Chaos, Solitons and Fractals 000 (2017) 1–11

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals
Nonlinear Science, and Nonequilibrium and Complex Phenomena

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

Frontiers

Identification and validation of stable ARFIMA processes with application to UMTS data

Krzysztof Burnecki*, Grzegorz Sikora

Faculty of Pure and Applied Mathematics, Hugo Steinhaus Center, Wroclaw University of Science and Technology, Wyspianskiego 27, 50-370 Wroclaw, Poland

A R T I C L E I N F O

Article history:
Received 5 January 2017
Revised 25 March 2017
Accepted 28 March 2017
Available online xxx

Keywords:
ARFIMA process
Stable distribution
Long memory
UMTS data

A B S T R A C T

In this paper we present an identification and validation scheme for stable autoregressive fractionally integrated moving average (ARFIMA) time series. The identification part relies on a recently introduced estimator which is a generalization of that of Kokoszka and Taqqu and a new fractional differencing algorithm. It also incorporates a low-variance estimator for the memory parameter based on the sample mean-squared displacement. The validation part includes standard noise diagnostics and backtesting procedure. The scheme is illustrated on Universal Mobile Telecommunications System (UMTS) data collected in an urban area. We show that the stochastic component of the data can be modeled by the long memory ARFIMA. This can help to monitor possible hazards related to the electromagnetic radiation.

1. Introduction

The concept of anomalous diffusion and fractional dynamics has deeply penetrated the statistical and chemical physics communities, yet the subject has also become a major field in mathematics [1,2]. Historically, fractional dynamical systems are related to the concept of fractional dynamic equations. This is an active field of study in physics, mechanics, mathematics, and economics investigating the behavior of objects and systems that are described by using differentiation of fractional orders. The celebrated fractional Fokker–Planck equation (FFPE), describing anomalous diffusion in the presence of an external potential was derived explicitly in [3], where methods of its solution were introduced and for some special cases exact solutions were calculated.

Derivatives and integrals of fractional orders can be used to describe random phenomena that can be characterized by long (power-like) memory or self-similarity [1,2]. Long memory (or long-range dependence) is a property of certain stationary stochastic processes describing phenomena, which concern the events that are arbitrarily distant still influence each other exceptionally strong. It has been associated historically with slow decay of correlations and a certain type of scaling that is connected to self-similar processes [4,5].

Recently, there has been a great interest in long-range dependent and self-similar processes, in particular fractional Brownian motion (FBM), fractional stable motion (FSM) and autoregressive fractionally integrated moving average (ARFIMA), which are also called fractional autoregressive integrated moving average (FARIMA) [6,7]. This importance can be judged, for example, by a very large number of publications having one of these notions in the title, in areas such as finance and insurance [8–15], telecommunication [16–21], hydrology [22], climate studies [23], linguistics [24], DNA sequencing [25] or medicine [26]. Long-range dependent and self-similar processes also appear widely in other areas like biophysics [7,27–32] or astronomy [33]. These publications address a great variety of issues: detection of long memory and self-similarity in the data, statistical estimation of parameters of long-range dependence and self-similarity, limit theorems under long-range dependence and self-similarity, simulation of long memory and self-similar processes, relations to ergodicity and many others [6,7,34–37].

The FBM, FSM and ARFIMA serve as basic stochastic models for fractional anomalous dynamics [7]. The former two models are self-similar and their increments form long-range dependent processes. The discrete-time ARFIMA process is stationary and generalizes both models since aggregated, in the limit, it converges to either fractional Brownian or stable motion. As a consequence, a partial sum ARFIMA process can be considered as a unified model for fractional anomalous diffusion in experimental data [38]. A type of anomaly of the process is controlled only by its the memory parameter regardless of the underlying distribution [28]. We also note that there is a relationship between the ARFIMA and continuous time random walk (CTRW) which is a classical model of anomalous diffusion [3,39]. The latter can be obtained by subordi-

* Corresponding author.
E-mail address: krzysztof.burnecki@pwr.edu.pl (K. Burnecki).

http://dx.doi.org/10.1016/j.chaos.2017.03.059
0960-0779/

nation of the Ornstein–Uhlenbeck process which discrete version is an autoregressive (AR) process, so a special case of the ARFIMA [40,41].

In contrast to FBM and FLSM, ARFIMA allows for different light- and heavy-tailed distributions, and both long (power-like) and short (exponential) dependencies [38]. Moreover, as a stationary process, it provides prediction tools.

It appears that the values of ARFIMA with Gaussian noise, for the memory parameter d greater than 0, have so slowly decaying autocovariance function that it is not absolutely summable. This behavior serves as a classical definition of the long-range dependence. However, it is also a well-known fact that the heavy-tailed probability distributions with diverging variance are ubiquitous in nature and finance [42–47].

The stable probability densities have the asymptotics decaying at infinity as $|x|^{-1-\alpha}$, where α is the index of stability varying between 0 and 2. They attract distributions having the same law of decay. On the contrary, the Gaussian distribution has the index of stability 2 and attracts all distributions with lighter tails [42,48,49].

Stably distributed random noises are observed in such diverse applications as plasma turbulence (density and electric field fluctuations [49–51]), stochastic climate dynamics [52–54], physiology (heartbeats [55]), electrical engineering [56], biology [28,30], and economics [57,58]. Heavy-tailed distributions govern circulation of dollar bills [59] and behavior of the marine vertebrates in response to patchy distribution of food resources [60].

In this paper we propose an identification and validation scheme for ARFIMA processes with noise in the domain of attraction of the stable law which is based on estimation algorithm introduced in [61]. The scheme is illustrated on the electromagnetic radiation data which shows long memory behavior which is also observed for telecommunication data in [19].

The paper is organized as follows: in Section 2 we recall basic facts about a prominent example of long memory processes, namely ARFIMA time series. In Section 3 we introduce a step by step procedure for identification of a ARFIMA process. The procedure involves (i) a method of preliminary estimation of the memory parameter based on the mean-squared displacement, (ii) a new method of fractional differencing which leads to model order estimation and (iii) the estimation formula for stable ARFIMA times series introduced in [61]. Section 4 is devoted to validation of the fitted model. It consists of analysis of residuals: testing their randomness and fitting a distribution which is done by standard statistical tests, and backtesting which involves prediction formula for ARFIMA time series. The identification and validation procedure is illustrated in Section 5 on electromagnetic field data collected in the vicinity of an Universal Mobile Telecommunications System (UMTS) station in Wroclaw. After removing deterministic seasonality and volatility from the data, a long memory ARFIMA process is identified and validated. In Section 6 a summary of the results is given.

2. ARFIMA process

In this section we briefly present the main facts about ARFIMA time series which were introduced in [62] and [63]. Such process $\{X_t\}$, denoted by ARFIMA(p, d, q), is defined by

$$\Phi_p(B)X_t = \Theta_q(B)(1-B)^{-d}Z_t, \tag{1}$$

where innovations (noise sequence) Z_t are i.i.d. random variables with either finite or infinite variance. We also assume that the innovations belong to the domain of attraction of an α-stable law with $0 < \alpha \le 2$. For the infinite variance case ($\alpha < 2$) this means that

$$P(|Z_t| > x) = x^{-\alpha}L(x), \ \ as\ x \to \infty, \tag{2}$$

where L is a slowly varying function at infinity, and

$$\frac{P(Z_t > x)}{P(|Z_t| > x)} \to a, \ \ \frac{P(Z_t < -x)}{P(|Z_t| > x)} \to b, \ \ as\ x \to \infty, \tag{3}$$

where a and b are nonnegative numbers such that $a + b = 1$. The finite variance case ($\alpha = 2$) leads the domain of attraction of Gaussian law. Polynomials Φ_p and Θ_q have classical forms, i.e. $\Phi_p(z) = 1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p$ is the autoregressive polynomial, $\Theta_q(z) = 1 + \theta_1 z + \theta_2 z^2 + \ldots + \theta_q z^q$ is the moving average polynomial. The operator B, called the backward operator, satisfies $BX_t = X_{t-1}$ and $B^j X_t = X_{t-j}$, $j \in \mathbb{N}$. The crucial part of ARFIMA Definition (1) is the operator $(1-B)^{-d}$ called the fractional integrating operator and the fractional number d called the memory parameter.

The operator $(1-B)^{-d}$ has the infinite binomial expansion

$$(1-B)^{-d}Z_t = \sum_{j=0}^{\infty} b_j(d) Z_{t-j}, \tag{4}$$

where the $b_j(d)'$s are the coefficients in the expansion of the function $f(z) = (1-z)^{-d}$, $|z| < 1$, i.e.

$$b_j(d) = \frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)}, \ \ j = 0, 1, \ldots, \tag{5}$$

where Γ is the gamma function. The series (4) is convergent almost surely and ARFIMA Definition (1) is correct if and only if

$$\alpha(d-1) < -1 \Leftrightarrow d < 1 - \frac{1}{\alpha}. \tag{6}$$

In the Gaussian case where $\alpha = 2$ we have $d < 1/2$. Under Assumption (6) and when polynomials Φ_p and Θ_q do not have common roots, and Φ_p has no roots in the closed unit disk $\{z\colon |z| \le 1\}$, the ARFIMA(p, d, q) time series defined by (1) has the causal moving average form

$$\underbrace{X_t}_{\text{ARFIMA }(p,d,q)} = \overbrace{(1-B)^{-d}}^{d-\text{fractional integrating}} \underbrace{\frac{\Theta_q(B)}{\Phi_p(B)} Z_t}_{\text{ARMA }(p,q)} = \sum_{j=0}^{\infty} c_j(d) Z_{t-j}, \tag{7}$$

where

$$c_j(d) = b_j(d) + \sum_{i=1}^{p} \phi_i c_{j-i}(d) + \sum_{k=1}^{q} \theta_k b_{j-k}(d), \tag{8}$$

for $j = 0, 1, \ldots$ Therefore ARFIMA(p, d, q) time series can be obtained by d-fractional integrating of ARMA(p, q) series. The d-fractional integrating through $(1-B)^{-d}$ operator builds the dependence between observations in a ARFIMA sequence, even as they are far apart in time.

When $d > -1 + 1/\alpha$ the ARFIMA(p, d, q) time series $\{X_t\}$ is invertible and Definition (1) can be rewritten in the equivalent form

$$\Phi_p(B) \underbrace{\overbrace{(1-B)^d}^{d-\text{frac. differencing}} \ \overbrace{X_t}^{\text{ARFIMA }(p,d,q)}}_{\text{ARMA }(p,q)} = \Theta_q(B) Z_t. \tag{9}$$

The operator $(1-B)^d$, called the fractional differencing operator, is the inverse operator of the fractional integrating operator $(1-B)^{-d}$. It has the infinite binomial expansion of the form (4) with the opposite d. Hence, according to (9), the ARMA(p, q) time series can be obtained after d–fractional differencing of ARFIMA(p, d, q) sequence.

When the memory parameter d is close to 1/2, all the coefficients $b_j(d)'$s are positive and converge to zero at a power rate. In view of series representation (4), ARFIMA(0, d, 0) observation X_t depends not only on the present noise observation Z_t, but also depends strongly on the whole history of the noise process. Hence