Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics and Data Analysis 52 (2008) 4487–4501
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Assessing influence in Gaussian long-memory models

Wilfredo Palma[a], Pascal Bondon[b,*], José Tapia[c]

[a] Pontificia Universidad Católica de Chile, Chile
[b] CNRS UMR 8506, Université Paris XI, France
[c] Universidad Adolfo Ibáñez, Chile

Received 28 April 2007; received in revised form 10 January 2008; accepted 11 January 2008
Available online 10 March 2008

Abstract

A statistical methodology for detecting influential observations in long-memory models is proposed. The identification of these influential points is carried out by case-deletion techniques. In particular, a Kullback–Leibler divergence is considered to measure the effect of a subset of observations on predictors and smoothers. These techniques are illustrated with an analysis of the River Nile data where the proposed methods are compared to other well-known approaches such as the Cook and the Mahalanobis distances.

1. Introduction

Influence assessment has been widely discussed in the linear model setting; see, for example, Belsley et al. (1980), Thomas and Cook (1990) and Lawrance (1995). On the other hand, there has been an increasing interest in influence diagnostics and outlier identification in the time series context; see, for instance, Bruce and Martin (1989), Peña (1991), Lefrançois (1991), Tsay et al. (2000), Cavanaugh and Oleson (2001), Choy (2001), and Cai and Davies (2003), among others. Furthermore, Kohn and Ansley (1987), Harrison and West (1991), Cavanaugh and Johnson (1999) and Proietti (2003) have investigated influence measures in state-space models. In particular, Kohn and Ansley (1987) extend the concept of leverage to the state-space framework and propose definitions for studentized and deleted residuals, while Harrison and West (1991) introduce an influence diagnostic in a Bayesian framework. Furthermore, following the work of Johnson and Geisser (1983), Cavanaugh and Johnson (1999) introduce a Kullback–Leibler-type influence measure on the smoothers for finite state-space models using fixed-interval smoothing and the expectation-maximization algorithm for parameter estimation.

Long-memory processes have been widely used to model long-range dependent observations in several disciplines; see, for example, Doukhan et al. (2003) and Palma (2007) for examples from hydrology to economics. For short-memory models, the influence of an observation can be expected to be only local. However, in long-memory models, the influence of a single or a group of observations might be profound and far-reaching. Consequently, there is particular interest in examining influential observations in long-memory models. There are several difficulties in

* Corresponding address: CNRS UMR 8506, Universite Paris XI, 3 rue Joliot-Curie, 91192 Gif-sur-Yvette, France. Tel.: +33 1 69 85 17 31; fax: +33 1 69 41 30 60.
E-mail address: bondon@lss.supelec.fr (P. Bondon).

0167-9473/$ - see front matter
doi:10.1016/j.csda.2008.01.030

implementing diagnostic measures in the analysis of long-range dependent data. Indeed, as pointed out by Beran (1994) and Doornik and Ooms (2003), calculation of estimates or predictors is highly computationally demanding. This is particularly cumbersome when implementing case-deletion techniques since the calculation of estimates or predictors must be carried out many times. A closely related problem is the need for appropriate methods for dealing with missing data. In this work, we solve these problems by developing efficient state-space techniques which have the additional advantage of allowing the easy handling of missing values.

The main contributions of this paper are the proposal and implementation of influence measures for long-memory processes. These influence measures assess the effect of one or several observations on the forecasts and the smoothers from state-space systems representing long-memory models. In contrast with other measures like the Mahalanobis distance or the Cook's distance, the Kullback–Leibler influence measures are based on different inferential objectives, that is, the recovery of the smoothed representation of the series and the forecasts of future values. Thus, the diagnostic techniques discussed are complementary, since they focus on distinct aspects of the analysis. The proposed measures are implemented by using Kalman filter equations for prediction, filtering and fixed point smoothing along with Newton-Raphson optimization for parameter estimation. The performance of the proposed influence measures is illustrated by means of a real-life example. *S-plus* and *Fortran* codes for carrying out the computations are available from the authors upon request.

The paper is structured as follows. Kullback–Leibler measures of the influence of a subset of observations on predictors and state-space smoothers are discussed in Sections 1.1 and 1.2. Since all state-space representations of long-range dependent processes are infinite dimensional and good finite approximations require a large state vector, we address the problem of reducing the dimensionality in Section 1.3. In Section 2, we apply the general methodology discussed in Section 1 for assessing influence on state-space representations of autoregressive fractional integrated moving average (ARFIMA) models which are an important class of long-memory processes. In Section 3, the detection techniques are illustrated by means of an application to the well-known River Nile data. Conclusions are presented in Section 4, and technical results are summarized in the Appendix.

1.1. Influence of cases on time-series predictors

Let $(Y_t)_{t\geq 1}$ be a discrete-time zero-mean Gaussian stationary time series, $Y = (Y_1, \ldots, Y_n)$ where $n \geq 1$, and let Y^K represent Y with cases indexed by $K = \{t_1, t_2, \ldots, t_k\}$ omitted (treated as missing values) where $1 \leq t_1 < t_2 < \cdots < t_k \leq n$. Let θ be a parameter vector that uniquely determines the time-series model $(Y_t)_{t\geq 1}$, and assume that maximum likelihood estimators (MLE) of θ exist, $\widehat{\theta}$ and $\widehat{\theta}^K$ based on Y and Y^K, respectively.

We want to assess the influence of the missing data $(Y_{t_1}, \ldots, Y_{t_n})$ on the prediction of Y_{n+h} for $h = 1, \ldots, H$ where $H \geq 1$ is a given forecast horizon. For this, we are interested in $f(y_{n+h}|y;\theta)$ and $f(y_{n+h}|y^K;\theta)$ which are the conditional densities of Y_{n+h} given $Y = y$ and $Y^K = y^K$, respectively. Even though considering the joint densities $f(y_{n+1}, \ldots, y_{n+H}|y;\theta)$ and $f(y_{n+1}, \ldots, y_{n+H}|y^K;\theta)$ could be more desirable, analysing the marginal densities $f(y_{n+h}|y;\theta)$ and $f(y_{n+h}|y^K;\theta)$ for $h = 1, \ldots, H$ is computationally more convenient. This aspect is crucial in analysing very long series as it is usually the case when fitting and predicting strongly-dependent processes. The impact of cases in K on the prediction of Y_{n+h} might be judged by measuring the discrepancy between $f(y_{n+h}|y;\theta)$ and $f(y_{n+h}|y^K;\theta)$. Since these densities depend on the unknown parameter θ, they are calculated in practice by replacing θ by $\widehat{\theta}$ in $f(y_{n+h}|y;\theta)$, and θ by $\widehat{\theta}^K$ in $f(y_{n+h}|y^K;\theta)$. We use the Kullback–Leibler divergence to measure this disparity; see, for example, Kullback (1997, page 5), Johnson and Geisser (1983), and Cavanaugh and Johnson (1999). It is defined by

$$I_h = \int_{\mathbb{R}} \log \frac{f(y_{n+h}|y;\widehat{\theta})}{f(y_{n+h}|y^K;\widehat{\theta}^K)} f(y_{n+h}|y;\widehat{\theta}) \mathrm{d}y_{n+h}.$$

The best mean-squared predictors of Y_{n+h} based on Y and Y^K are

$$Y_n^h = \mathrm{E}[Y_{n+h}|Y], \qquad Y_n^{h,K} = \mathrm{E}[Y_{n+h}|Y^K],$$

and the prediction error variances are

$$\Sigma_n^h = \mathrm{E}[(Y_{n+h} - Y_n^h)^2], \qquad \Sigma_n^{h,K} = \mathrm{E}[(Y_{n+h} - Y_n^{h,K})^2].$$

These quantities depend on the unknown parameter θ, and we denote by $Y_n^h(\widehat{\theta})$, $\Sigma_n^h(\widehat{\theta})$, $Y_n^{h,K}(\widehat{\theta}^K)$ and $\Sigma_n^{h,K}(\widehat{\theta}^K)$ the corresponding quantities when θ is replaced by its estimate $\widehat{\theta}$ or $\widehat{\theta}^K$. A standard derivation of the Kullback–Leibler information for Gaussian distributions gives

$$I_h = \frac{1}{2}\left\{ \frac{\Sigma_n^h(\widehat{\theta})}{\Sigma_n^{h,K}(\widehat{\theta}^K)} - \log \frac{\Sigma_n^h(\widehat{\theta})}{\Sigma_n^{h,K}(\widehat{\theta}^K)} + \frac{(Y_n^h(\widehat{\theta}) - Y_n^{h,K}(\widehat{\theta}^K))^2}{\Sigma_n^{h,K}(\widehat{\theta}^K)} - 1 \right\},$$

and this expression measures the influence of observations indexed by K on the predictor and the prediction error variance simultaneously. In order to measure the influence on the set of predictors Y_n^h for $h = 1, \dots, H$, we propose the global divergence measure

$$D = \sum_{h=1}^{H} I_h. \tag{1}$$

1.2. Influence of cases on the smoothers from state-space models

In what follows, we assume that the time series $(Y_t)_{t\geq 1}$ is generated by a Gaussian linear state-space model. Following Cavanaugh and Johnson (1999), we focus on the state smoothers which are the best mean-squared linear estimates of the state vectors given the observations. A fairly general form of a linear state-space model is

$$X_{t+1} = F X_t + V_t,$$
$$Y_t = G X_t + W_t,$$

where $(X_t)_{t\geq 1}$ is a sequence of m-dimensional Gaussian states with $X_t = (X_t(1), \dots, X_t(m))'$, F is a matrix, G is a vector, and $U_t = \begin{bmatrix} V_t \\ W_t \end{bmatrix}$ is a zero-mean Gaussian multivariate process such that $(X_1, U_t)_{t\geq 1}$ is an independent sequence and $\mathrm{E}[U_t U_t'] = \begin{bmatrix} Q & S \\ S' & R \end{bmatrix}$. The matrices F, G, Q, S, R depend on the parameter θ of the time series model. For any fixed t, $1 \leq t \leq n$, we define the best mean-squared linear estimates

$$X_{t|n} = \mathrm{E}[X_t | Y], \qquad X_{t|n}^K = \mathrm{E}[X_t | Y^K],$$

with error covariance matrices

$$\Omega_{t|n} = \mathrm{E}[(X_t - X_{t|n})(X_t - X_{t|n})'], \qquad \Omega_{t|n}^K = \mathrm{E}[(X_t - X_{t|n}^K)(X_t - X_{t|n}^K)'].$$

The conditional densities which determine $X_{t|n}$ and $X_{t|n}^K$ are respectively $f(x_t|y;\theta)$ and $f(x_t|y^K;\theta)$ where $x_t = (x_t(1), \dots, x_t(m))$. Thus, the influence of cases indexed by K on the smoother $X_{t|n}$ might be judged by measuring the disparity between $f(x_t|y;\widehat{\theta})$ and $f(x_t|y^K;\widehat{\theta}^K)$. Using the directed Kullback–Leibler divergence

$$J_t = \int_{\mathbb{R}^m} \log \frac{f(x_t|y;\widehat{\theta})}{f(x_t|y^K;\widehat{\theta}^K)} f(x_t|y;\widehat{\theta}) \mathrm{d}x_t,$$

we obtain that

$$J_t = \frac{1}{2}\left\{ \mathrm{tr}\left[\Omega_{t|n}(\widehat{\theta})\Omega_{t|n}^K(\widehat{\theta}^K)^{-1}\right] - \log \frac{\det(\Omega_{t|n}(\widehat{\theta}))}{\det(\Omega_{t|n}^K(\widehat{\theta}^K))} \right.$$
$$\left. + (X_{t|n}(\widehat{\theta}) - X_{t|n}^K(\widehat{\theta}^K))'\Omega_{t|n}^K(\widehat{\theta}^K)^{-1}(X_{t|n}(\widehat{\theta}) - X_{t|n}^K(\widehat{\theta}^K)) - m \right\}, \tag{2}$$

where $X_{t|n}(\widehat{\theta})$ and $\Omega_{t|n}(\widehat{\theta})$ are the estimates of X_t and the corresponding error-covariance matrix based on the complete data Y through $\widehat{\theta}$, while $X_{t|n}^K(\widehat{\theta}^K)$ and $\Omega_{t|n}^K(\widehat{\theta}^K)$ denote the same quantities when the estimation of X_t is based on Y^K through $\widehat{\theta}^K$.

1.3. Reducing dimensionality

The evaluation of J_t in (2) requires the calculation of the inverse matrix $\Omega^K_{t|n}(\widehat{\theta}^K)^{-1}$ which is usually a very demanding task. For example, assume that $n = 1000$, $m = 100$ and $K = \{t_1\}$ (we measure the influence of a single case). For each t and t_1, we have to invert a matrix of order 100×100. This task must be done 10^6 times to measure the effect on all the smoothers $X_{t|n}$ of all observations Y_{t_1} considered individually. A simpler problem from a computational point of view, is to assess the influence of cases indexed by K on a linear combination $c'X_{t|n}$ instead of $X_{t|n}$, for some conveniently chosen vector $c = (c_1, \dots, c_m)'$. In this case, the Kullback–Leibler divergence is

$$J_t^c = \int_{\mathbb{R}} \log \frac{f^c(x_t|y;\widehat{\theta})}{f^c(x_t|y^K;\widehat{\theta}^K)} f^c(x_t|y;\widehat{\theta})\mathrm{d}x_t,$$

where $f^c(x_t|y;\theta)$ and $f^c(x_t|y^K;\theta)$ are the conditional density of $c'X_t$ given $Y = y$ and $Y^K = y^K$, respectively. For the Gaussian process $(Y_t)_{t\geq 1}$, J_t^c takes the form

$$J_t^c = \frac{1}{2}\left\{ \frac{c'\Omega_{t|n}(\widehat{\theta})c}{c'\Omega^K_{t|n}(\widehat{\theta}^K)c} - \log \frac{c'\Omega_{t|n}(\widehat{\theta})c}{c'\Omega^K_{t|n}(\widehat{\theta}^K)c} + \frac{(c'X_{t|n}(\widehat{\theta}) - c'X^K_{t|n}(\widehat{\theta}^K))^2}{c'\Omega^K_{t|n}(\widehat{\theta}^K)c} - 1 \right\}.$$

Note that this expression for J_t^c resembles (2), but it avoids the calculation of the inverse matrix $\Omega^K_{t|n}(\widehat{\theta}^K)^{-1}$. There are many possible choices for c. For instance in Section 2, we take $c = (1, 0, \dots, 0)'$ in the context of the state-space representation (7) and (8) and we explain how to calculate $c'X_{t|n}$ and $c'X^K_{t|n}$. Another alternative is to choose c as the first principal component c_t of the estimation error $X_t - X_{t|n}$. Thus, $c_t'(X_t - X_{t|n})$ is the linear combination with maximal variance subject to $c_t'c_t = 1$, and therefore c_t is an eigenvector associated to the largest eigenvalue λ_t of $\Omega_{t|n}$ (Shumway and Stoffer, 2000, page 466). Nevertheless, observe that in this case c depends on both t and $\widehat{\theta}$, and therefore, may not be an attractive choice from a computational point of view. Many alternative methods to reduce dimensionality may be found in the literature, see for instance the special issue on statistical learning methods including dimensionality reduction edited by Bock and Vichi (2007).

In order to measure the global effect of the cases in K on the set of smoothed estimates $X_{t|n}$ or $c'X_{t|n}$ for $t = 1, \dots, n$, we define the divergence measure

$$J = \sum_{t=1}^{n} J_t^c. \tag{3}$$

Our interest is to apply the influence measures D and J for long-memory ARFIMA models. To this end, a state-space representation of an ARFIMA process and calculations of the divergence measures D and J using outputs from Kalman recursions are presented in the next section.

2. Influence in ARFIMA models

A Gaussian sequence (Z_t) is called an ARFIMA(p, d, q) process if (Z_t) satisfies the equation

$$(1 - \phi_1 B - \cdots - \phi_p B^p)Z_t = (1 + \theta_1 B + \cdots + \theta_q B^q)(1 - B)^{-d}\epsilon_t, \tag{4}$$

where B is the backward shift operator $BZ_t = Z_{t-1}$, (ϵ_t) is a sequence of zero-mean uncorrelated Gaussian random variables with the same variance σ^2_ϵ, $d \in (-1, 1/2)$, and the polynomials $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ and $\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$ have no common zeros and neither $\phi(z)$ nor $\theta(z)$ has zeros in the closed unit disk $\{z \in \mathbb{C} : |z| \leq 1\}$. The unique causal moving average satisfying (4) is the process

$$Z_t = \sum_{k=0}^{\infty} \psi_k \epsilon_{t-k}, \tag{5}$$

where $(\psi_k)_{k\in\mathbb{N}}$ are called the MA(∞) parameters of (Z_t) and are the coefficients in the Taylor series expansion of $\psi(z) = \theta(z)\phi(z)^{-1}(1 - z)^{-d}$ for $|z| < 1$. The process (Z_t) admits the mean-squared convergent autoregressive representation

$$Z_t = \epsilon_t - \sum_{k=1}^{\infty} \pi_k Z_{t-k},$$

where $(\pi_k)_{k\in\mathbb{N}}$ are called the AR(∞) parameters of (Z_t) and are the coefficients in the Taylor series expansion of $\pi(z) = \theta(z)^{-1}\phi(z)(1-z)^d$ for $|z| < 1$. If $d \in (0, 1/2)$, the covariance function of (Z_t) is not summable, the spectral density of (Z_t) has a pole at the origin, and (Z_t) is a long-memory process.

Since long-memory processes are not Markovian, all linear state-space representations of the ARFIMA process (Z_t) are infinite-dimensional. Despite this, exact Kalman filter computations can be made based on the finite sample $(Y_1, \ldots, Y_n)$. However, the exact recursions are computationally intensive. To avoid such calculations, the infinite-dimensional state-space system can be very well approximated by a finite representation (Chan and Palma, 1998). For this, we consider the approximation of (5) given by

$$Y_t = \sum_{k=0}^{m} \psi_k \epsilon_{t-k}, \tag{6}$$

which corresponds to a MA(m) process in contrast to the MA(∞) process (5). Taking $X_t = (X_t(1), \ldots, X_t(m))'$, where

$$X_t(i) = \mathrm{E}[Y_{t+i-1}|I_{t-1}] = \sum_{k=i}^{m} \psi_k \epsilon_{t+i-1-k}$$

for $1 \le i \le m$ and $I_{t-1} = \overline{\mathrm{sp}}\{\epsilon_{t-i}; i \ge 1\}$, (Y_t) has the state-space representation (Brockwell and Davis, 1991, Example 12.1.6)

$$X_{t+1} = FX_t + H\epsilon_t, \tag{7}$$

$$Y_t = GX_t + \epsilon_t, \tag{8}$$

with

$$F = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}, \qquad G = [1, 0, \ldots, 0], \qquad H = \begin{bmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_{m-1} \\ \psi_m \end{bmatrix}. \tag{9}$$

In this case, we have

$$Q = \sigma_\epsilon^2 HH', \qquad S = \sigma_\epsilon^2 H', \qquad R = \sigma_\epsilon^2. \tag{10}$$

The approximate representation (6) of the ARFIMA(p, d, q) process (Z_t) has computational advantages over the exact one. Indeed, the matrices involved in the truncated Kalman equations are of order $m \times m$, and then only m^2 evaluations are required for each iteration. Therefore, the order of the MLE algorithm is reduced from $\mathcal{O}(n^3)$ to $\mathcal{O}(n)$. The consistency, efficiency and asymptotic normality of the MLE obtained with the approximate state-space representation are established in Chan and Palma (1998). Moreover, it is reported there that the truncated MLE works well with small truncation parameter m relative to the sample size n, for example, $m = 6$ for $n = 100$ and $m = 30$ for $n = 1000$. In what follows, we focus our analysis on the state-space system described by (7) and (8).

In order to compute the influence measure D defined by (1), we need to calculate the h-step predictors Y_n^h and $Y_n^{h,K}$ as well as their prediction error variances Σ_n^h and $\Sigma_n^{h,K}$. Since (Y_t) is a MA(m) process, Y_{n+h} is uncorrelated with Y_t for all $1 \le t \le n$ when $h > m$, and then $Y_n^h = Y_n^{h,K} = 0$ and $\Sigma_n^h = \Sigma_n^{h,K} = \mathrm{E}[Y_{n+h}^2] = \sigma_\epsilon^2 \sum_{k=0}^m \psi_k^2$ when $h > m$. Now, when $1 \le h \le m$, both predictors can be obtained by solving linear orthogonality relations. In the case of Y_n^h, Bondon (2001) has established simple relations between the prediction coefficients which are not only recursive with respect to the order n but also recursive with respect to the step h. In the case of $Y_n^{h,K}$, the covariance matrix of Y^K is not Toeplitz in general, and then recursive methods based on the Levinson algorithm are useless. For this reason, we use the Kalman recursions and we show in the Appendix that

$$Y_n^h = \widehat{X}_{n+1}(h), \qquad Y_n^{h,K} = \widehat{X}_{n+1}^K(h), \tag{11}$$

where $\widehat{X}_{n+1}(h)$ and $\widehat{X}_{n+1}^K(h)$ are the h-th component of the state predictors $\widehat{X}_{n+1} = \mathrm{E}[X_{n+1}|Y]$ and $\widehat{X}_{n+1}^K = \mathrm{E}[X_{n+1}|Y^K]$, respectively. Furthermore,

$$\Sigma_n^h = \Omega_{n+1}(h,h) + \sigma_\epsilon^2 \sum_{k=0}^{h-1} \psi_k^2, \qquad \Sigma_n^{h,K} = \Omega_{n+1}^K(h,h) + \sigma_\epsilon^2 \sum_{k=0}^{h-1} \psi_k^2, \tag{12}$$

where $\Omega_{n+1}(h,h)$ and $\Omega_{n+1}^K(h,h)$ are the h-th diagonal elements of the error covariance matrices $\Omega_{n+1} = \mathrm{E}[(X_{n+1} - \widehat{X}_{n+1})(X_{n+1} - \widehat{X}_{n+1})']$ and $\Omega_{n+1}^K = \mathrm{E}[(X_{n+1} - \widehat{X}_{n+1}^K)(X_{n+1} - \widehat{X}_{n+1}^K)']$, respectively. Now, $\widehat{X}_{n+1}$ and Ω_{n+1} are obtained through the Kalman prediction recursions detailed in Brockwell and Davis (1991, Proposition 12.2.2) with the initial condition $X_1 = \sum_{k=0}^{\infty} F^k H \epsilon_{-k}$. To obtain $\widehat{X}_{n+1}^K$ and Ω_{n+1}^K, we use the same recursions but the state-space model needs to be modified as explained by Shumway and Stoffer (1982) and Brockwell and Davis (1991, Section 12.3). Specifically, the state equation (7) is unchanged while the observation equation (8) is replaced by

$$Y_t^\star = G_t X_t + \epsilon_t^\star, \tag{13}$$

where

$$G_t = \begin{cases} G & \text{if } t \notin K, \\ 0 & \text{if } t \in K, \end{cases} \qquad \epsilon_t^\star = \begin{cases} \epsilon_t & \text{if } t \notin K, \\ N_t & \text{if } t \in K, \end{cases}$$

where (N_t) is a sequence of independent zero-mean Gaussian random variables with the same variance σ_N^2 which is independent of X_1 and (ϵ_t). Accordingly, the noise covariance matrices are now

$$Q = \sigma_\epsilon^2 H H', \qquad S_t = \begin{cases} \sigma_\epsilon^2 H' & \text{if } t \notin K, \\ 0 & \text{if } t \in K, \end{cases} \qquad R_t = \begin{cases} \sigma_\epsilon^2 & \text{if } t \notin K, \\ \sigma_N^2 & \text{if } t \in K. \end{cases} \tag{14}$$

The calculations of $(\widehat{X}_{n+1}, \Omega_{n+1})$ and $(\widehat{X}_{n+1}^K, \Omega_{n+1}^K)$ are detailed in the Appendix. The procedure applies to any set K, and therefore it can be used for both single-case deletion and multiple-case deletion diagnostics.

Consider now the influence measure J defined by (3) where $c = (1, 0, \ldots, 0)'$. According to (8), $c' X_t = X_t(1) = Y_t - \epsilon_t$. If $\sum_{k=0}^m \psi_k z^k \neq 0$ for all $z \in \mathbb{C}$ such that $|z| = 1$, a well-known result from complex analysis guarantees the existence of $r > 1$ such that

$$\left(\sum_{k=0}^m \psi_k z^k \right)^{-1} = \sum_{k=-\infty}^{\infty} \alpha_k z^k, \qquad r^{-1} < |z| < r,$$

the Laurent series being absolutely convergent in the specified annulus. Then, according to Brockwell and Davis (1991, Proposition 3.1.2), (6) can be inverted to give $\epsilon_t = \sum_{k=-\infty}^{\infty} \alpha_k Y_{t-k}$. Therefore,

$$X_t(1) = \sum_{k=-\infty}^{\infty} \beta_k Y_{t-k},$$

where

$$\beta_k = \begin{cases} -\alpha_k & \text{if } k \neq 0, \\ 1 - \alpha_0 & \text{if } k = 0. \end{cases}$$

Observe that $X_t(1)$ can also be written as

$$X_t(1) = \sum_{k=-\infty}^{-m} \beta_{t-k} Y_k + \sum_{k=1-m}^{n+m} \beta_{t-k} Y_k + \sum_{k=n+m+1}^{\infty} \beta_{t-k} Y_k. \tag{15}$$

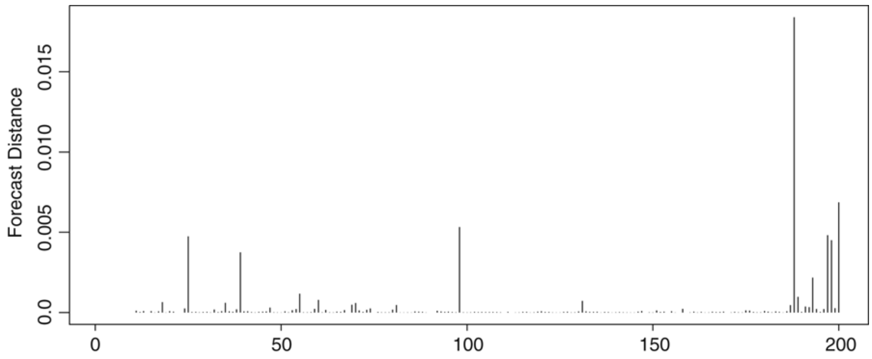

Fig. 1. Nile River influence measure D (out-of-sample predictors).

Since (Y_t) is a MA(m) process, the first and the third term on the right of (15) are uncorrelated with Y_t for all $1 \le t \le n$, and then

$$X_{t|n}(1) = \sum_{k=1-m}^{n+m} \beta_{t-k} \mathrm{E}[Y_k | Y], \qquad X_{t|n}^K(1) = \sum_{k=1-m}^{n+m} \beta_{t-k} \mathrm{E}[Y_k | Y^K],$$

where

$$\mathrm{E}[Y_k|Y] = Y_k \quad \text{if } 1 \le k \le n, \qquad \mathrm{E}[Y_k|Y^K] = Y_k \quad \text{if } k \in \{1, \dots, n\} \setminus K.$$

Again, since the covariance matrix of Y^K is not Toeplitz in general, the calculation of $\mathrm{E}[Y_k|Y^K]$ is a demanding task. Therefore, we use the Kalman fixed point smoothing recursions presented in Brockwell and Davis (1991, Proposition 12.2.4) with the state-space model (7) and (8) to get $X_{t|n}$ and $\Omega_{t|n}$, and with the state-space model (7) and (13) to get $X_{t|n}^K$ and $\Omega_{t|n}^K$. The calculations are detailed in the Appendix for any set K.

3. Illustration

In this section, we illustrate the use of influence measures in the context of long-memory series. To this end, consider the well-known River Nile data consisting of yearly minimal water levels for the years 622 A.D. – 821 A.D. recorded at the Rhoda Gauge. Following Beran (1994), an ARFIMA$(0, d, 0)$ model is fitted to this series by MLE after subtracting the sample mean. The estimated parameters are $\widehat{d} = 0.336$ with estimated asymptotic standard deviation $\widehat{\sigma}_d = 0.056$, and estimated noise standard deviation $\widehat{\sigma}_\epsilon = 0.898$. The computation of the estimates and the influence measures are carried out by means of the approximate state-space models (7)–(8), and (7), (13) with $m = 80$. After fitting the model, we calculate the following four influence measures where the set K corresponds to the singleton $K = \{t_1\}$ for each time t_1 in the data set. For greater clarity, the time scale in years is replaced by the index $k = 1, \dots, 200$.

(a) *Influence on out-of-sample predictors.* Fig. 1 displays the influence measure D as a function of k. Then D assesses the effect of deleting the k-th observation on the set of out-of-sample forecasts $\{Y_{200}^1, \dots, Y_{200}^{50}\}$. We use $H = 50$ since it represents 25% of the sample. From Fig. 1, the most influential observations seem to be $k = 25, 39, 98, 188, 197, 198, 200$. However, we would expect observations closer to the end of the series to be more influential on the out-of-sample forecasts than observations far back into the past. Some theoretical justifications can be found in Pourahmadi (2001, Section 8.3) and Bondon (2005). For instance, consider for a stationary time series (Z_t) the simple problem of predicting Z_{n+1} from $\{Z_t; t \le n, t \ne 0\}$, i.e. when only Z_0 from the past is missing. It was shown by Pourahmadi and Soofi (2000) that the increase in variance due to Z_0 being missing or excluded is

$$\delta_n = \frac{\pi_{n+1}^2}{\sum_{k=0}^{n} \pi_k^2} \sigma^2,$$

where σ^2 is the innovation variance of (Z_t) and (π_k) are the AR(∞) parameters of (Z_t). Hence, when the sequence $(|\pi_k|)$ is decreasing, the influence of Z_0 on the prediction of Z_{n+1} decreases as n increases. For instance, when (Z_t)

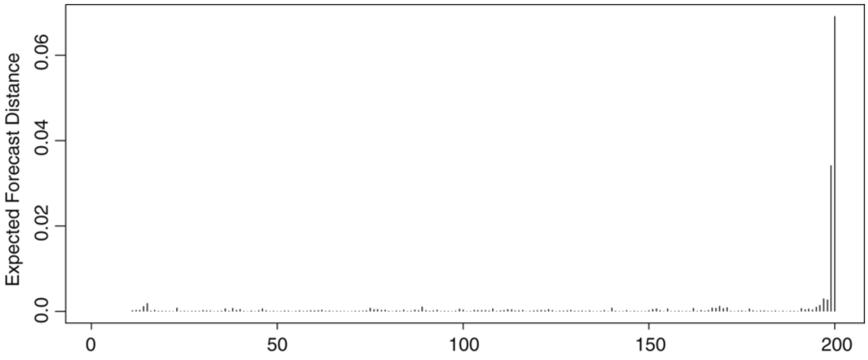

Fig. 2. Expected influence measure *ED* (out-of-sample predictors).

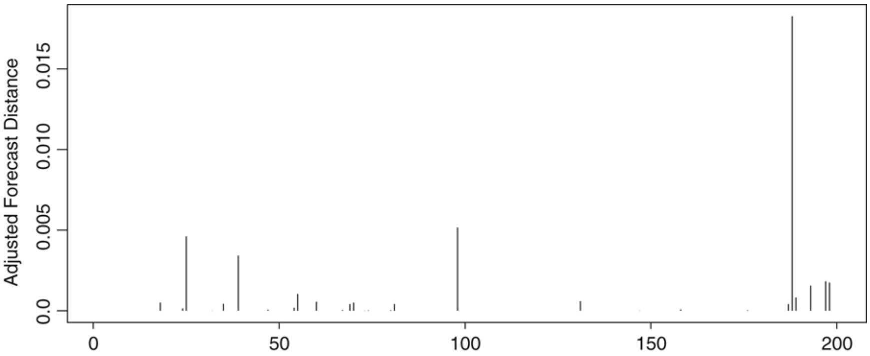

Fig. 3. Nile River influence measure *A* (out-of-sample predictors).

is an ARFIMA$(0, d, 0)$ process, we have

$$\pi_k = \frac{\Gamma(k-d)}{\Gamma(k+1)\Gamma(-d)}, \tag{16}$$

where Γ is the Gamma function, and then $|\pi_{k+1}| = |\pi_k|(k-d)/(k+1) < |\pi_k|$. This aspect is shown in Fig. 2 where we plot the expected D (ED) measure. This measure is computed as the average of D over 100 simulations of ARFIMA$(0, d, 0)$ processes with $d = 0.33$, $\sigma_\epsilon = 0.90$ and $n = 200$. In order to discount this effect, we compute the adjusted measure

$$A = \max(D - ED, 0),$$

so that A indicates the presence of overinfluential observations. Fig. 3 displays A and suggests that observations $k = 25, 39, 98, 188$ may be over-influential. In particular, the observations $k = 25$ and $k = 39$ which are close to the beginning of the series appear to have a large impact on what happens beyond the end of the series.

Fig. 4 illustrates how the out-of-sample predictors are numerically changed by the deletion of an influential observation as opposed to the deletion of a noninfluential case. We plot the measure

$$\Delta_h = 100 \left| 1 - \frac{Y_n^{h,K}(\widehat{\theta}^K)}{Y_n^h(\widehat{\theta})} \right|,$$

for $h = 1, \ldots, 50$, $n = 200$, $K = \{25\}$ and $K = \{150\}$. The estimates of the long-memory parameter are $\widehat{d} = 0.336$ (full sample), $\widehat{d}^{(25)} = 0.354$ (observation $k = 25$ deleted), and $\widehat{d}^{(150)} = 0.334$ (observation $k = 150$ deleted). It is clear that the deletion of the influential case $k = 25$ modifies the predictors much more than the deletion of the noninfluential case $k = 150$, even though the case $k = 25$ is farther into the past than the case $k = 150$. Moreover, we see that the effect of the observation $k = 25$ on the h-step predictor increases with h which shows the long-term influence.

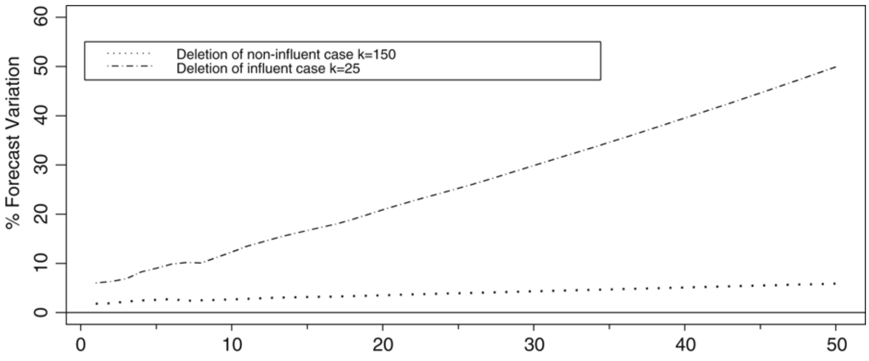

Fig. 4. Effect of an influential/noninfluential observation on the h-step predictor.

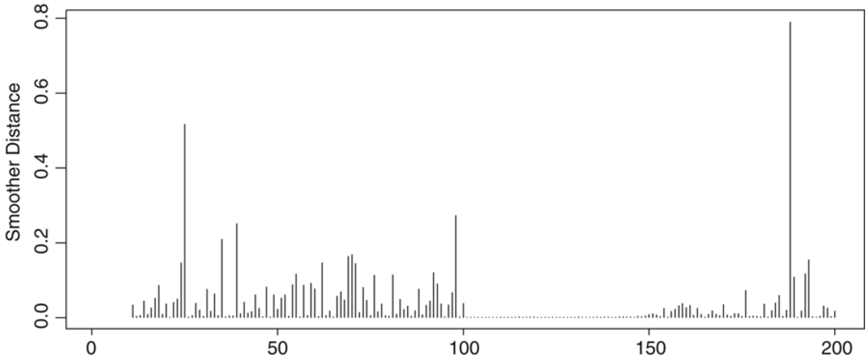

Fig. 5. Nile River influence measure J (smoothers).

(b) *Influence on the state smoothers.* The influence measure J defined by (3) where $c = (1, 0, \ldots, 0)'$ is considered in this illustration and is displayed in Fig. 5. The results suggest that the most influential observations are $k = 25, 35, 39, 98, 188$. Thus, the smoother influence measure seems to indicate one more observation as potentially influential than A, namely the case $k = 35$.

(c) *Mahalanobis distance.* This metric assesses the impact of observations in K on the in-sample predictions of the process. Following Peña (1991), we may write the Mahalanobis distance as

$$M = \frac{\left(\mathfrak{p}(\widehat{d}) - \mathfrak{p}(\widehat{d}^K)\right)' \widehat{\Sigma}_{\mathfrak{p}}^{-1} \left(\mathfrak{p}(\widehat{d}) - \mathfrak{p}(\widehat{d}^K)\right)}{n\widehat{\sigma}_\epsilon^2},$$

where $\mathfrak{p}(\widehat{d})$ and $\mathfrak{p}(\widehat{d}^K)$ are the vector $(\pi_1, \ldots, \pi_n)'$ where d is replaced in (16) by $\widehat{d}$ and $\widehat{d}^K$, respectively, and $\widehat{\sigma_\epsilon}^2 \widehat{\Sigma}_{\mathfrak{p}}$ is the estimated asymptotic covariance matrix of $\mathfrak{p}(\widehat{d})$. According to this distance, the most influential observations seem to be $k = 25, 35, 39, 98, 188$, see Fig. 6. Consequently, in this case the Mahalanobis distance indicates similar influential observations as the smoother measure J.

(d) *Cook-type distance.* This influence measure assesses the impact of observations in K on the estimate of the long-memory parameter d, and following Chatterjee and Hadi (1988, page 151) it can be written as

$$C = \frac{|\widehat{d} - \widehat{d}^K|}{\widehat{\sigma}_d}.$$

Fig. 7 displays the measure C for the Nile River data. Due to the lack of theoretical baselines for this measure in the case of long-memory time-series models, we have selected potentially overinfluential cases as follows. Observe that the cases $k = 25, 188$ have been also flagged as highly influential by all the previous measures. Combining this information with the Cook's distance, we have chosen these two cases as overinfluential observations. However, according to Fig. 7, we should also select the case $k = 70$ among the most influential.

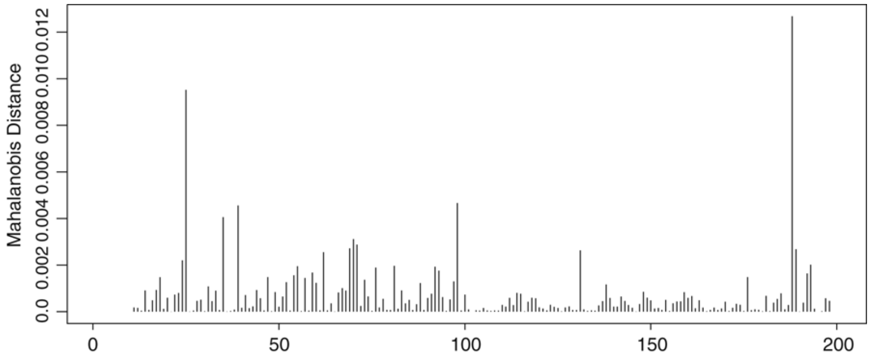

Fig. 6. Nile river influence measure M (Mahalanobis).

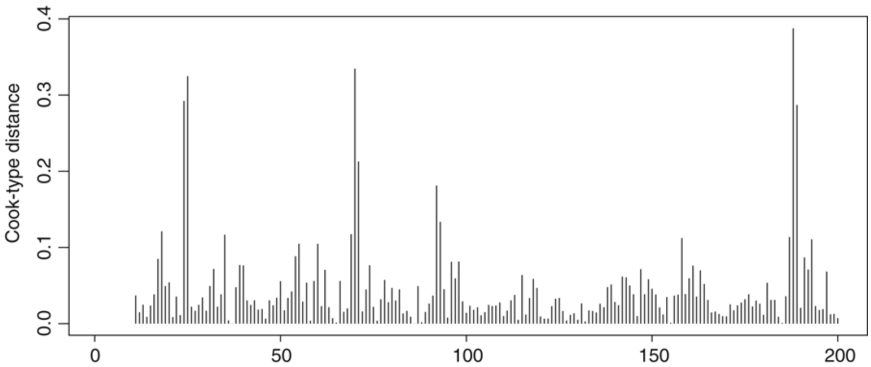

Fig. 7. Nile River influence measure C (Cook).

Table 1
Nile River data: Summary of influential observations

Influence measure	Observation								
	25	35	39	70	98	188	197	198	200
Out-of-sample predictors (D)	√		√		√	√	√	√	√
Adjusted D (A)	√		√		√	√			
State smoothers (J)	√	√	√		√	√			
In-sample predictors (M)	√	√	√		√	√			
Parameter estimate (C)	√			√		√			

Table 1 summarizes the results. Observations $k = 25$ and $k = 188$ seem to be overinfluential in terms of both in-sample and out-of-sample predictors, state smoothers and estimation of the long-memory parameter. Observation $k = 35$ is flagged as influential by both the smoothers and the Mahalanobis distances. Observations $k = 39$ and $k = 98$ seem to be overinfluential in terms of both in-sample and out-of-sample predictors and smoothers, while observation $k = 70$ appears to have substantial influence on the parameter estimation. Finally, observations $k = 197, 198, 200$ are flagged as influential by the out-of-sample prediction influence measure D.

In order to gain insight about the reasons why some observations are flagged as influential by different diagnostic techniques, we plot the squared differences and the cumulative squared differences of the Nile River data in Figs. 8 and 9, and the original series and the cumulative corrected mean series in Fig. 10.

In Figs. 8 and 9, the dotted vertical lines represent the locations of potentially influential observations. The legends on each case indicate the corresponding measure. From these plots, note that some cases ($k = 25, 188$) are associated with large jumps in the data while other flagged cases ($k = 197, 198, 200$) are related to their position at the end of the series. As discussed above, we would expect some impact of these observations on out-of-sample predictors.

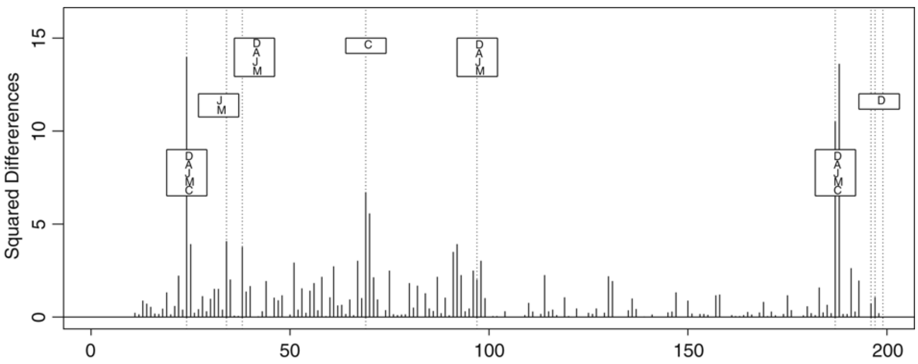

Fig. 8. Squared differences of the Nile River series with all the flagged locations (vertical dotted lines). The legends on each case indicate the corresponding influence measure.

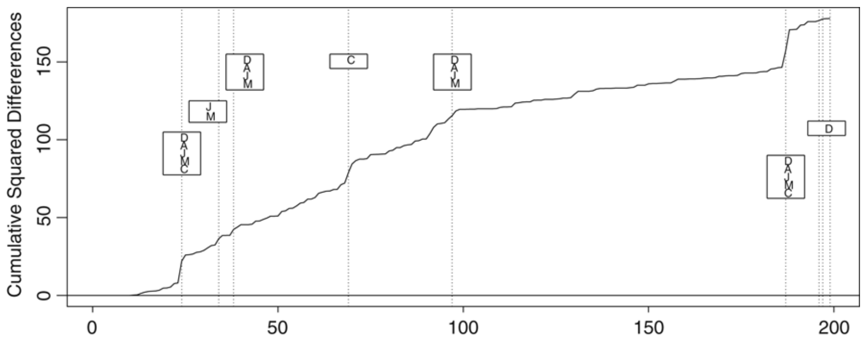

Fig. 9. Cumulative squared differences of the Nile River series with all the flagged locations (vertical dotted lines). The legends on each case indicate the corresponding influence measure.

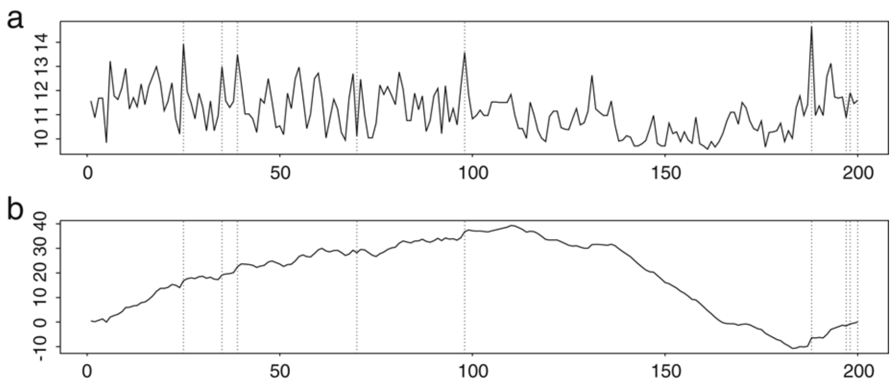

Fig. 10. Flagged locations (vertical dotted lines) (a) Original Nile River series, (b) Cumulative corrected mean Nile River series.

We observe in Fig. 10(a) that most flagged locations are either in the first half or by the end of the series. In the second half of the time series, there is a stretch of data with relatively low levels and no flagged locations. Analogously, we notice in Fig. 10(b) that all the flagged points seem to be located in upper-trend periods, or equivalently, periods with high average water levels.

Finally, observe that most of the techniques discussed in this paper suggest that around the year 720 A.D. ($k = 98$), there are some possible highly influential observations. To some extent this result coincides with a structural change previously noted by Beran (1994, Section 10.3).

The procedure to calculate the quantities in the influence measures discussed in this section may be summarized in the following two steps.

Step I. Calculation of parameter estimates, Kalman predictors and smoothers with the full sample:

(i) Select an initial value $\theta^{(0)}$ for the parameter vector θ and deduce initial values $\sigma_\epsilon^{2(0)}$ for σ_ϵ^2, and $\psi_k^{(0)}$ for the MA(∞) parameters ψ_k for $k = 1, \ldots, m$.

(ii) For $s = 1, \ldots, n$, run the Kalman recursions (18) to obtain $\widehat{X}_s(1)^{(0)}$ and $\Omega_s^{(0)}$, and compute $\ell(\theta^{(0)})$ where

$$\ell(\theta) = -\frac{1}{2}\sum_{s=1}^{n}\left\{\log(\Omega_s(1,1)+\sigma_\epsilon^2) + \frac{(Y_s - \widehat{X}_s(1))^2}{\Omega_s(1,1)+\sigma_\epsilon^2}\right\}$$

is the log-likelihood function of Y (up to a constant).

(iii) Run one iteration of the maximization procedure for $\ell(\theta)$ to obtain a new set of estimates, $\theta^{(1)}$.

(iv) Iterate the steps (i)–(iii) until convergence replacing $\theta^{(j-1)}$ by $\theta^{(j)}$ at the j-th iteration. Denote the resulting parameter vector by $\widehat{\theta}$.

(v) Calculate $Y_n^h(\widehat{\theta})$ and $\Sigma_n^h(\widehat{\theta})$ for $h = 1, \ldots, H$ using (11) and (12), respectively, and determine $X_{t|n}(\widehat{\theta})$ and $\Omega_{t|n}(\widehat{\theta})$ for $t = 1, \ldots, n$ using (20).

Step II. Calculation of parameter estimates, Kalman predictors and smoothers with observations in K treated as missing:

(i) Take $\widehat{\theta}$ obtained in step I(iv) as an initial value $\theta^{(0)}$ for the parameter vector and apply step I(i).

(ii) For $s = 1, \ldots, n$, run the Kalman recursions Eq. (18) with the missing data modifications Eq. (19) to obtain $\widehat{X}_s^K(1)^{(0)}$ and $\Omega_s^{K(0)}$, and compute $\ell^K(\theta^{(0)})$ where

$$\ell^K(\theta) = -\frac{1}{2}\sum_{s=1}^{n}\left\{\log(\Omega_s^K(1,1)+\sigma_\epsilon^2) + \frac{(Y_s^\star - \widehat{X}_s^K(1))^2}{\Omega_s^K(1,1)+\sigma_\epsilon^2}\right\}$$

is the log-likelihood function of Y^K (up to a constant).

(iii) Proceed as in steps I(iii)–(iv) where $\ell(\theta)$ is replaced by $\ell^K(\theta)$. Denote by $\widehat{\theta}^K$ the resulting parameter vector.

(iv) Calculate $Y_n^{h,K}(\widehat{\theta}^K)$ and $\Sigma_n^{h,K}(\widehat{\theta}^K)$ for $h = 1, \ldots, H$ using (11) and (12), respectively, and determine $X_{t|n}^K(\widehat{\theta}^K)$ and $\Omega_{t|n}^K(\widehat{\theta}^K)$ for $t = 1, \ldots, n$ using (21).

4. Conclusion

In this paper, we have proposed and implemented statistical methodologies to assess the influence of observations in long-memory models. The measures discussed are based on the Kullback–Leibler discrepancy between two densities and they assess the impact of deleting one or more observations on out-of-sample forecasts or on state smoothers. The diagnostic techniques discussed in this work are complementary since they focus on assessing influence in different aspects of the model, namely out-of-sample predictors, state smoothers, in-sample predictors and parameter estimates.

In order to make these methodologies applicable in practice, we have explained how to compute the diagnostic measures by using efficient Kalman filter techniques. These state-space methods are crucial for solving two main problems related to the calculation of influence measures for long-memory models, namely the need for efficient computation of estimates, predictors and smoothers along with their variances, and the necessity of appropriate techniques for dealing with missing values. The methodology can be summarized as follows: (a) Represent the long-memory process in terms of a state-space model, (b) Run Kalman recursions to obtain parameter estimates, predictors and state smoothers with all the data and then with deleted cases, (c) Obtain the influence measure and (d) Detect over-influential observations.

We have illustrated the use of the diagnostic techniques with an influence analysis of the well-known River Nile data. From this application, we may conclude that for long-memory processes, some observations may have not only a local influence, but also a large impact, far beyond the end of the series.

At this stage of the development, it is more appropriate to view these detection techniques as exploratory tools for flagging some observations as potentially influential. This information can be used for verifying data integrity (e.g. consistency, accuracy, correctness), checking for possible changes of the process (e.g. structural changes), warning about possible problems with predictions of future values, etc. However, further work for deriving practical levels of overinfluence for these measures is necessary. Along this line, large-scale Monte Carlo studies may prove to be helpful. Extensions of these techniques to the detection of influential observations in regression models with long-range dependent errors are being explored along with other related methodologies such as local-influence measures.

Acknowledgment

This research was supported by grants ECOS/CONICYT and FONDECYT 1040934.

Appendix

Calculations of (Y_n^h, Σ_n^h) **and** $(Y_n^{h,K}, \Sigma_n^{h,K})$. According to Brockwell and Davis (1991, Eqs. (12.2.11–12)), we have $Y_n^h = GF^{h-1}\widehat{X}_{n+1}$ and $Y_n^{h,K} = GF^{h-1}\widehat{X}_{n+1}^K$, and then (11) follows from (9). It results from Brockwell and Davis (1991, Eqs. (12.2.13–14)) and (9) and (10) that $\Sigma_n^h = \Omega_n^h(1,1) + \sigma_\epsilon^2$, where

$$\Omega_n^h(i,j) = \Omega_n^{h-1}(i+1,j+1) + \sigma_\epsilon^2 \psi_i \psi_j \tag{17}$$

for $h \geq 2$ and $1 \leq i, j \leq m$ with the convention $\Omega_n^h(i,j) = 0$ if $i > m$ or $j > m$, and the initial condition $\Omega_n^1 = \Omega_{n+1}$. Applying (17) recursively, we get $\Omega_n^h(1,1) = \Omega_n^1(h,h) + \sigma_\epsilon^2 \sum_{k=1}^{h-1} \psi_k^2$, and therefore we get the expression of Σ_n^h given in (12). To obtain the expression of $\Sigma_n^{h,K}$, we note that $\Sigma_n^{h,K}$ satisfies the same recursions that Σ_n^h but with the initial condition $\Omega_n^1 = \Omega_{n+1}^K$.

Calculations of $(\widehat{X}_{n+1}, \Omega_{n+1})$. Applying Brockwell and Davis (1991, Eqs. (12.2.6–7)) to the state-space model (7) and (8), we deduce from (9) and (10) that $\widehat{X}_{n+1}$ and Ω_{n+1} are determined by the following recursions for $s = 1, \ldots, n$,

$$\begin{aligned}
\widehat{X}_{s+1}(i) &= \widehat{X}_s(i+1) + \frac{\Theta_s(i)}{\Omega_s(1,1)+\sigma_\epsilon^2}(Y_s - \widehat{X}_s(1)), \\
\Theta_s(i) &= \Omega_s(i+1,1) + \sigma_\epsilon^2 \psi_i, \\
\Pi_{s+1}(i,j) &= \Pi_s(i+1,j+1) + \sigma_\epsilon^2 \psi_i \psi_j, \\
\Upsilon_{s+1}(i,j) &= \Upsilon_s(i+1,j+1) + \frac{\Theta_s(i)\Theta_s(j)}{\Omega_s(1,1)+\sigma_\epsilon^2}, \\
\Omega_{s+1} &= \Pi_{s+1} - \Upsilon_{s+1},
\end{aligned} \tag{18}$$

where $1 \leq i, j \leq m$ with the convention $\widehat{X}_s(i) = \Omega_s(i,1) = \Pi_s(i,j) = \Upsilon_s(i,j) = 0$ if $i > m$ or $j > m$, and the initial conditions $\widehat{X}_1 = \mathrm{E}(X_1)$, $\Pi_1 = \mathrm{E}(X_1 X_1')$, $\Upsilon_1 = \mathrm{E}(X_1)\mathrm{E}(X_1)'$, $\Omega_1 = \Pi_1 - \Upsilon_1$. Since $X_1 = \sum_{k=0}^{\infty} F^k H \epsilon_{-k} = \sum_{k=0}^{m-1} F^k H \epsilon_{-k}$, we have $\widehat{X}_1 = \Upsilon_1 = 0$, $\Omega_1 = \Pi_1$, and

$$\Pi_1(i,j) = \Pi_1(j,i) = \sigma_\epsilon^2 \sum_{k=0}^{m-j} \psi_{i+k}\psi_{j+k}, \qquad 1 \leq i \leq j \leq m.$$

Calculations of $(\widehat{X}_{n+1}^K, \Omega_{n+1}^K)$. Applying Brockwell and Davis (1991, Eqs. (12.2.6–7)) to the state-space model (7) and (13), it results from (14) that $\widehat{X}_{n+1}^K$ and Ω_{n+1}^K are determined by the recursions (18) with the same conventions and initial conditions, but where $\widehat{X}_s$, $\widehat{X}_{s+1}$, Ω_s and Ω_{s+1} are replaced by $\widehat{X}_s^K$, $\widehat{X}_{s+1}^K$, Ω_s^K and Ω_{s+1}^K, respectively, and equation defining Θ_s is replaced by

$$\Theta_s(i) = \begin{cases} \Omega_s^K(i+1,1) + \sigma_\epsilon^2 \psi_i & \text{if } s \notin K, \\ 0 & \text{if } s \in K. \end{cases} \tag{19}$$

Calculations of $(X_{t|n}, \Omega_{t|n})$. Applying Brockwell and Davis (1991, Eqs. (12.2.18–20)) to the state-space model (7) and (8), we deduce from Eqs. (9) and (10) that $X_{t|n}$ and $\Omega_{t|n}$ are determined for fixed t by the following recursions, which are solved sequentially for $s = t, t+1, \ldots, n$:

$$
\begin{aligned}
X_{t|s}(i) &= X_{t|s-1}(i) + \frac{\Omega_{t,s}(i,1)}{\Omega_s(1,1)+\sigma_\epsilon^2}(Y_s - \widehat{X}_s(1)),\\
\Omega_{t,s+1}(i,j) &= \Omega_{t,s}(i,j+1) - \frac{\Omega_{t,s}(i,1)\Theta_s(j)}{\Omega_s(1,1)+\sigma_\epsilon^2},\\
\Omega_{t|s}(i,j) &= \Omega_{t|s-1}(i,j) - \frac{\Omega_{t,s}(i,1)\Omega_{t,s}(j,1)}{\Omega_s(1,1)+\sigma_\epsilon^2},
\end{aligned} \tag{20}
$$

where $1 \le i, j \le m$ with the convention $\Omega_{t,s}(i,j) = 0$ if $j > m$, $\widehat{X}_s$, Θ_s and Ω_s are obtained by (18), and the initial conditions are $X_{t|t-1} = \widehat{X}_t$ and $\Omega_{t,t} = \Omega_{t|t-1} = \Omega_t$.

Calculations of $(X^K_{t|n}, \Omega^K_{t|n})$. Using Brockwell and Davis (1991, Eqs. (12.2.18–20)) with the state-space model (7) and (13), it results from (14) that $X^K_{t|n}$ and $\Omega^K_{t|n}$ are determined for fixed t by the following recursions, which are solved successively for $s = t, t+1, \ldots, n$:

$$
\begin{aligned}
X^K_{t|s}(i) &= \begin{cases} X^K_{t|s-1}(i) + \dfrac{\Omega_{t,s}(i,1)}{\Omega^K_s(1,1)+\sigma_\epsilon^2}(Y_s - \widehat{X}^K_s(1)) & \text{if } s \notin K,\\ X^K_{t|s-1}(i) & \text{if } s \in K,\end{cases}\\
\Omega^K_{t|s}(i,j) &= \begin{cases} \Omega^K_{t|s-1}(i,j) - \dfrac{\Omega_{t,s}(i,1)\Omega_{t,s}(j,1)}{\Omega^K_s(1,1)+\sigma_\epsilon^2} & \text{if } s \notin K,\\ \Omega^K_{t|s-1}(i,j) & \text{if } s \in K,\end{cases}
\end{aligned} \tag{21}
$$

where $\Omega_{t,s}$ is defined in (20) in which Ω_s is replaced by Ω^K_s and Θ_s is given by (19), and the initial conditions are $X^K_{t|t-1} = \widehat{X}^K_t$ and $\Omega_{t,t} = \Omega^K_{t|t-1} = \Omega^K_t$.

References

Belsley, D.A., Kuh, E., Welsch, R.E., 1980. Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. John Wiley, New York.

Beran, J., 1994. Statistics for Long-Memory Processes. Chapman and Hall, New York.

Bock, H.-H., Vichi, M., 2007. Special issue on statistical learning methods including dimensionality reduction. Comput. Statist. Data Anal. 52 (1).

Bondon, P., 2001. Recursive relations for multistep prediction of a stationary time series. J. Time Ser. Anal. 22 (4), 399–410.

Bondon, P., 2005. Influence of missing values on the prediction of a stationary time series. J. Time Ser. Anal. 26 (4), 519–525.

Brockwell, P.J., Davis, R.A., 1991. Time Series: Theory and Methods, second edition. Springer Verlag, New York.

Bruce, A.G., Martin, R.D., 1989. Leave-*k*-out diagnostics for time series. J. Roy. Statist. Soc. Ser. B 51 (3), 363–424. With discussion and a reply by the authors.

Cai, Y., Davies, N., 2003. A simple diagnostic method of outlier detection for stationary Gaussian time series. J. Appl. Stat. 30 (2), 205–223.

Cavanaugh, J.E., Johnson, W.O., 1999. Assessing the predictive influence of cases in a state space process. Biometrika 86 (1), 183–190.

Cavanaugh, J.E., Oleson, J.J., 2001. A diagnostic for assessing the influence of cases on the prediction of missing data. The Statistician 50 (4), 427–440.

Chan, N.H., Palma, W., 1998. State space modeling of long-memory processes. Ann. Statist. 26 (2), 719–740.

Chatterjee, S., Hadi, A.S., 1988. Sensitivity Analysis in Linear Regression. John Wiley, New York.

Choy, K., 2001. Outlier detection for stationary time series. J. Statist. Plann. Inference 99 (2), 111–127.

Doornik, J.A., Ooms, M., 2003. Computational aspects of maximum likelihood estimation of autoregressive fractionally integrated moving average models. Comput. Statist. Data Anal. 42 (3), 333–348.

Doukhan, P., Oppenheim, G., Taqqu, M.S. (Eds.), 2003. Theory and Applications of Long-Range Dependence. Birkhäuser Boston Inc., Boston, MA.

Harrison, J., West, M., 1991. Dynamic linear model diagnostics. Biometrika 78 (4), 797–808.

Johnson, W., Geisser, S., 1983. A predictive view of the detection and characterization of influential observations in regression analysis. J. Amer. Statist. Assoc. 78 (381), 137–144.

Kohn, R., Ansley, C.F., 1987. Signal extraction for finite nonstationary time series. Biometrika 74 (2), 411–421.

Kullback, S., 1997. Information theory and statistics. Dover Publications Inc., Mineola, NY. Reprint of the second (1968) edition.

Lawrance, A.J., 1995. Deletion influence and masking in regression. J. Roy. Statist. Soc. Ser. B 57 (1), 181–189.

Lefrançois, B., 1991. Detecting over-influential observations in time series. Biometrika 78 (1), 91–99.

Palma, W., 2007. Long-Memory Time Series: Theory and Methods. In: Wiley Series in Probability and Statistics, John Wiley & Sons, Hoboken, NJ.

Peña, D., 1991. Measuring influence in dynamic regression models. Technometrics 33, 93–101.
Pourahmadi, M., 2001. Foundations of Time Series Analysis and Prediction Theory. Wiley, New York.
Pourahmadi, M., Soofi, E.S., 2000. Prediction variance and information worth of observations in time series. J. Time Ser. Anal. 21 (4), 413–434.
Proietti, T., 2003. Leave-K-out diagnostics in state-space models. J. Time Ser. Anal. 24 (2), 221–236.
Shumway, R.H., Stoffer, D.S., 1982. An approach to time series smoothing and forecasting using the EM algorithm. J. Time Ser. Anal. 3, 253–264.
Shumway, R.H., Stoffer, D.S., 2000. Time Series Analysis and its Applications. Springer-Verlag, New York.
Thomas, W., Cook, R.D., 1990. Assessing influence on predictions from generalized linear models. Technometrics 32, 59–65.
Tsay, R.S., Peña, D., Pankratz, A.E., 2000. Outliers in multivariate time series. Biometrika 87 (4), 789–804.