because if $N(1) < N$, then $\hat{a}_N$ cannot be computed from the data.

Let $\tau = \{\tau_i; i \geq 0\}$ be a strictly increasing sequence of random points over time, the distribution function of which depends on N (to avoid superscript, the dependence on N is through the distribution function), where N represents the sampling frequency or sampling rate, that is, the average number of samples obtained in $[0, 1]$.

The sequence τ, defined by the renewal process (RP), is given as follows:

$$\tau_i = \sum_{j=0}^{i} t_j \quad i \geq 0, \tag{2.2}$$

where $\{t_j, j \geq 0\}$ is a sequence of independent and identically distributed random variables (i.i.d.), with a common distribution function $G_N(\cdot)$, that depend on N with support in $[0, \infty)$, and are absolutely continuous with density g_N, such that $G_N(0) = 0$, satisfying the following hypothesis:

H1 $\mathbb{E}[t_i] = \frac{1}{N}$ for all $i \geq 0$.

H2 $\mathbb{E}[t_i^2] = \frac{\kappa_1}{N^\alpha}$, $0 < \alpha \leq 2$.

H3 $\mathbb{E}[t_i^4] = \frac{\kappa_2}{N^\beta}$, $0 < \beta \leq 2\alpha$.

Here κ_1 and κ_2 are constants not depending on N. Note that the conditions $\alpha \leq 2$ and $\beta \leq 2\alpha$ come from the Cauchy inequality.

Henceforth, $G_{N,i}$ denotes the probability distribution function associated with τ_i and its density functions $g_{N,i}$, and $N(1)$ is the number of observations needed to sample up to one. Examples of distributions satisfying **H1** to **H3** are: the beta prime distribution, with parameters $(1, N + 1)$, and the exponential distribution with parameter $\lambda = N$. This distribution is a limit case for $\alpha = 2$.

Because we proved the a.s. convergence of τ_N to one, we can similarly get the a.s. convergence of $\tau_{N+N\epsilon} - \tau_N$ to ϵ, for $\alpha > 3/2$ and $\beta > 2$, as shown below. Using this fact and recalling that the random variables τ_N and $\tau_{N+N\epsilon} - \tau_N$ are independent.

3. Convergence results

In this section, we provide our main result.

We give the main properties of the process $B^H = \{B^H_t, t \geq 0\}$ with zero mean, the increments of which are considered as the noise in model (2.1).

N1 Covariance structure: $R_H(t, s) = \mathbb{E}(B^H_t B^H_s) = \frac{1}{2}\left(t^{2H} + s^{2H} - |t - s|^{2H}\right)$.

N2 We consider a finite-variance process that is self-similar with stationary increments.

N3 The random time sequence τ, which depends on N, and the long-memory noise B^H are independent.

We prove that the LSE is an unbiased and strongly consistent estimator for a, the drift parameter of the random sampled regression model with long-memory noise. To estimate the parameter of interest in the model (2.1), the LSE is computed and is determined by

$$\hat{a}_{N(1)} = \frac{\sum_{i=0}^{N(1)} \tau_i Y_{\tau_{i+1}}}{\sum_{i=0}^{N(1)} \tau_i^2}. \tag{3.1}$$

Recall that, from (2.1) and (3.1), we have

$$\hat{a}_{N(1)} - a = \frac{\sum_{i=0}^{N(1)} \tau_i \Delta B^H_{\tau_{i+1}}}{\sum_{i=0}^{N(1)} \tau_i^2}. \tag{3.2}$$

Let $\hat{a}_N$ be the LS estimator obtained by replacing $N(1)$ with $N - 1$ in (3.1); that is,

$$\hat{a}_N = \frac{\sum_{i=0}^{N-1} \tau_i Y_{\tau_{i+1}}}{\sum_{i=0}^{N-1} \tau_i^2}. \tag{3.3}$$

We consider the following decomposition from (3.1) and (3.3): $\hat{a}_{N(1)} - a = \hat{a}_{N(1)} - \hat{a}_N + \hat{a}_N - a$. Then, the proof of the main Theorem 1 is given in two steps

- 1st, we prove in Prop. 1 that $\hat{a}_N$ converges a.s. to a.
- 2nd, we control the difference $\hat{a}_{N(1)} - \hat{a}_N$ a.s. in Prop. 2.

Theorem 1. *Let τ be the random time defined in* (2.2)*, and let the process $B^H = \{B^H_t, t \geq 0\}$ with a zero mean, and with increments that are considered as the noise. These satisfy hypotheses* **H1** *to* **H3** *and* **N1** *to* **N3***, respectively. Then, for $\alpha > \max\{3/2, 1/H\}$ and $\beta > 2$, the LS estimator $\hat{a}_{N(1)}$ given in (3.1) of the drift parameter a in model (2.1) is strongly consistent,*

$$\hat{a}_{N(1)} \xrightarrow[N\to\infty]{a.s.} a.$$

For $\alpha > 1$ and $\beta > 1$, convergence in probability is ensured.

Proposition 1. *Let τ be the random time defined in* (2.2)*, and let the process $B^H = \{B^H_t, t \geq 0\}$ with a zero mean, and with increments that are considered as the noise. These satisfy hypotheses* **H1** *to* **H3** *and* **N1** *to* **N3***, respectively, for $\alpha > \max\{3/2, 1/H\}$ and $\beta > 2$. Then, the LS estimator $\hat{a}_N$ given in (3.3) of the drift parameter a in model (2.1) is strongly consistent,*

$$\hat{a}_N \xrightarrow[N\to\infty]{a.s.} a.$$

For $\alpha > 1$ and $\beta > 1$, the convergence in probability is ensured.

Proposition 2. *Let τ be the random time defined in* (2.2)*, and let the process $B^H = \{B^H_t, t \geq 0\}$ with a zero mean, and with increments that are considered as the noise. These satisfy hypotheses* **H1** *to* **H3** *and* **N1** *to* **N3***, respectively, for $\alpha > \max\{3/2, 1/H\}$ and $\beta > 2$. Consider the LS estimators $\hat{a}_N$ and $\hat{a}_{N(1)}$ of the drift parameter a given in* (3.3) *and* (3.1)*, respectively, for the model (2.1). Then,*

$$|\hat{a}_{N(1)} - \hat{a}_N| \xrightarrow[N\to\infty]{a.s.} 0. \tag{3.4}$$

For $\alpha > 1$ and $\beta > 1$, the convergence in probability is ensured.

4. Simulation Study

We develop a Monte Carlo simulation study to assess the finite-sample properties for the LS estimator in the linear regression model (2.1). The long-memory noise driven is by a fractional Brownian motion evaluated at deterministic times and two different random times defined by Equation (2.2).

The deterministic case: We consider the model defined by equation (2.1) observed at equally spaced times, that is, $\tau_i = i/N$, for $i = 1, \dots, N$. We consider $N = 200$.

The exponential and beta prime case: The most studied renewal process is the Poisson process, which appears when t_i has an exponential distribution (λ). We consider $\lambda = 200$. For the beta prime distribution, we consider a distribution with parameters $(1, 201)$.

For all the simulations shown, we consider $M = 1000$ replicates of the model with the parameters $a = 0.2$ and $a = 2$. For the exponential and beta prime cases, the number of observations is a random variable $N(1)$, representing how many observations are within the interval $[0, 1]$. We also consider different values of the Hurst parameter: $H = 0.05$, $H = 0.25$ and $H = 0.45$ (anti-persistent cases); and $H = 0.55$, $H = 0.75$, and 0.95 (long-memory cases).

- From the results of equation (**??**) and the result of simulation, we obtain an upper bound on the convergence rate of $|a_{N(1)} - a|$.
- When an exponential distribution of parameter $\lambda = N$ is considered ($\alpha = 2$ and $\beta = 4$), the upper bound is given by C/N for the convergence in probability. For a.s. convergence, the upper bound is C/N^{2H-1}. The latter bound coincides when the times are considered to be equally spaced.
- As H increases, the slope fit improves.
- The values of the mean show that the estimator is unbiased. Note that the SD decreases as the value of H approaches one, which is expected, because the conditional variance of the noise decreases as H approaches one.

Reference: Statistica Sinica Preprint No: SS-2020-0457 (in press).