Journal of Multivariate Analysis 146 (2016) 119–137

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Detecting and estimating intensity of jumps for discretely observed ARMA$D(1, 1)$ processes

D. Blanke [a,*], D. Bosq [b]

[a] *Avignon University, LMA EA2151, 33 rue Louis Pasteur, F-84000 Avignon, France*

[b] *Sorbonne Universités, UPMC Univ Paris 06, EA 3124, Laboratoire de Statistique Théorique et Appliquée, F-75005, 4 place Jussieu, Paris, France*

ARTICLE INFO

Article history:
Received 31 December 2014
Available online 5 September 2015

AMS 2010 subject classifications:
primary 62M
60F17

Keywords:
Functional linear processes
ARMA$D(1, 1)$ processes
Jumps
Estimation of intensity
Discrete data

ABSTRACT

We consider n equidistributed random functions, defined on [0,1], and admitting fixed or random jumps, the context being $D[0, 1]$-valued ARMA(1, 1) processes. We begin with properties of ARMA$D(1, 1)$ processes. Next, different scenarios are considered: fixed instants with a given but unknown probability of jumps (the deterministic case), random instants with ordered intensities (the random case), and random instants with non ordered intensities (the completely random case). By using discrete data and for each scenario, we identify the instants of jumps, whose number is either random or fixed, and then estimate their intensity.

1. Introduction

1.1. There is an abundant literature concerning functional data analysis (FDA) and prediction of stochastic processes in infinite dimensional spaces. In particular, the books by Ramsay and Silverman [50], Ferraty and Vieu [31], Ferraty and Romain [30], Horváth and Kokoszka [35] and the recent book edited by Bongiorno et al. [9] contain interesting theoretical and practical results. See also [11,13]. In general, X takes its value in $L^2 = L^2([0, h])$ or in $C = C([0, h])$, but, in some situations, one may consider that a *jump* does exist if there is a large peak: see, for example the annual sediment in [9, p. 8]. Thus, it is perhaps more natural to consider the space $D = D([0, h])$ which is càdlàg and equipped with the Skorohod metric d° (see [6, p. 125]): with that metric, D becomes a separable complete metric space. Note that this metric is not easy to compute. In this paper, we consider càdlàg processes from a functional point of view: by this way, we work in the context of FDA with jumps.

1.2. Works dedicated to *jumps in stochastic processes* appear very often: actually, there are more than 1200 papers concerning them. Thus, we may only give recent and limited references. For example, processes with jumps are widely used in finance: we may refer to [20,54]; [39, part 2], [29, ch. 10]; [49], etc.; but applications can also be found in fields as varied as the environment, medicine, reliability, etc., see e.g. [33,4,16,10]. Many mathematical models have been proposed and studied [25,42,34,24], and statistical estimation appears e.g. in [18,17,26], etc. Note that the pioneer paper concerning jumps appears in Lévy [44]. Other references of interest will appear below.

* Corresponding author.
E-mail addresses: delphine.blanke@univ-avignon.fr (D. Blanke), denis.bosq@upmc.fr (D. Bosq).

http://dx.doi.org/10.1016/j.jmva.2015.08.014
0047-259X/

1.3. Now, here and in the books quoted in Section 1.1, our purpose is somewhat different since we want to observe a process over a *sequence of time intervals*. More precisely, let $(\xi_t,\ t \in \mathbb{R})$ be a real measurable continuous time process. We put

$$X_n(t) = \xi_{(n-1)h+t}, \quad 0 \le t \le h,\ n \in \mathbb{Z} \tag{1.1}$$

where $h > 0$ is a time interval. The process may contain some jumps and we envisage to *detect* them and to estimate *intensity of jumps*, given the data $X_1, \ldots, X_n$.

Another motivation should be *prediction of* X_{n+1} *over the time interval* $[nh, (n+1)h]$. One way to predict X_{n+1} would be to treat *continuous time and jumps separately* (see [51,1,52], etc.). As an example, consider the functional autoregressive process of order 1 (ARD(1)):

$$X_{n+1}(t) = \rho(X_n)(t) + Z_{n+1}(t), \quad 0 \le t \le h,\ n \in \mathbb{Z},$$

where ρ is a continuous linear operator with respect to the sup-norm. Then, in order to separate the continuous part from the jump's part, we may suppose that $\rho(D) \subset C$. That condition is satisfied by the *Ornstein–Uhlenbeck process driven by a Levy process*, cf Example 2.1. Another classical example is given by: $\rho_r(x)(t) = \int_0^h r(s,t)x(s)\,ds$, $0 \le t \le h, x \in D$ where r satisfies Example 2.2, see also [21,35], etc. Thus, $\rho_r(x) \in C$. Finally, the condition $\rho(D) \subset C$ seems quite standard and characterizes the unpredictability of jumps by confining them in the innovation process. Now, the best probabilistic predictor of X_{n+1} is $\rho(X_n)$ and it can be approximated by using an estimator of ρ. An exponential rate is obtained in [11, p. 222–235], when the detector and intensity of jumps appear in the current paper. One direction (currently under development) will consist in combining the two approaches to improve the prediction.

1.4. A more general model should be the ARMA$D(p,p)$ process defined by

$$X_n - \rho_1(X_{n-1}) - \cdots - \rho_p(X_{n-p}) = Z_n - \rho'_1(Z_{n-1}) - \cdots - \rho'_p(Z_{n-p}), \quad n \in \mathbb{Z},$$

where X_n and Z_n are D-valued and where ρ_j, $\rho'_j, j, j' = 1, \ldots, p$ are continuous linear operators with respect to the sup-norm. In order to study this process, it should be possible to work in the space $D([0,h]^p)$ (cf [43]). Note that if ρ_j, $\rho'_j, j, j' = 1, \ldots, p$ are C-valued, X_n and Z_n have again the same jumps.

Now, since this model is difficult to handle, and in order to simplify the exposition, we take $p = 1$ and write

$$X_n - \rho(X_{n-1}) = Z_n - \rho'(Z_{n-1}), \quad n \in \mathbb{Z},$$

note that, Z_{n-1} may be replaced with an exogenous variable (see for example [32]).

1.5. We now give some practical examples of *jumps over time intervals*:

- a patient's *electrocardiogram* at each minute [46,48,45];
- the *temperature* day by day [56];
- *El Niño* southern oscillation (ENSO): a prediction over one year shows a jump in May [5];
- *wave amplitude* [55];
- *pollution* day by day [35];
- *credit cards* transaction and its prediction [35];
- another example is *electricity consumption*: it admits a jump early in the morning and in the evening (see [3,27,28]);
- administration of a *drug treatment*: each day produces a shock at time intervals (see [40]);
- *astronomical time series* with 100000 data (see [48]);
- *earthquake* and *explosion*: [46];
- predicting *ozone* [36,22,15,23];
- predicting the *euro–dollar rate* [41];
- finally, the *mistral gust* during one day or one week is one of our objective for prediction: 240000 data are at our disposal. Predicting the greatest jump should be of interest, see [37].

1.6. In our considered framework, preliminary results were first obtained by Bosq [12] and, the case of observations in continuous time also appears in [7]. Here, we use *high frequency data* (HFD); this scheme appears in many situations (see [8,19,2] among others). Concerning prediction with HFD, practical results will be studied later with combined predictors. In particular, we will apply the results to the mistral gusts with big data.

1.7. In Section 2, we introduce the ARMA$D(1,1)$ model which is connected with FDA:

$$X_n - m - \rho(X_{n-1} - m) = Z_n - \rho'(Z_{n-1}), \quad n \in \mathbb{Z} \tag{1.2}$$

where m is a trend and $\rho(D) \subset C$, $\rho'(D) \subset C$ so that (X_n) and (Z_n) have the same jumps. We give several properties of (1.2) as well as examples. In the following, we study various types of jumps.

In Section 3, we consider data of the form $X_i(\frac{\ell}{q_n})$, $\ell = 0, \ldots, q_n$, $q_n \ge 1$, $i = 1, \ldots, n$; where ℓ and q_n are integers and $(X_1, \ldots, X_n)$ are D-valued realizations of (1.2). We consider the case of fixed but unknown instants of jumps $t_1, \ldots, t_k$, where t_j, denotes the jth jump, $j = 1, \ldots, k$ and k is unknown too. In this part, each jump may occur randomly at time t_j with unknown probability $p_j \in]0,1]$, $j = 1, \ldots, k$, so the number of jumps is a random variable depending on $i = 1, \ldots, n$.