Research Article

Applied Stochastic Models in Business and Industry

Received 18 December 2009, Revised 29 January 2010, Accepted 29 January 2010 Published online 25 May 2010 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/asmb.847

Fitting non-Gaussian persistent data

Wilfredo Palma[a∗†] and Mauricio Zevallos[b]

This paper discusses a new methodology for modeling non-Gaussian time series with long-range dependence. The class of models proposed admits continuous or discrete data and considers the conditional variance as a function of the conditional mean. These types of models are motivated by empirical properties exhibited by some time series. The proposed methodology is illustrated with the analysis of two real-life persistent time series. The first application is concerned with the modeling of stock market daily trading volumes, whereas the second application consists of a study of mineral deposit measurements. Copyright © 2010 John Wiley & Sons, Ltd.

Keywords: ARFIMA models; conditional variance; long-range dependence; persistence; quasi-maximum likelihood; prediction

1. Introduction

Long-memory data arise in many fields such as finance, economics, politics, geophysics, hydrology, among many others. These types of time series are characterized by a sample autocorrelation function with hyperbolic decay and significant autocorrelations even at large lags. Most of the literature on modeling long-memory is concerned with the Gaussian processes. For a recent review see, for example, Palma [1] and references therein.

Nevertheless, many empirical studies have shown the presence of time series exhibiting long-range dependence and non-Gaussianity. For example, persistence has been observed in stock market daily trading volumes, see for example Lobato and Velasco [2] and references therein. On the other hand, there are several studies of geophysical data exhibiting long-range dependence, see for instance Dmowska and Saltzman [3] and Shumway and Stoffer [4].

In those studies, the authors use well-known Gaussian ARFIMA models, but they do not use potentially more appropriate specific distributions in the case, for example, of trading volumes or geological data. As far as we know, one of the few works dealing with non-Gaussian long-range dependent time series is provided by Brockwell [5]. That paper models long-memory in the observed data by means of a latent long-range-dependent process, all this within a generalized linear model (GLM) framework where the estimation is carried out via the Markov Chain Monte Carlo (MCMC) methods.

Unlike the previous studies of trading volume and geological data mentioned above, this paper proposes a statistical framework that allows the choice of a specific data distribution conditional on their past. For example, in the analysis of the IBM daily trading volumes we use a Poisson distribution and for the geological data we use a Gamma distribution, as described in Section 5. On the other hand, this paper provides an alternative approach to Brockwell [5]. In our setting, we propose a long-memory time series model in which the distribution of the observations is specified conditionally on their past. One benefit of this approach is that since there is no latent process, the likelihood function of the model can be directly computed from the data.

This paper is organized as follows. Section 2 describes the data sets studied in this work. In Section 3, a family of models with conditional mean, conditional heteroscedasticity and long-memory is proposed and some of its properties are established. Section 4 deals with the estimation, diagnostics, model building and prediction of the proposed model. This methodology is applied to the analysis of a discrete and a continuous time series data in Section 5 and final remarks are

[a]*Department of Statistics, Pontificia Universidad Católica de Chile, Santiago, Chile*
[b]*Department of Statistics, University of Campinas, Brazil*
∗*Correspondence to: Wilfredo Palma, Department of Statistics, Pontificia Universidad Católica de Chile, Santiago, Chile.*
†*E-mail: wilfredo@mat.puc.cl*

Contract/grant sponsor: Fondecyt; contract/grant number: 1085239
Contract/grant sponsor: FAPESP
Contract/grant sponsor: FAEPEX

given in Section 6. The proofs of the results presented in this paper are given in the technical appendix at the end of the paper.

2. The data

Two time series exhibiting non-Gaussianity and long-memory are analyzed in this paper. The first data set consists of a time series of counts corresponding to daily trading volume of IBM stocks. These data have been already analyzed in several studies, see, for example, Lobato and Velasco [2] and the references therein, but by taking a logarithmic transformation of the trading volume. See also Andersen [6] for another modeling approach of these data. Here we deal directly with the series of counts. The second series discussed in this work corresponds to glacial varves, which are sedimentary deposits of layers of sand and silt deposited yearly during the spring melting seasons.

2.1. Stock market trading volume data

The data analyzed in this work correspond to daily trading volume of IBM stock for the period 2 January 1986–31 December 1993; see Figure 1. We use this period to illustrate the proposed methodology as the analysis is simplified here by not having to explicitly model complex long-term trends in the data. Furthermore, unlike most of the previous studies, we will model this time series as counts, avoiding the log-transformation of the original data to achieve normality. The original series, displayed in Figure 1, exhibits several peaks whereas the histogram, see the left panel of Figure 2, indicates the asymmetry of the distribution. In addition, the autocorrelation function shown on the right panel of Figure 2 indicates the persistence of this series.

2.2. Sedimentary deposits data

Figure 3 displays the thicknesses of the yearly varves at one location in Massachusetts for the period 11,833–11,200 BC, see Shumway and Stoffer [4] for further details. This data set is posted at www.stat.pitt.edu/~stoffer/tsa.html. Several peaks are observed in the time series plot of Figure 3. In addition, the histogram of the data, shown on the left panel of Figure 4, reveals that the distribution is very asymmetric. The evidence of long-memory behavior can be obtained from the sample autocorrelation function displayed on the right panel of Figure 4.

The two time series of the previous examples seem to display non-Gaussian distribution and long-range dependence. A family of models that can capture these features is proposed in the next section.

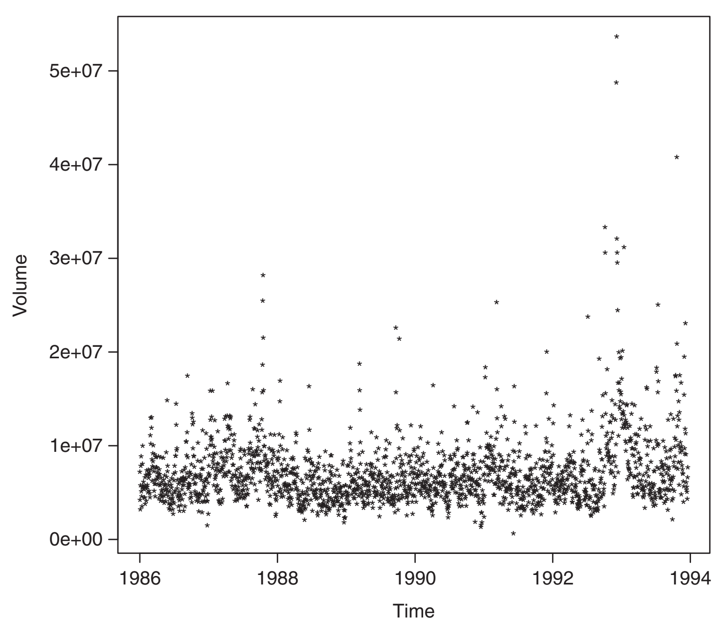

Figure 1. IBM daily trading volume data, 1986–1993.

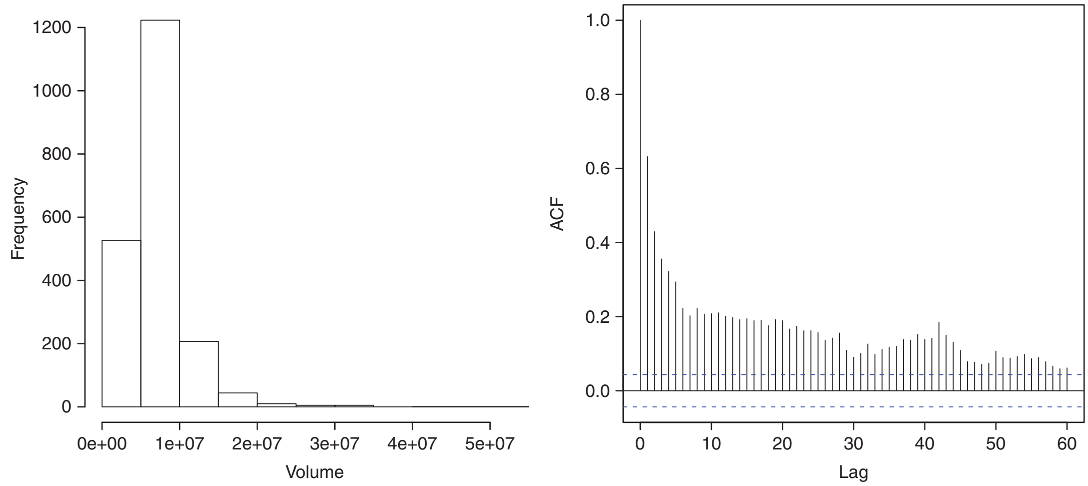

Figure 2. IBM daily trading volume data: histogram (left panel) and sample autocorrelation function (right panel).

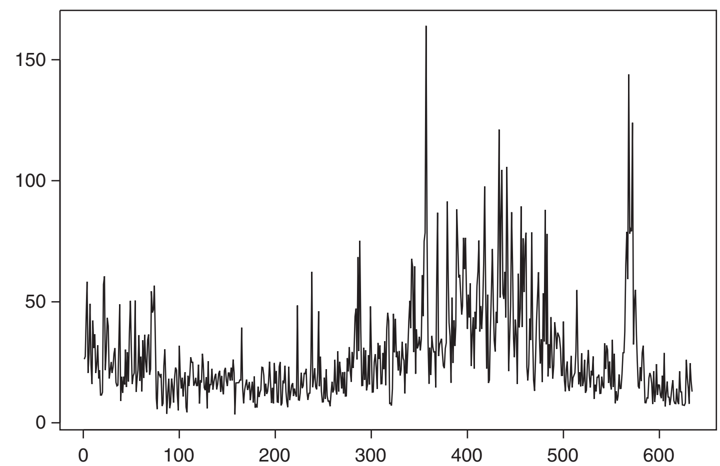

Figure 3. Varve glacial data.

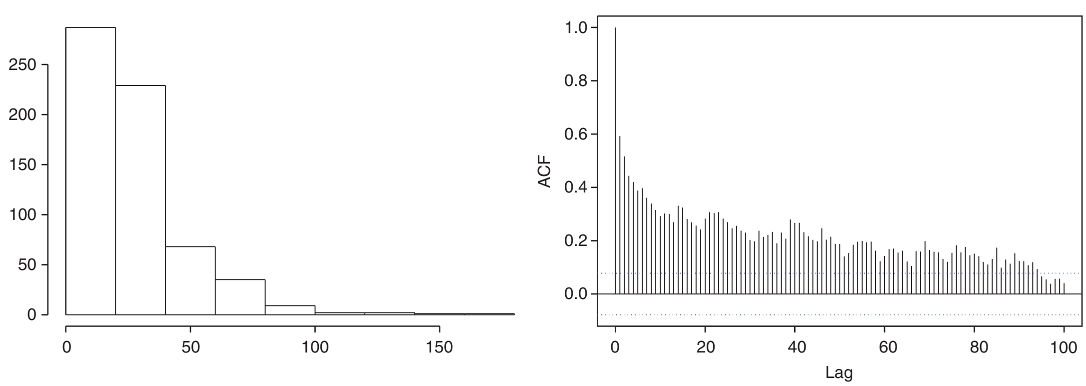

Figure 4. Varve glacial data: histogram (left panel) and sample autocorrelation function (right panel).

3. Methodology

In order to analyze the time series data presented in the previous section, we introduce next a class of conditional long-memory models (CLMs). In addition, a number of important statistical properties are derived and some specific models are discussed in detail.

3.1. Model

Let $G(\alpha, \beta)$ be a distribution corresponding to a discrete or continuous nonnegative random variable with both finite mean α and variance β. Let g be a positive function, μ be a constant and $\{\pi_j\}_{j\geqslant 0}$ be an absolutely summable sequence of real numbers, that is $\sum_{j=0}^{\infty}|\pi_j|<\infty$ such that $\pi_0=1$ and $\pi_j \approx Cj^{-d-1}$ for large positive j and some $d<\frac{1}{2}$. The coefficient d is the so-called long-memory parameter and it is related to the self-similar Hurst parameter, H, by the formula $d=H-\frac{1}{2}$. A conditional long-memory process $\{y_t\}$ with values in $\mathscr{Y}$ is defined as

$$y_t|\mathscr{F}_{t-1} \sim G(\lambda_t, g(\lambda_t)), \tag{1}$$

$$\lambda_t = \mu \sum_{j=0}^{\infty} \pi_j - \sum_{j=1}^{\infty} \pi_j y_{t-j}, \tag{2}$$

where $\mathscr{F}_t$ is the σ-field generated by $\{y_t, y_{t-1}, \ldots\}$, the information up to instant t. In addition, the conditional distribution function, G, may depend on other parameters besides π_j and μ. These parameters will be denoted by the vector η. In (1), symbol $\sim$ denotes that, conditional on the information $\mathscr{F}_{t-1}$, y_t has distribution G with variance $\mathrm{Var}[y_t|\mathscr{F}_{t-1}]=g(\lambda_t)$, which is a function of the conditional mean $\mathbf{E}[y_t|\mathscr{F}_{t-1}]=\lambda_t$.

Observe that there are a number of time series methodologies for handling non-Gaussian long-memory data, including the approaches proposed by Stanislavky *et al.* [7], Samorodnitsky and Taqqu [8] and Chechkin and Gonchar [9], among others. For example, the ARFIMA model described in Equations (2)–(3) of Stanislavky *et al.* [7] is represented by a linear Wold expansion where the noise can have a non-Gaussian distribution. This formulation is very helpful for handling, for example, processes with infinite variance. A similar approach is considered in the self-similar processes discussed by Samorodnitsky and Taqqu [8]. However, it seems to be hard with this formulation to handle, for example, positive observations or count data. The approach taken in our paper simplifies this issue since we specify explicitly the conditional distribution of the data given a data-driven parameter. In Chechkin and Gonchar [9] paper, the authors consider a persistent Levy process that extends the Fractional Gaussian Noise model to other distributions. Similar to the approach by Stanislavky *et al.* [7], handling positive or count data may be difficult with this technique since the process modeled in this context comes from the increments of the Levy motion.

Models (1)–(2) can be written in different ways. For instance, if we define the sequence $\varepsilon_t = y_t - \mathbf{E}[y_t|\mathscr{F}_{t-1}] = y_t - \lambda_t$ then $\mathbf{E}(\varepsilon_t)=0$ and, as shown in the next lemma, if $\mathbf{E}[g(\lambda_t)]$ is finite and constant, then $\{\varepsilon_t\}$ is an *innovation* process, that is a zero-mean, uncorrelated sequence with finite constant variance.

Lemma 1

Consider the model described by (1)–(2). Then,

(a) $\mathrm{Var}(\varepsilon_t)=\mathbf{E}[g(\lambda_t)]$.
(b) $\mathrm{cov}(\varepsilon_t, \varepsilon_s)=0$, for all $t \neq s$.
(c) If $\mathrm{Var}(\varepsilon_t)$ is a finite constant, then $\{\varepsilon_t\}$ is an innovation process.

Based on $\{\varepsilon_t\}$, we can write the model as follows: replacing $\lambda_t = y_t - \varepsilon_t$ in (2) and since $\pi_0=1$, we obtain

$$\Pi(B)(y_t - \mu) = \varepsilon_t, \tag{3}$$

where $\Pi(B)=\sum_{j=0}^{\infty}\pi_j B^j$, B is the backshift operator such that $By_t = y_{t-1}$. This representation is familiar in time series analysis with independent perturbations ε_t. A widely used model for capturing both short- and long-range dependencies, is the ARFIMA(p,d,q) model proposed by Granger and Joyeux [10] and Hosking [11]. In this case,

$$\Pi(B) = \Phi(B)(1-B)^d \Theta(B)^{-1}, \tag{4}$$

where $\Phi(B)=1-\phi_1 B - \cdots - \phi_p B^p$, $\Theta(B)=1-\theta_1 B - \cdots - \theta_q B^q$, d is the fractional differencing parameter and $(1-B)^d = \sum_{j=0}^{\infty} b_j B^j$ with $b_0=1$ and $b_j = \Gamma(j-d)/\{\Gamma(j+1)\Gamma(-d)\}$ for $j \geqslant 1$.

Thus, model (1)–(2) with parametrization (4) satisfies

$$\Phi(B)(1-B)^d (y_t - \mu) = \Theta(B)\varepsilon_t, \tag{5}$$

$$\varepsilon_t|\mathscr{F}_{t-1} \sim (0, g(\lambda_t)). \tag{6}$$

To ensure the causality and the invertibility of the filter (4), it is assumed that the polynomials $\Phi(B)$ and $\Theta(B)$ have no common roots. These are all located outside the unit circle, and $|d|<\frac{1}{2}$.

Note that in (5), although the sequence $\{\varepsilon_t\}$ is not a strict white noise (independent), it is uncorrelated under some conditions, cf. Lemma 1. Therefore, the process $\{y_t\}$ and the ARFIMA model with independent input error sequence $\{\varepsilon_t\}$ have the same autocorrelation function, as stated in Theorem 2. For convenience, a conditional long-memory process (1)–(2) with an ARFIMA(p,d,q) filter will be referred to as CLM–ARFIMA process. When $\Pi(B)=1-\phi B$ in (3), we obtain the model proposed by Grunwald *et al.* [12].

3.2. Properties

The following theorems, which are proved in the technical appendix, establish some properties of the conditional long-memory process defined in Section 3.1. These results are related to second-order stationarity, named simply as stationarity, and correlation structure of the process $\{y_t\}$. In what follows, we will focus on processes $\{y_t\}\subset\mathbb{R}_+$. This corresponds to a technical condition required for an application of the monotone convergence theorem to show the mean-stationarity of the process, see proof of Theorem 1.

Theorem 1
Consider a process $\{y_t\}\subset\mathbb{R}_+$ defined by (5)–(6) where $\pi_j\leqslant 0$ for $j\geqslant 1$. Then, the process $\{y_t, t\in\mathbb{Z}\}$ is mean-stationary for any positive function g.

Theorem 2
Consider the CLM (5)–(6) with parametrization (4) and assume that the roots of $\Phi(z)=0$ lie outside the unit circle. Then, under conditions of Theorem 1,

(a) If g is a positive concave function, then $\{\varepsilon_t\}$ is an innovation sequence.
(b) Assume that $\mathrm{Var}(\varepsilon_t)$ is a finite constant. If $d<0.5$,

$$\mathrm{Var}(y_t)=\mathbf{E}\{g(\lambda_t)\}\sum_{j=0}^{\infty}\psi_j^2,$$

where $\Psi(B)=\Phi(B)^{-1}(1-B)^{-d}\Theta(B)=\sum_{j=0}^{\infty}\psi_j B^j$.
(c) Under conditions of part (b), the CLM–ARFIMA(p,d,q) process and the standard ARFIMA(p,d,q) process have the same autocorrelation function.
(d) Under conditions of parts (a) and (b), the process $\{y_t, t\in\mathbb{Z}\}$ is stationary.

In terms of modeling the data discussed in Section 2, Theorem 2 can be used as follows: specify g as a positive concave function and then, by part (a) $\{\varepsilon_t\}$ is an innovation sequence. Therefore, by part (c), both CLM-ARFIMA and standard ARFIMA processes share the same correlation structure. Thus, the identification stage for both types of time series models are similar.

3.3. Examples

The setup (5)–(6) is general enough to allow for modeling data with diverse distributions. For instance, the conditional distribution G may belong to the exponential family with support in $\mathbb{R}_+$ such as Binomial, Gamma or Poisson distributions. In what follows, we discuss briefly these examples corresponding to continuous and discrete conditional distributions G, respectively.

(a) *Conditional Poisson*: Define the model as $y_t|\mathscr{F}_{t-1}\sim\mathrm{Poi}(\lambda_t)$ where $\mathscr{Y}=\{0,1,2,\ldots\}$, η is null and $g(\lambda_t)=\lambda_t$. In this case, $\mathbf{E}\{g(\lambda_t)\}=\mathbf{E}[\lambda_t]=\mu$.
(b) *Conditional Binomial*: Consider the model $y_t|\mathscr{F}_{t-1}\sim\mathrm{Bin}(m,p_t)$ where n is fixed, $\mathscr{Y}=\{0,1,2,\ldots\}$, η is null, $\lambda_t=mp_t$ and $g(\lambda_t)=\lambda_t(m-\lambda_t)/m$. In this case, g is concave and bounded by $m/4$.
(c) *Conditional Gamma*: Let $y_t|\mathscr{F}_{t-1}\sim\mathrm{Gamma}(\lambda_t/\beta,\beta)$, with $\eta=\beta>0$, $\mathscr{Y}=(0,\infty)$ and $g(\lambda_t)=\beta\lambda_t$. For this distribution we have $\mathbf{E}\{g(\lambda_t)\}=\beta\mathbf{E}[\lambda_t]$.

As stated in Theorems 1 and 2, the correlation structure depends on both, the parametrization and the function g. Two parameterizations that satisfy conditions of Theorem 1 are

CLM–ARFIMA(0,d,0) with $d\in(0,\frac{1}{2})$. Here, $\Pi(B)=(1-B)^d$ is such that $\sum_{j=0}^{\infty}\pi_j=0$ and $\pi_j\leqslant 0$, $\forall j\geqslant 1$.
CLM–ARFIMA(1,d,0). Here $\pi_j=\pi_{j-1}\{(1-\phi)-(1+d)/j\}/\{1-\phi(j-1)/(j-2-d)\}$ for $j\geqslant 1$. In this case, it can be shown that $\pi_j\leqslant 0$ for $j=1,2,\ldots$, if $d\in(0,\frac{1}{2})$ and $-d\leqslant\phi\leqslant(1-d)/2$.

In addition, note that in the previous examples (a),(b) and (c), $g(\cdot)$ is a concave function in λ_t. Therefore, by Theorem 2 (a), $\varepsilon_t = y_t - \lambda_t$ is an innovation sequence. Hence, the time series described in examples above have the same autocorrelation structure of an ARFIMA process with independent noise.

4. Fitting the data

In this section, we discuss a procedure for modeling and forecasting the stationary model (1)–(2) with parametrization (4). For illustration purposes we regard G as a member of the exponential family, but the procedure can be adapted to handle other conditional distributions. In order to assess the performance of estimation methodology, some Monte Carlo experiments are presented in Section 4.2.

4.1. Estimation

Assume that the time series data $\{y_1, \ldots, y_n\}$ are generated by model (1)–(2) with parametrization (4). The vector of unknown parameters is denoted by $\theta = (\mu, \delta, \eta)'$ where μ is the level, $\delta = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, d)'$ and η is associated with the conditional distribution G as described in Section 3.1.

First, we estimate the mean of the processes. A simple estimator for the level of the process is the arithmetic mean, $\hat{\mu}_n = (1/n)\sum_{t=1}^{n} y_t$. As shown by the next lemma, under some mild conditions, this estimate is consistent.

Lemma 2
Consider the process described by (1)–(2) with $0 < d < \frac{1}{2}$. If $g(\cdot)$ is a positive concave function, then $\hat{\mu}_n$ is a $n^{1/2-d}$ consistent estimate of μ.

Observe that although this estimate is consistent, its asymptotic distribution may not be Gaussian, cf. Brockwell and Davis [13, p. 527].

Once the mean μ is estimated by $\hat{\mu}_n$, the parameters δ and η may be estimated by using the maximum likelihood method. For computing the likelihood, we replace μ by $\hat{\mu}_n$. The *conditional* pseudo log-likelihood, see Box *et al.* [14, pp. 226–28], is given by

$$L(\delta, \eta) = \sum_{t=2}^{n} l_t, \tag{7}$$

where $l_t = \log f_\theta(y_t|\mathscr{F}_{t-1})$ and the contribution of the first observation, usually negligible for long time series, has been removed. Given that the conditional distribution G is a member of the exponential family we write

$$f_\theta(y_t|\mathscr{F}_{t-1}) = a^*(\lambda_t, \eta)\psi^*(y_t)\exp\left\{\sum_{i=1}^{m} b_i(\lambda_t, \eta)R_i(y_t)\right\}, \tag{8}$$

where the functions $a^*(\cdot)$ and $b_i(\cdot)$ depend on the information $\mathscr{F}_{t-1}$ only through λ_t and the functions ψ^* and R_i do not depend on δ and η. Then,

$$l_t = a(\lambda_t, \eta) + \psi(y_t) + \sum_{i=1}^{m} b_i(\lambda_t, \eta)R_i(y_t), \tag{9}$$

where $a = \log(a^*)$ and $\psi = \log(\psi^*)$. In order to obtain the maximum likelihood estimator $(\hat{\delta}, \hat{\eta})$ and its precision, we need to calculate the *score* and the Hessian. The *score* is given by

$$\frac{\partial L(\delta, \eta)}{\partial(\delta, \eta)} = \left(\sum_{t=2}^{T} \frac{\partial l_t}{\partial \delta}, \sum_{t=2}^{T} \frac{\partial l_t}{\partial \eta}\right)', \tag{10}$$

where from (9),

$$\frac{\partial l_t}{\partial \delta} = \frac{\partial a(\lambda_t, \eta)}{\partial \delta} + \sum_{i=1}^{m} \frac{\partial b_i(\lambda_t, \eta)}{\partial \delta} R_i(y_t), \tag{11}$$

$$\frac{\partial l_t}{\partial \eta} = \frac{\partial a(\lambda_t, \eta)}{\partial \eta} + \sum_{i=1}^{m} \frac{\partial b_i(\lambda_t, \eta)}{\partial \eta} R_i(y_t). \tag{12}$$

Expressions $\partial a(\lambda_t,\eta)/\partial\delta$ and $\partial b_i(\lambda_t,\eta)/\partial\delta$ are functions of $\partial\lambda_t/\partial\delta$, which in turn depend on the model through the sequence $\{\pi_j\}$. For instance, from (2),

$$\frac{\partial\lambda_t}{\partial\delta}=\mu\sum_{j=0}^{\infty}\frac{\partial\pi_j}{\partial\delta}-\sum_{j=1}^{\infty}y_{t-j}\frac{\partial\pi_j}{\partial\delta}. \tag{13}$$

The following examples illustrate the form of π_j and $\partial\pi_j/\partial\delta$ for two useful parameterizations:

(a) *Example 1*: for a CLM–ARFIMA$(0,d,0)$ model.
In this case $\delta=d$, $\partial\pi_1/\partial\delta=-1$ and $\pi_j=\pi_{j-1}(j-d-1)/j$ for $j\geqslant 2$. Then,

$$\frac{\partial\pi_j}{\partial\delta}=\frac{\partial\pi_{j-1}}{\partial\delta}\left(\frac{j-d-1}{j}\right)-\frac{\pi_{j-1}}{j}. \tag{14}$$

(b) *Example 2*: for a CLM–ARFIMA$(1,d,0)$ model.
Here $\delta=(\phi,d)$ and from Hosking [11], $\pi_j=\{(1-\phi)-(1+d)/j\}\Gamma(j-d-1)/\{\Gamma(j)\Gamma(-d)\}$ for $j\geqslant 1$. Then,

$$\pi_j=\pi_{j-1}A_j, \tag{15}$$

$$A_j=\left(\frac{j-d-2}{j}\right)\left\{\frac{j(1-\phi)-(1+d)}{(j-1)(1-\phi)-(d+1)}\right\}. \tag{16}$$

Hence, $\partial\pi_1/\partial d=-1$ and

$$\frac{\partial\pi_j}{\partial d}=\pi_{j-1}\frac{\partial A_j}{\partial d}+A_j\frac{\partial\pi_{j-1}}{\partial d},\qquad j\geqslant 2, \tag{17}$$

where after some algebra,

$$\frac{\partial A_j}{\partial d}=\left(\frac{j-d-2}{j}\right)\left[\frac{(1-\phi)}{\{(j-1)(1-\phi)-(d+1)\}^2}\right]-\left[\frac{j(1-\phi)-(1+d)}{j\{(j-1)(1-\phi)-(d+1)\}}\right],\qquad j\geqslant 2. \tag{18}$$

In addition, $\partial\pi_1/\partial\phi=-1$ and

$$\frac{\partial\pi_j}{\partial\phi}=\frac{\partial\pi_{j-1}}{\partial\phi}\left(\frac{j-d-2}{j-1}\right),\quad j\geqslant 2. \tag{19}$$

Finally,

$$\frac{\partial\pi_j}{\partial\delta}=\left(\frac{\partial\pi_j}{\partial\phi},\frac{\partial\pi_j}{\partial d}\right)' \tag{20}$$

where $\partial\pi_j/\partial\phi$ and $\partial\pi_j/\partial d$ are defined in (19) and (17), respectively.

In practice, only n observations $y_1,\ldots,y_n$ are available. But λ_t and the *score* depend on the infinite past of the process $\{y_t\}$. Therefore, the following approximations may be used:

$$\lambda_t\approx\mu\sum_{j=0}^{t-1}\pi_j-\sum_{j=1}^{t-1}\pi_j y_{t-j}, \tag{21}$$

$$\frac{\partial\lambda_t}{\partial\delta}\approx\mu\sum_{j=0}^{t-1}\frac{\partial\pi_j}{\partial\delta}-\sum_{j=1}^{t-1}\frac{\partial\pi_j}{\partial\delta}y_{t-j}. \tag{22}$$

Furthermore, in order to maximize the log-pseudo-likelihood, the Davidon–Fletcher–Powell algorithm is used in this work, see Press *et al.* [15, Ch.10]. This is an attractive scheme because it avoids calculating the Hessian matrix at each iteration of the optimization procedure. A single evaluation of the Hessian is needed to obtain the parameter variance–covariance matrix.

4.2. Simulations

In order to gain some insight on the finite sample performance of the quasi maximum likelihood estimation (QMLE) described in Section 4.1, several Monte Carlo experiments were carried out. Table I shows simulation results from a Poisson conditional distribution, with $\mu=10$, several values of the long-memory parameter d for fractional noise processes with filter $\Pi(B)=(1-B)^d$ and two sample sizes $n=200$ and $n=800$. The simulations results are based on 1000 repetitions.

From Table I, observe that the estimates are reasonably close to the true parameter for both μ and d. For a linear Gaussian ARFIMA process, the theoretical standard deviation of $\hat{d}$ is 0.055 for $n=200$ and 0.028 for $n=800$, not depending on d. These values are close to the results obtained from the simulations. Note that for fixed n, the estimated standard deviations $\sigma(\hat{d})$ seem to be similar across the different values of the long-memory parameter d. This is particularly noticeable for the case $n=800$, suggesting that the behavior of the estimation standard error for the conditional Poisson model seems to be similar to the standard Gaussian case. However, some asymptotic properties of the estimates for CLM, such as consistency and normality, have not been formally established yet. A difficulty with the asymptotic analysis of these non-Gaussian processes is that the noise sequence is not necessarily independent. Thus, standard asymptotic results, see for example Section 5.2 of Taniguchi and Kakizawa [16], are not applicable. On the other hand, Hosoya [17] relaxes the independence assumption, replacing it by mixing conditions on $\{\varepsilon_t\}$. However, it seems that these properties have not been established for the innovations from a CLM satisfying (1)–(2). Another path is provided by Ling and Li [18]. They proved a central limit theorem for the maximum likelihood estimate for conditionally heteroscedastic long-memory time series. In this case, the estimate might have an asymptotic variance greater than the non-conditional Gaussian long-memory case, as suggested by Li *et al.* [19].

In order to assess the quality of QMLE estimators in the ARFIMA$(1,d,0)$ parametrization, we conducted several Monte Carlo simulations with two sample sizes, $n=500$ and $n=1000$. The main purpose of this study is to evaluate the performance of the method for autoregressive parameters ϕ closer to the region of non-stationarity. Recall that from Section 3.3, $-d\leqslant\phi\leqslant(1-d)/2$. In addition, we calculated the Haslett and Raftery estimator implemented in the *R* package and the Whittle estimator for these simulated series. The results are summarized in Tables II and III. For $n=500$, the QMLE are slightly biased. Compared with QMLE, the Whittle estimators present more bias for ϕ and almost the same bias for d. However, the QMLE is much better in terms of standard deviation and MSE. In addition, the Haslett and Raftery method is highly biased. For $n=1000$ and compared with Table II, for each combination the bias of QMLE estimators is small. Compared with QMLE, the Whittle estimators present better results in terms of bias, almost the same performance in terms of RMSE for the parameter combinations ($d=0.4$, $\phi=0.27$) and ($d=0.45$, $\phi=0.25$). However, for the combination ($d=0.2$, $\phi=0.35$), the QMLE is better in terms of RMSE. Compared with Table II, the performance of the Haslett and Raftery estimator improves in both bias and RMSE.

4.3. Modeling and diagnostics

As observed by Tsay [20, p. 161], nonlinear time series modeling involves both the experience of the analyst and the type of the problem under study. In practice, the nature of the variables defines the sample space $\mathscr{Y}$. For instance, counting processes lead naturally to discrete positive data. Besides, the distribution of data can be specified by means of tools such as histograms and q–q plots. In some cases, the conditional distribution defines the form of $g(\lambda_t)$, for example, for a Poisson distribution, $g(\lambda_t)=\lambda_t$. But, in other situations we have some flexibility when defining $g(\lambda_t)$. The task of determining the conditional variance can be helped by observing the patterns of data and correlation structure from simulated time series. The sample autocorrelation function of both the observations and their squares may give some clues about the underlying dependence structure of the data, see for instance Baillie and Chung [21]. Furthermore, the residuals $e_t=y_t-\hat{\lambda}_t$ can be used for assessing the goodness of fit, by checking the absence of correlation on the residuals.

Table I. Monte Carlo experiments for conditional long-memory Poisson models.

	$n=200$				$n=800$			
d	$\hat{\mu}$	$\hat{d}$	$\sigma(\hat{d})$	RMSE$(\hat{d})$	$\hat{\mu}$	$\hat{d}$	$\sigma(\hat{d})$	RMSE$(\hat{d})$
0.10	10.03	0.087	0.049	0.051	10.00	0.094	0.029	0.030
0.20	10.04	0.186	0.057	0.059	10.01	0.194	0.030	0.030
0.30	10.02	0.278	0.062	0.066	10.03	0.295	0.030	0.030
0.40	10.10	0.379	0.061	0.064	10.03	0.395	0.030	0.030
0.45	10.00	0.426	0.057	0.062	10.09	0.445	0.028	0.029

Applied Stochastic Models in Business and Industry

Table II. Monte Carlo experiments. One thousand replications of CLM-ARFIMA(1,d,0) Poisson time series of size $n=500$.

Parameters			QMLE		HR		Whittle	
d	ϕ		$\widehat{d}$	$\widehat{\phi}$	$\widehat{d}$	$\widehat{\phi}$	$\widehat{d}$	$\widehat{\phi}$
0.20	0.35	Mean	0.1559	0.3839	0.1375	0.4029	0.1429	0.4009
		Bias	−0.0441	0.0339	−0.0625	0.0529	−0.0571	0.0509
		SD	0.0845	0.0947	0.0941	0.1101	0.1133	0.1251
		RMSE	0.0953	0.1006	0.1129	0.1221	0.1268	0.1350
0.40	0.27	Mean	0.3577	0.3103	0.3134	0.3541	0.3656	0.3131
		Bias	−0.0423	0.0403	−0.0866	0.0841	−0.0344	0.0431
		SD	0.0865	0.0986	0.1045	0.1191	0.1279	0.1353
		RMSE	0.0962	0.1065	0.1357	0.1457	0.1324	0.1420
0.45	0.25	Mean	0.4042	0.2957	0.3440	0.3549	0.4061	0.3064
		Bias	−0.0458	0.0457	−0.1060	0.1049	−0.0439	0.0564
		SD	0.0804	0.0950	0.1149	0.1326	0.1276	0.1355
		RMSE	0.0925	0.1054	0.1562	0.1690	0.1349	0.1467

Table III. Monte Carlo experiments. One thousand replications of CLM-ARFIMA(1,d,0) Poisson time series of size $n=1000$.

Parameters			QMLE		HR		Whittle	
d	ϕ		$\widehat{d}$	$\widehat{\phi}$	$\widehat{d}$	$\widehat{\phi}$	$\widehat{d}$	$\widehat{\phi}$
0.20	0.35	Mean	0.1756	0.3706	0.1684	0.3777	0.1758	0.3731
		Bias	−0.0244	0.0206	−0.0316	0.0277	−0.0242	0.0231
		SD	0.0617	0.0704	0.0680	0.0781	0.0771	0.0856
		RMSE	0.0663	0.0733	0.0749	0.0828	0.0808	0.0886
0.40	0.27	Mean	0.3744	0.2953	0.3565	0.3117	0.3956	0.2797
		Bias	−0.0256	0.0253	−0.0435	0.0417	−0.0044	0.0097
		SD	0.0599	0.0710	0.0598	0.0713	0.0693	0.0780
		RMSE	0.0651	0.0753	0.0739	0.0825	0.0694	0.0785
0.45	0.25	Mean	0.4225	0.2781	0.3973	0.3017	0.4427	0.2648
		Bias	−0.0275	0.0281	−0.0527	0.0517	−0.0073	0.0148
		SD	0.0573	0.0695	0.0598	0.0703	0.0619	0.0696
		RMSE	0.0635	0.0750	0.0797	0.0872	0.0623	0.0711

4.4. Prediction

A major goal in the statistical analysis of time series concerns with forecasting. For the class of CLMs defined in Section 3, the distribution of the process $\{y_t\}$ conditional on the past information, $\mathscr{F}_{t-1}$, has mean λ_t. Therefore a natural one-step predictor of λ_t is $\hat{\lambda}_t$, which is based on (21),

$$\hat{\lambda}_t=\hat{\mu}\sum_{j=0}^{t-1}\hat{\pi}_j-\sum_{j=1}^{t-1}\hat{\pi}_j y_{t-j}, \tag{23}$$

where each $\hat{\pi}_j$ depends on the parameter estimates $\hat{\delta},\hat{\eta}$. Hence, the estimate conditional distribution is $y_t|\mathscr{F}_{t-1}\sim G(\hat{\lambda}_t, g(\hat{\lambda}_t))$ and the construction of conditional prediction intervals for one-step forecasts is a simple task.

5. Results

The trading volumes and the varve sedimentary data presented in Section 2 are analyzed next by means of the methodology proposed in Section 4.

5.1. Stock-market trading volume data

Let us begin by modeling the time series of counts consisting of IBM daily trading volumes presented in Section 2. This series has 2024 observations, with mean 7 174 449 and standard deviation 3 773 536.

The results from the exploratory analysis, see Figure 2, suggest that a candidate model for the data may be a conditional long-memory Poisson, see Example (a) in Section 3.3, with an ARFIMA parametrization. The order of the model, (1,d,0), was selected using AIC and the estimate obtained is $\widehat{d}=0.278$ with standard deviation SD($\widehat{d}$)$=0.033$, and $\widehat{\phi}=0.231$ with standard deviation SD($\widehat{\phi}$)$=0.042$. Therefore, the corresponding t-tests are $t_d=8.42$ and $t_{\phi}=5.50$, both highly significant.

From Figures 5–6, the fitted conditional mean series closely resembles the evolution of the data. Besides, the sample autocorrelation function of the residuals, see Figure 7, indicates almost no correlation. Note, however, that other whiteness tests could be further considered for these non-Gaussian residuals, as for example the robustified Portmanteau tests by Escanciano and Lobato [22]. Prediction bands at the level 95% for one-step volume forecasts are given in Figures 8–9. They are based on the Poisson distribution with 2.5% each tail. Note that the time-varying confidence prediction bands allow us to capture the time series evolution. Moreover, as expected, most of the observations are included by these confidence bands

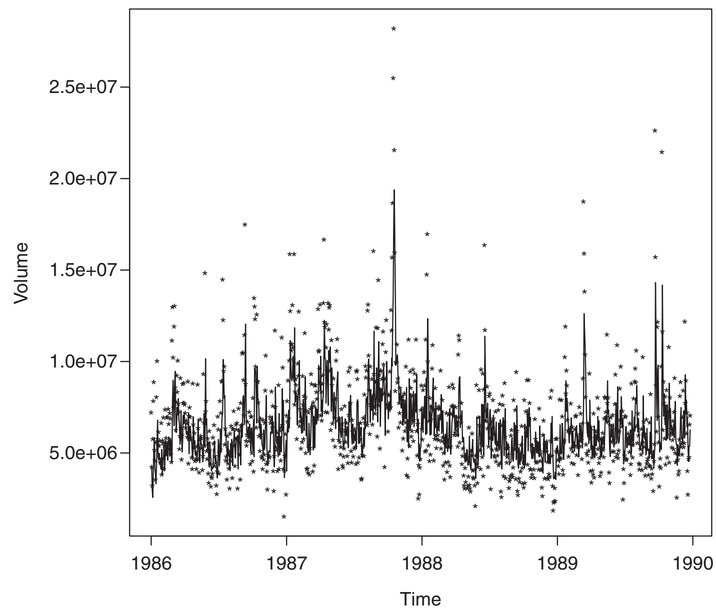

Figure 5. IBM daily trading volume data and fitted conditional mean. Period 1986–1989.

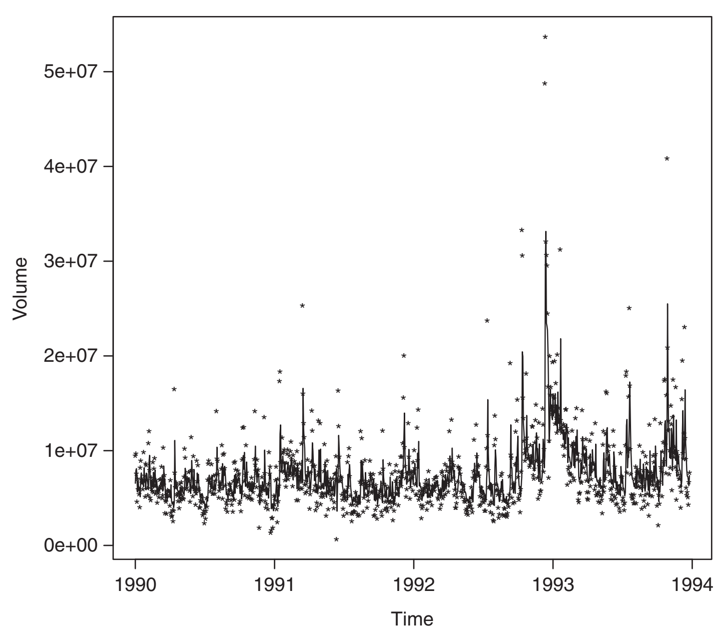

Figure 6. IBM daily trading volume data and fitted conditional mean. Period 1990–1993.

Applied Stochastic Models in Business and Industry

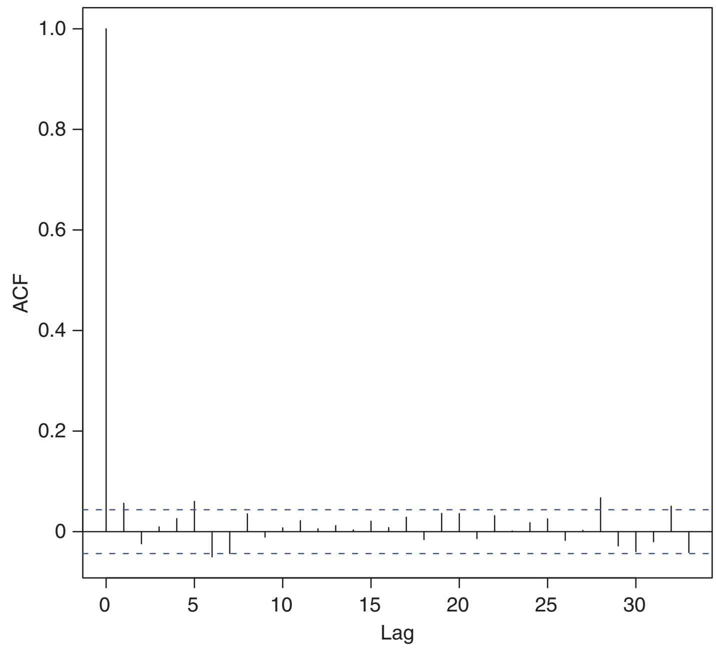

Figure 7. IBM daily trading volume data: sample autocorrelation function of residuals.

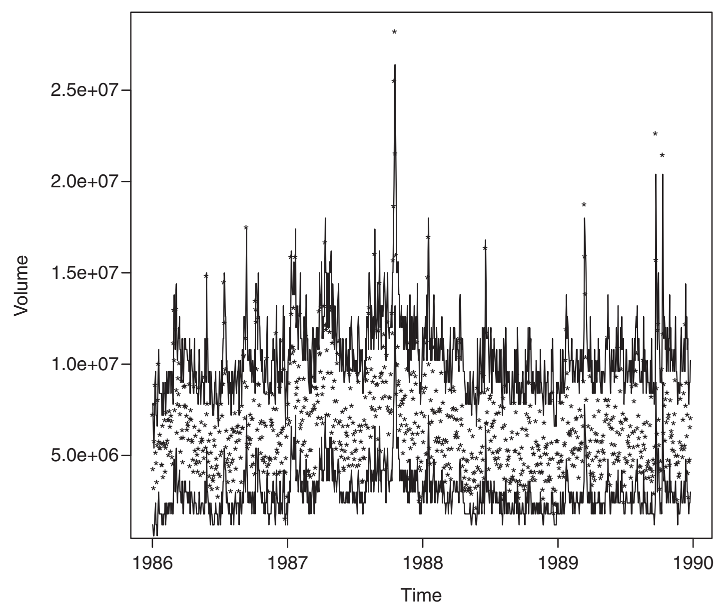

Figure 8. IBM daily trading volume data: one-step conditional prediction bands (95%). Period 1986–1989.

(93%). Thus, the number of observations included within the predictions bands is close to its theoretical value of 95%. On the whole, these results permit to conclude that the fit is very good.

5.2. Glacial varves data

In this case we are dealing with a continuous variable and given the features of Figures 3 and 4, a CLM-ARFIMA Gamma, see Example (c) in Section 3.3, with parametrization $(0,d,0)$ selected by AIC is proposed for this time series data.

The maximum likelihood estimates are given in Table IV. As observed, both estimates are highly significant. The fitted conditional mean series provided in Figure 10 describes the evolution of the data quite well and the empirical autocorrelation function of residual innovations, see the left panel of Figure 11, shows no significant correlation.

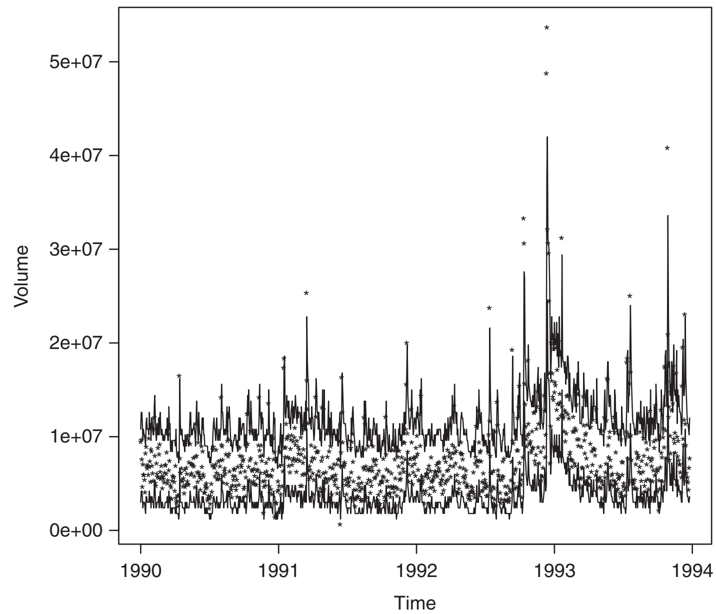

Figure 9. IBM daily trading volume data: one-step conditional prediction bands (95%). Period 1990–1993.

Table IV. Estimates for glacial varve data.

Parameter	Estimate	S.D.	t
d	0.337	0.0262	12.86
β	0.159	0.0086	18.49

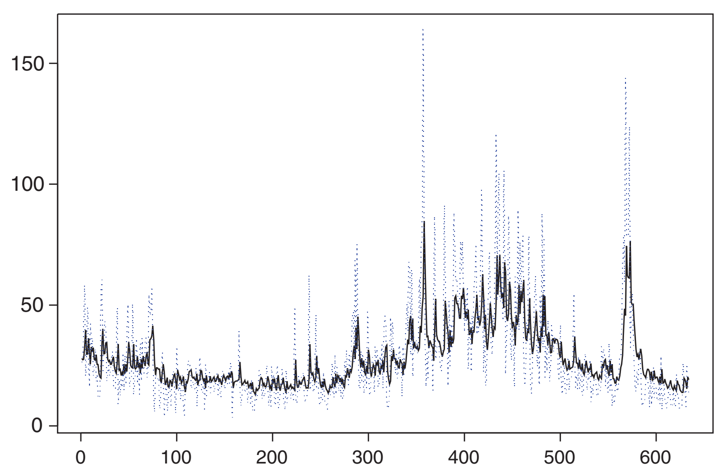

Figure 10. Varve data (dotted line) and fitted conditional mean (heavy line).

On the other hand, since log-varve series exhibits a symmetric histogram, see the right panel of Figure 11, it is illustrative to compare our analysis with the linear ARFIMA Gaussian methodology applied to the transformed data. Thus, Shumway and Stoffer [4] obtain the approximated Gaussian maximum likelihood estimate $\hat{d} = 0.384$. However, the corresponding standard error is not provided by those authors. In addition, by applying the state-space methodology described in Chan and Palma [23] we obtain the estimate $\hat{d} = 0.3878$ with standard deviation $\mathrm{SD}(\hat{d}) = 0.0133$ and Student statistic $t_d = 14.2$ being highly significant. After obtaining by exponentiation the one-step predictors in the original scale, the estimated innovation variance is 242.76. This is slightly bigger than the one obtained from the CLM methodology, 240.51. Therefore, in terms of prediction something is gained. However, the main advantage of the proposed methodology is that no transformation of the data is required since the observations are directly modeled.

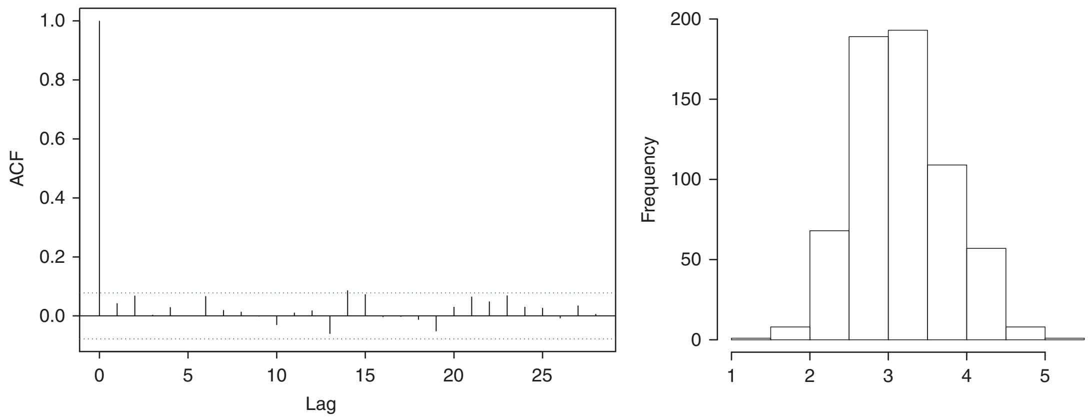

Figure 11. Varve data: autocorrelation function of varve residuals (left panel). Histogram of the log-transformed data (right panel).

6. Conclusions

In this paper, we have analyzed two types of non-Gaussian time series data exhibiting long-range dependence. To this end, we have proposed a family of CLMs, which are capable of handling data with those features. As shown in the results in Section 5, these models seem to fit the data under study very well. Furthermore, the methodology applied to these observations can be a useful tool for developing point forecasts and prediction bands. It is worth noting that unlike most of the currently available techniques, the methodology applied in this work does not require transformation of the data. As it is well known from the time series literature, transforming the data may create problems with the interpretation of the fitted model and potentially inappropriate predictions bands.

Appendix A

Proof of Lemma 1
(a) Since $\varepsilon_t = y_t - \mathbf{E}(y_t|\mathscr{F}_{t-1})$, $\mathbf{E}(\varepsilon_t)=0$ and $\mathrm{Var}[\varepsilon_t]=\mathbf{E}[\mathbf{E}\{(y_t-\lambda_t)^2\}|\mathscr{F}_{t-1}]=\mathbf{E}\{\mathbf{E}(y_t^2|\mathscr{F}_{t-1})-\lambda_t^2\}=\mathbf{E}\{g(\lambda_t)\}$. (b) Since $\mathbf{E}(\varepsilon_t)=0$, $\mathrm{cov}(\varepsilon_t,\varepsilon_{t-k})=\mathbf{E}(\varepsilon_t\varepsilon_{t-k})$. Now,

$$\mathbf{E}(\varepsilon_t\varepsilon_{t-k})=\mathbf{E}\{\mathbf{E}(\varepsilon_t\varepsilon_{t-k}|\mathscr{F}_{t-k})\}=\mathbf{E}\{\varepsilon_{t-k}\mathbf{E}(\varepsilon_t|\mathscr{F}_{t-k})\},$$

but, as shown below, $\mathbf{E}(\varepsilon_t|\mathscr{F}_{t-k})=0$ for all $k\geqslant 1$. Consequently, $\mathbf{E}(\varepsilon_t\varepsilon_{t-k})=0$ and the result follows. The proof of $\mathbf{E}(\varepsilon_t|\mathscr{F}_{t-k})=0$ can be made by induction: (i) for $k=1$, $\mathbf{E}(\varepsilon_t|\mathscr{F}_{t-1})=0$ by the definition of ε_t. (ii) Then, it is assumed that $\mathbf{E}(\varepsilon_t|\mathscr{F}_{t-k})=0$ holds for some k. (iii) Now, for $k+1$, $\mathbf{E}(\varepsilon_t|\mathscr{F}_{t-(k+1)})=\mathbf{E}\{\mathbf{E}(\varepsilon_t|\mathscr{F}_{t-k})|\mathscr{F}_{t-(k+1)}\}$ because $\mathscr{F}_{t-k}\supset\mathscr{F}_{t-(k+1)}$ by properties of conditional expectation, see for example Durrett [24, p. 226]. Finally from (ii) $\mathbf{E}(\varepsilon_t|\mathscr{F}_{t-k})=0$ and therefore $\mathbf{E}(\varepsilon_t|\mathscr{F}_{t-(k+1)})=0$, as required. (c) This follows directly from (a) and (b). □

Proof of Theorem 1
Note that $\mathbf{E}(y_t)=\mathbf{E}[\mathbf{E}(y_t|\mathscr{F}_{t-1})]=\mathbf{E}(\lambda_t)$. But, $\lambda_t=\mu\sum_{j=0}^{\infty}\pi_j+\sum_{j=1}^{\infty}(-\pi_j)y_{t-j}$ and by taking expectation on both sides we have $\mathbf{E}[\lambda_t]=\mu\sum_{j=0}^{\infty}\pi_j+\mathbf{E}[\sum_{j=1}^{\infty}(-\pi_j)y_{t-j}]$. Now, since $\pi_j\leqslant 0$ for $j=0,1,\ldots,$ by Theorem 16.6 of [25] the expectation and the infinite sum can be commuted. Thus, $\mathbf{E}[\lambda_t]=\mu\sum_{j=0}^{\infty}\pi_j+\sum_{j=1}^{\infty}(-\pi_j)\mathbf{E}[y_{t-j}]$.

Let $a_t=\mathbf{E}[y_t]$. Then, we have $a_t=\mu\sum_{j=0}^{\infty}\pi_j+\sum_{j=1}^{\infty}(-\pi_j)a_{t-j}$. Since $\sum_{j=0}^{\infty}|\pi_j|<\infty$, we may write $\sum_{j=0}^{\infty}\pi_j(a_{t-j}-\mu)=0$, or equivalently,

$$\Phi(B)\Theta(B)^{-1}(1-B)^d(a_t-\mu)=0.$$

Since $\Phi(B)$, $\Theta(B)^{-1}$, $(1-B)^d$ are invertible filters for $|d|<\frac{1}{2}$, we obtain $a_t=\mu+c$ for all t, where c is an arbitrary constant. Thus, the process $\{y_t\}$ has constant finite expected value. For convention we will assume that the mean is μ, i.e. c is set to be zero. □

Proof of Theorem 2
(a) Since g is concave, $-g$ is convex. Therefore, by Jensen's inequality, $\mathbf{E}[g(\lambda_t)]\leqslant g(\mathbf{E}[\lambda_t])$. Then, by Theorem 1, $\mathbf{E}[\lambda_t]=\mu$ for all t and, hence, $\mathbf{E}[g(\lambda_t)]\leqslant g(\mu)<\infty$. Consequently, by Lemma 1(a) $\mathrm{Var}(\varepsilon_t)<\infty$, and by Lemma 1(c) $\{\varepsilon_t\}$ is an innovation

process. In addition, from expression (2.6) and Theorem 13.2.2 of Brockwell and Davis [13] we obtain parts (b) and (c). Finally, item (d) follows directly from (a)–(c). □

Proof of Lemma 2
Given that $g(\cdot)$ is a positive concave function, by Theorem 2, the process $\{y_t\}$ is stationary. Now, an application of Theorem 10.4(b) and Section 10.4 of Palma [1] yields $\mathrm{Var}(\hat{\mu}_n)\leqslant Cn^{2d-1}$. Hence, by the Chebyshev inequality, $\mathbb{P}(|\hat{\mu}_n-\mu|>\varepsilon)\leqslant Cn^{2d-1}/\varepsilon^2$, as required. □

Acknowledgements

This research was partially supported by Grant 1085239 from Fondecyt for the first author and grants from FAPESP and FAEPEX for the second author.

References

1. Palma W. *Long-memory Time Series*: *Theory and Methods*. Wiley: Hoboken, NJ, 2007.
2. Lobato I, Velasco C. Long-memory in stock-market trading volume. *Journal of Business and Economic Statistics* 2000; **18**:410–427.
3. Dmowska R, Saltzman B (eds). *Long-Range Persistence in Geophysical Time Series*. Advances in Geophysics, vol. 40. Academic Press: New York, 1999.
4. Shumway RH, Stoffer DS. *Time Series Analysis and Its Applications*. Springer: Berlin, 2000.
5. Brockwell A. Likelihood-based analysis of a class of generalized long-memory time series models. *Journal of Time Series Analysis* 2006; **28**:386–407.
6. Andersen TG. Return volatility and trading volume: an information flow interpretation of stochastic volatility. *Journal of Finance* 1996; **60**:169–204.
7. Stanislavsky AA, Burnecki K, Magdziarz M, Weron A, Weron K. FARIMA modeling of solar flare activity from empirical time series of soft x-ray solar emission. *The Astrophysical Journal* 2009; **693**:1877–1882.
8. Samorodnitsky G, Taqqu MS. *Stable Non-Gaussian Random Processes*. Chapman & Hall: New York, 1994.
9. Chechkin AV, Gonchar VYu. A model for persistent Levy motion. *Physica A*: *Statistical Mechanics and Its Applications* 2000; **277**:312–326.
10. Granger CWJ, Joyeux R. An introduction to long-range time series models and fractional differencing. *Journal of Time Series Analysis* 1980; **1**:15–30.
11. Hosking JRM. Fractional differencing. *Biometrika* 1981; **68**:165–176.
12. Grunwald GK, Hyndman RJ, Tedesco L, Tweedie RL. Non-Gaussian conditional linear AR(1) models. *The Australian & New Zealand Journal of Statistics* 2000; **42**:479–495.
13. Brockwell P, Davis R. *Time Series Theory and Methods* (2nd edn). Springer: New York, 1991.
14. Box GEP, Jenkins GM, Reinsel GC. *Time Series Analysis* (3rd edn). Prentice-Hall: Englewood Cliffs, NJ, 1994.
15. Press WH, Teukolski SA, Vetterling WT, Flannery BP. *Numerical Recipes. The Art of Scientific Computing* (2nd edn). Cambridge University Press: Cambridge, 1992.
16. Taniguchi M, Kakizawa Y. *Asymptotic Theory of Statistical Inference for Time Series*. Springer: New York, 2000.
17. Hosoya Y. Limit theory with long-range dependence and statistical inference of related models. *Annals of Statistics* 1997; **25**:105–137.
18. Ling S, Li WK. On fractionally integrated autoregressive time series models with conditionally heteroscesdaticity. *Journal of the American Statistical Association* 1997; **92**:1184–1194.
19. Li WK, Ling S, McAleer M. A survey of recent theoretical results for time series models with GARCH errors. *Journal of Economic Survey* 2002; **16**:245–269.
20. Tsay RS. *Analysis of Financial Time Series*. Wiley: New York, 2002.
21. Baillie RT, Chung H. Estimation of GARCH models from the autocorrelations of the squares of a process. *Journal of Time Series Analysis* 2001; **22**:631–650.
22. Escanciano JC, Lobato IN. An automatic Portmanteau test for serial correlation. *Journal of Econometrics* 2009; **151**:140–149.
23. Chan NH, Palma W. State-space modeling of long-memory processes. *Annals of Statistics* 1998; **26**:719–740.
24. Durrett R. *Probability*: *Theory and Examples* (2nd edn). Duxbury Press: Belmont, CA, 1996.
25. Billingsley P. *Probability and Measure*. Wiley: New York, 1995.