INFLUENCE OF MISSING VALUES ON THE PREDICTION OF A STATIONARY TIME SERIES

By Pascal Bondon

CNRS UMR 8506

First Version received May 2003

Abstract. The influence of missing observations on the linear prediction of a stationary time series is investigated. Simple bounds for the prediction error variance and asymptotic behaviours for short and long-memory processes respectively are presented.

Keywords. Prediction theory; missing value problems; asymptotic behaviour; long-memory.
AMS 2000 subject classification. Primary 62M10, Secondary 60G25.

1. INTRODUCTION

In many practical situations, the values of a time series are not observed at equally spaced times either because some data are missing or because prudent statistical analysis of the data leads to discard some observed suspect values. In these situations, a parametric model can be fitted to the data either by maximizing the exact Gaussian likelihood function of the observed data or by using an expectation-maximization (EM) algorithm (see for instance, Shumway and Stoffer, 2000, Ch. 4).

When the parametric model admits a finite dimensional state space representation, as this is the case for instance for an autoregressive moving average (ARMA) model, this representation can be modified to accommodate the missing data, and the exact Gaussian likelihood function can readily and conveniently be computed using the associated Kalman prediction recursions (see Brockwell and Davis, 1991, Ch. 12). Then linear estimates of the missing values and their corresponding prediction error variances can be evaluated by application of the Kalman fixed point smoothing algorithm to the modified state space model. When the parametric model does not admit a finite dimensional state space representation, as this is the case for example for any long-memory process in the class of fractionally integrated ARMA models, this approach is not applicable directly.

An alternative to estimate the missing values consists in using explicit formulae of these estimates as proposed by Brubacher and Wilson (1976) and Ljung (1989). As noticed by Ferreiro (1987), this direct approach is often more efficient numerically, especially if the number of missing observations is small. Furthermore, explicit formulae are useful for analysing theoretically the influence of the number and of the positions of the missing values on the interpolation error variances (see Pourahmadi, 2001, Ch. 8).

An algebraic expression for the best linear interpolation of a finite number of missing data with arbitrary pattern of any (minimal) stationary time series, whose past and future are observed indefinitely was given by Rozanov (1967, p. 101). This formula involves the inverse autocorrelations of the series, which can be expressed themselves in terms of the AR(∞) parameters (see Masani, 1960, Lem. 2.7). In the case where the past is infinite but the future is finite, the same interpolation problem was studied by Pourahmadi (1989), who established elegant closed form expressions.

A problem, which is closely related to the interpolation of missing data is the prediction of future data in the presence of an incomplete past. Within the framework of the state space formulation, future data can be evaluated by applying the Kalman prediction algorithm to the appropriately modified model. For a stationary process, the predictors obtained by iterating indefinitely in the Kalman recursions converge in mean square to the best linear infinite past predictor. Without resorting to a state space formulation, Cheng and Pourahmadi (1997) have proposed an algorithm to compute the best linear predictor. This algorithm applies to any stationary time series with finite or infinite state space representation, and is a generalization of the innovation algorithm presented in Brockwell and Davis (1991, Prop. 5.2.2). In the case where the past is altered by a finite number of missing data with arbitrary pattern, a closed form expression for the prediction error variance was given by Grenander and Rosenblatt (1954), and an explicit formula for the predictor was established in Bondon (2002).

In this paper, we investigate the influence of missing observations on the linear prediction of a stationary time series. In Theorem 1, lower and upper bounds for the prediction error variance are established. These bounds reveal the important role played by the AR(∞) parameters. In Theorem 2, some properties of the predictor are presented, and in Theorem 3, asymptotic behaviours for the prediction error variance are obtained for short and long-memory processes respectively.

2. PRELIMINARIES

We denote by L_2, the space of the equivalence classes of square integrable real valued random variables defined on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$. The space L_2 is a Hilbert space with inner product $\langle X, Y \rangle = \mathrm{E}XY$ and norm $\|X\| = \sqrt{\mathrm{E}X^2}$, where E stands for the expectation operator. In this paper, convergence of a series of random variables will always be in L_2. We consider a zero mean and (weakly) stationary time series $(X_t)_{t \in \mathbb{Z}}$, i.e., $(X_t) \subset L_2$ and $\langle X_s, X_t \rangle$ depends only on $t - s$ for any $(s,t) \in \mathbb{Z}^2$. We set $\sigma_X = \|X_t\|$.

For a collection S of random variables in L_2, $\overline{\text{sp}}\, S$ denotes the closed linear span of elements of S in the norm of L_2. For any $t \in \mathbb{Z}$, we set

$$H_t = \overline{\text{sp}}\{X_s; s \leq t\},$$

the past and present subspace of the process (X_t) at time t, P_t denotes the orthogonal projection operator onto H_t, and $\epsilon_t = X_t - P_{t-1}X_t$. The sequence (ϵ_t) is zero mean, uncorrelated, stationary, and is called the innovation process of (X_t). We set $\sigma_\epsilon = \|\epsilon_t\|$. The process (X_t) is said to be nondeterministic if $\epsilon_t \neq 0$ for some $t \in \mathbb{Z}$. It was shown by Kolmogorov (1941, Thm. 23) that (X_t) is nondeterministic if and only if (iff) the logarithm of its spectral density f is integrable with respect to Lebesgue measure on $[-\pi, \pi]$, $\mathrm{d}\lambda$. In this case, we have

$$\sigma_\epsilon^2 = 2\pi \exp\left\{ \int_{-\pi}^{\pi} \ln f(\lambda) \mathrm{d}\lambda / 2\pi \right\}.$$

The Wold decomposition theorem states that every nondeterministic process (X_t) can be written as

$$X_t = \sum_{i=0}^{\infty} c_i \epsilon_{t-i} + V_t, \tag{1}$$

where the coefficients (c_i) are unique, $c_0 = 1$, $\sum_{i=0}^{\infty} c_i^2 < \infty$, (V_t) is deterministic and is uncorrelated with (ϵ_t), and $(V_t) \subset \cap_{t \in \mathbb{Z}} H_t$. The coefficients (c_i) are called the MA(∞) parameters of (X_t). The AR(∞) parameters (a_i) are defined recursively by

$$a_0 = -1, \qquad a_i = -\sum_{j=0}^{i-1} a_j c_{i-j}, \quad i \geq 1. \tag{2}$$

Both (a_i) and (c_i) only depend on the Fourier coefficients of $\ln f$.

For any $t \in \mathbb{Z}$, we set

$$I_t = \overline{\text{sp}}\{X_s; s \neq t\},$$

the past and future subspace of the process (X_t) at time t, P_{I_t} denotes the orthogonal projection operator onto I_t, and $\eta_t = X_t - P_{I_t} X_t$. The process (η_t) is zero mean, stationary, and is called the interpolation error or the two-sided innovation process of (X_t). We set $\sigma_\eta = \|\eta_t\|$. The process (X_t) is said to be minimal if $\eta_t \neq 0$ some $t \in \mathbb{Z}$. Since $H_{t-1} \subset I_t$, every minimal stationary process is nondeterministic. It was shown by Kolmogorov (1941, Thm. 24) that (X_t) is minimal iff f^{-1} is Lebesgue integrable on $[-\pi, \pi]$. According to Masani (1960, Lem. 2.7), minimality is also equivalent to $\sum_{i=0}^{\infty} a_i^2 < \infty$. In this case, we have

$$\sigma_\eta^2 = \frac{4\pi^2}{\int_{-\pi}^{\pi} f^{-1}(\lambda) \mathrm{d}\lambda} = \frac{\sigma_\epsilon^2}{\sum_{i=0}^{\infty} a_i^2}.$$

In the following, we assume that the data $X_{-t_1}, \ldots, X_{-t_N}$ are missing, where N is a given integer and $0 < t_1 < \cdots < t_N$, and we consider the problem of predicting X_0 from its incomplete past

$$H'_{-1} = \overline{\mathrm{sp}}\{X_t; t \leq -1, t \neq -t_1, \ldots, -t_N\}.$$

We denote by P'_{-1} the orthogonal projection operator onto H'_{-1}, and we set $M = \{t_1, \ldots, t_N\}$.

3. BOUNDS FOR THE PREDICTION ERROR VARIANCE

In order to quantify the worth of the data $X_{-t_1}, \ldots, X_{-t_N}$ in the linear prediction of X_0, lower and upper bounds for the increase of the prediction error variance $\|X_0 - P'_{-1}X_0\|^2 - \sigma_\epsilon^2$ are established in Theorem 1. Note that $P'_{-1}X_0 \in H'_{-1} \subset H_{-1}$, and $\epsilon_0 \perp H_{-1}$. Therefore, $P_{-1}X_0 - P'_{-1}X_0 \perp \epsilon_0$, and we have

$$\|X_0 - P'_{-1}X_0\|^2 = \|\epsilon_0 + P_{-1}X_0 - P'_{-1}X_0\|^2 = \sigma_\epsilon^2 + \|P_{-1}X_0 - P'_{-1}X_0\|^2.$$

THEOREM 1. *For any nondeterministic stationary process* (X_t), *we have*

$$\sigma_\eta \max_{i \in M} |a_i| \leq \|P_{-1}X_0 - P'_{-1}X_0\| \leq \sigma_\epsilon \sum_{i \in M} \varphi_i |a_i| \leq \sigma_X \sum_{i \in M} |a_i|, \tag{3}$$

where $\varphi_i = \left(\sum_{j=0}^{t_N - i} c_j^2\right)^{1/2}$.

PROOF. According to (1), we have

$$X_0 - \sum_{i=1}^{t_N} a_i X_{-i} = -\sum_{i=0}^{t_N} a_i X_{-i} = -\sum_{i=0}^{t_N} a_i \sum_{j=0}^{\infty} c_j \epsilon_{-i-j} - \sum_{i=0}^{t_N} a_i V_{-i}$$

$$= -\sum_{k=0}^{\infty} s_k \epsilon_{-k} - \sum_{i=0}^{t_N} a_i V_{-i},$$

where $s_k = \sum_{i=0}^{\min(t_N, k)} a_i c_{k-i}$. According to (2), $s_0 = -1$ and $s_k = 0$ for $k = 1, \ldots, t_N$. Therefore,

$$X_0 = \sum_{i=1}^{t_N} a_i X_{-i} + \epsilon_0 + U_0,$$

where

$$U_0 = -\sum_{k=t_N+1}^{\infty} s_k \epsilon_{-k} - \sum_{i=0}^{t_N} a_i V_{-i} \in H_{-t_N-1}.$$

Consequently,

$$P_{-1}X_0 = \sum_{i=1}^{t_N} a_i X_{-i} + U_0,$$

$$P'_{-1}X_0 = \sum_{\substack{i=1 \\ i \notin M}}^{t_N} a_i X_{-i} + \sum_{i \in M} a_i P'_{-1} X_{-i} + U_0,$$

$$P_{-1}X_0 - P'_{-1}X_0 = \sum_{i \in M} a_i (X_{-i} - P'_{-1}X_{-i}). \tag{4}$$

Since $H_{-t_N-1} \subset H'_{-1}$, we have

$$\begin{aligned} \left\| \sum_{i \in M} a_i (X_{-i} - P'_{-1}X_{-i}) \right\| &\le \sum_{i \in M} |a_i| \|X_{-i} - P'_{-1}X_{-i}\| \\ &\le \sum_{i \in M} |a_i| \|X_{-i} - P_{-t_N-1}X_{-i}\|. \end{aligned} \tag{5}$$

According to (1), for any $t \in \mathbb{Z}$, we have

$$P_{-t_N-1}X_t = \sum_{i=t_N+t+1}^{\infty} c_i \epsilon_{t-i} + V_t,$$

$$X_t - P_{-t_N-1}X_t = \sum_{i=0}^{t_N+t} c_i \epsilon_{t-i}.$$

Hence, $\|X_{-i} - P_{-t_N-1}X_{-i}\| = \sigma_\epsilon \varphi_i$, and the first upper bound in (3) follows from (4) and (5). Since $\varphi_i \le (\sum_{j=0}^{\infty} c_j^2)^{1/2}$ and (1) implies that $\sigma_X^2 = \sigma_\epsilon^2 \sum_{j=0}^{\infty} c_j^2 + \|V_t\|^2$, the second upper bound in (3) follows immediately from the first one. Let $(i, j) \in M^2$. We have

$$\langle X_{-i}, \eta_{-j} \rangle = \begin{cases} \sigma_\eta^2 & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$

On the other hand, since $H'_{-1} \subset I_{-j}$ and $\eta_{-j} \perp I_{-j}$, we have $\eta_{-j} \perp H'_{-1}$. Therefore,

$$\left\langle \sum_{i \in M} a_i (X_{-i} - P'_{-1}X_{-i}), \eta_{-j} \right\rangle = a_j \sigma_\eta^2,$$

and using Schwarz inequality, we get

$$|a_j| \sigma_\eta \le \left\| \sum_{i \in M} a_i (X_{-i} - P'_{-1}X_{-i}) \right\|,$$

which, using (4), gives the lower bound in (3). □

For a minimal stationary process, the lower bound in (3) shows that the degradation of the prediction due to the missing data increases with $\max_{i \in M} |a_i|$. In the particular case, where $M = \{m\}$, $m > 0$, inequalities (3) give

$$\frac{\sigma_\epsilon |a_m|}{(\sum_{i=0}^{\infty} a_i^2)^{1/2}} \le \|P_{-1}X_0 - P'_{-1}X_0\| \le \sigma_\epsilon |a_m|, \tag{6}$$

and it results from Pourahmadi and Soofi (2000, Thm. 3.1) that

$$\|P_{-1}X_0 - P'_{-1}X_0\| = \frac{\sigma_\epsilon |a_m|}{(\sum_{i=0}^{m-1} a_i^2)^{1/2}}.$$

Therefore, the larger m is, the better is the preciseness of the lower bound in (6), while the lower m is, the better is the preciseness of the upper bound in (6).

Theorem 2 is an immediate consequence of Theorem 3.3 and Remark 3.3b in Bondon (2002).

THEOREM 2. *Let* (X_t) *be a nondeterministic stationary process. Then*
(a) $P'_{-1}X_0 = P_{-1}X_0$ *iff* $a_i = 0$ *for all* $i \in M$,
(b) $P'_{-1}X_0 \to P_{-1}X_0$ *in* L_2 *as* $t_1 \to \infty$.

For a minimal stationary process, Theorem 2 follows from inequalities (3). Indeed, (a) is immediate and to show (b), note that minimality is equivalent to $\sum_{i=0}^{\infty} a_i^2 < \infty$, which implies that $a_i \to 0$ as $i \to \infty$. Therefore, $a_i \to 0$ for each i in M as $t_1 \to \infty$, and the upper bound in (3) implies (b).

It is worth noticing that Theorem 1 gives rates of convergence of $P'_{-1}X_0$ to $P_{-1}X_0$ as t_1 tends to infinity. In Theorem 3, two typical asymptotic behaviours are presented.

THEOREM 3. *Let* (X_t) *be a nondeterministic stationary process. Then, as* $t_1 \to \infty$,
(a) $\|P_{-1}X_0 - P'_{-1}X_0\| = O(\alpha^{t_1})$ *if* $|a_i| = O(\alpha^i)$ as $i \to \infty$ *for some* $\alpha \in (0,1)$,
(b) $\|P_{-1}X_0 - P'_{-1}X_0\| \asymp t_1^{\alpha}$ *if* $|a_i| \sim ci^{\alpha}$ *as* $i \to \infty$ *where* $c > 0$ *and* $\alpha < -1/2$.

PROOF. (a) follows from $\sum_{i\in M}|a_i| = O(\alpha^{t_1})$ and from the upper bound in (3). Since $\alpha < -1/2$ and $|a_i| \sim ci^{\alpha}$ as $i \to \infty$, $\sum_{i=0}^{\infty} a_i^2 < \infty$, and therefore, (X_t) is minimal. Hence, we deduce from (3) that, for any $\epsilon \in (0,1)$, we have

$$0 < \sigma_\eta c(1-\epsilon) \le \frac{\|P_{-1}X_0 - P'_{-1}X_0\|}{t_1^{\alpha}} \le \sigma_X c(1+\epsilon)N < \infty,$$

for t_1 sufficiently large, which shows (b). □

According to Theorem 3(a), $P'_{-1}X_0$ tends to $P_{-1}X_0$ at least exponentially if (X_t) is an ARMA process with a positive spectral density. Theorem 3(b) implies that $P'_{-1}X_0$ tends to $P_{-1}X_0$ hyperbolically if (X_t) is a fractionally integrated ARMA process with long-memory parameter $d \in (-1/2, 1/2)\backslash\{0\}$. More precisely, in this case we have $|a_i| \sim ci^{-(1+d)}$ (see Kokoszka and Taqqu, 1995, Cor. 3.1), and then $\|P_{-1}X_0 - P'_{-1}X_0\| \asymp t_1^{-(1+d)}$. Therefore, the convergence of $P'_{-1}X_0$ to $P_{-1}X_0$ is faster in the long-memory case corresponding to $d \in (0,1/2)$, than in the intermediate memory case, where $d \in (-1/2, 0)$. At first glance, this may seem nonintuitive because $\mathrm{E}(X_{-t_1} X_0) \sim C t_1^{2d-1}$ and then, X_{-t_1} and X_0 are more correlated in the long-memory case. Nevertheless, the worth of X_{-t_1} in the linear prediction of X_0 depends crucially on the AR(∞) parameter a_{t_1}, and this parameter tends to zero more quickly, when d is positive.

REFERENCES

Bondon, P. (2002) Prediction with incomplete past of a stationary process. *Stochastic Processes and their Applications* 98(1), 67–76.

Brockwell, P. J. and Davis, R. A. (1991) *Time Series: Theory and Methods* (2nd edn). New York: Springer Verlag.

Brubacher, S. R. and Wilson, G. T. (1976) Interpolating time series with application to the estimation of holiday effects on electricity demand. *Journal of Applied Statistics* 25(2), 107–16.

Cheng, R. and Pourahmadi, M. (1997) Prediction with incomplete past and interpolation of missing values. *Statistics and Probability Letters* 33, 341–6.

Ferreiro, O. (1987) Methodologies for the estimation of missing observations in time series. *Statistics and Probability Letters* 5, 65–9.

Grenander, U. and Rosenblatt, M. (1954) An extension of a theorem of G. Szegö and its application to the study of stochastic processes. *Transactions of the American Mathematical Society* 76, 112–26.

Kokoszka, P. S. and Taqqu, M. S. (1995) Fractional ARIMA with stable innovations. *Stochastic Processes and their Applications* 60(1), 19–47.

Kolmogorov, A. N. (1941) Stationary sequences in Hilbert space. *Bulletin of the Moscow State University* 2(6), 1–40.

Ljung, G. M. (1989) A note on the estimation of missing values in time series. *Communications in Statistics. Simulation and Computation* 18(2), 459–65.

Masani, P. (1960) The prediction theory of multivariate stochastic processes, III. *Acta Mathematica* 104, 141–62.

Pourahmadi, M. (1989) Estimation and interpolation of missing values of a stationary time series. *Journal of Time Series Analysis* 10(2), 149–69.

Pourahmadi, M. (2001) *Foundations of Time Series Analysis and Prediction Theory*. New York: Wiley.

Pourahmadi, M. and Soofi, E. S. (2000) Prediction variance and information worth of observations in time series. *Journal of Time Series Analysis* 21(4), 413–34.

Rozanov, Yu. A. (1967) *Stationary Random Processes*. San Francisco: Holden-Day.

Shumway, R. H. and Stoffer, D. S. (2000) *Time Series Analysis and its Applications*. New York: Springer Verlag.