Estimation and Forecasting of Long-memory Processes with Missing Values

WILFREDO PALMA[1*] AND NGAI HANG CHAN[2]

[1] *Pontificia Universidad Católica de Chile, Santiago, Chile*
[2] *Carnegie Mellon University, Pittsburgh, USA and HKUST, Hong Kong*

ABSTRACT

This paper addresses the issues of maximum likelihood estimation and forecasting of a long-memory time series with missing values. A state-space representation of the underlying long-memory process is proposed. By incorporating this representation with the Kalman filter, the proposed method allows not only for an efficient estimation of an ARFIMA model but also for the estimation of future values under the presence of missing data. This procedure is illustrated through an analysis of a foreign exchange data set. An investment scheme is developed which demonstrates the usefulness of the proposed approach. © 1997 John Wiley & Sons, Ltd.

J. forecast, Vol. **16**, 395–410 (1997)

No. of Figures: 12. No. of Tables: 8. No. of References: 14.

KEY WORDS long memory; ARFIMA models; forecasting; maximum likelihood estimation; missing values; foreign exchange data

Long-range dependent models have been playing an important role in diverse fields ranging from economics to oceanography. A comprehensive review on this subject can be found in the recent monograph by Beran (1994) and the references therein. Data with long-range behaviour is often modelled by means of the so-called fractionally integrated autoregressive moving average (ARFIMA) process. An ARFIMA (p, d, q) process $\{y_t\}$ is defined by

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

where $\{\varepsilon_t\}$ is a sequence of uncorrelated random variables with zero means and constant variances σ^2 (white noise), B is the backshift operator such that $By_t = y_{t-1}$, $\Phi(B) = 1 + \phi_1 B + \cdots + \phi_p B^p$ is the autoregressive operator, $\Theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$ is the moving average operator, and $(1 - B)^d$ is the fractional difference operator. For $d \in (0, 1/2)$, model (1) defines a *long-memory*

* Correspondence to: Wilfredo Palma, Departamento de Probabilidad y Estadística, Universidad Católica de Chile, Casilla 306, Santiago 22, Chile.

Contract grant sponsor: Office of Naval Research; Contract grant number: N00014-89-J-1851
Contract grant sponsor: Research Grant Council of Hong Kong

CCC 0277–6693/97/060395–16$17.50

Received July 1996
Accepted April 1997

process with non-summable autocorrelations, that is, $\sum_{k=0}^{\infty} |\rho_k| = \infty$, where ρ_k is the kth lag autocorrelation function of $\{y_t\}$. For $d < 0$, model (1) is an *antipersistent* or an *intermediate-memory* process, with zero spectral density at frequency zero and summable autocorrelations, $\sum_{k=0}^{\infty} |\rho_k| < \infty$. If d is a positive integer, model (1) corresponds to an autoregressive integrated moving average (ARIMA) model and if $d = 0$, model (1) is the usual ARMA process.

ARFIMA models are said to have *long-memory* because their autocorrelations decay to zero at a hyperbolic rate, that is, $\rho_k \sim |k|^{-\alpha}$, $\alpha > 0$, for large k. On the other hand, ARMA models are called *short-memory* processes since their autocorrelations converge to zero at an exponential rate, i.e. $\rho_k \sim \mathrm{e}^{-a|k|}$, $a > 0$, for large k. Thus, the correlation between past and present observations vanishes at a faster rate for ARMA models than for ARFIMA processes. This unusual feature makes the estimation of parameters for an ARFIMA model more difficult but helps in the prediction of future values (see, for example, p. 11 of Beran, 1994). As a consequence, fitting ARFIMA models to real-life data has been a subtle and difficult task for practitioners.

This paper has three main objectives. First, it develops an efficient state-space algorithm to compute the maximum likelihood (ML) estimates for ARFIMA models via truncation. This procedure facilitates the fitting of ARFIMA models in practice. Second, it proposes a modification to the Kalman filter equations which allows for missing values. When these techniques are applied to analyse a foreign exchange data set which consists of missing values and exhibits long-memory behaviour, it is found that both model estimation and model forecast can be achieved reliably and efficiently. The state space approach proposed in this paper not only provides means to analyse long-range dependent time series but also helps to deal with missing values for long-range dependent models. Third, using these techniques, an investment scheme is developed to analyse the foreign exchange data set which demonstrates the effectiveness of the procedure.

This paper is organized as follows. In the next section the state-space models and the Kalman filter are introduced. The third section addresses the problem of missing data and presents a solution by a modified Kalman filter. Applications of the state-space systems and the Kalman filter to a foreign exchange data set are discussed in the fourth section. Conclusions are presented in the final section.

STATE-SPACE MODELS

Although a standard ARIMA model always has a finite-dimensional state-space representation, an ARFIMA process can only be written in terms of an infinite-dimensional state-space system (see Chan and Palma, 1996). A moving average representation of the ARFIMA(p, d, q) process (1) is given by:

$$y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \qquad t = 1, \ldots, n \tag{2}$$

where ψ_j are the coefficients of $\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \Theta(z)/\Phi(z)(1-z)^{-d}$. (Formulae for evaluating the ψ_j are given in the Appendix.) From equation (2), an infinite-dimensional state-space representation may be written as (see p. 22 of Hannan and Deistler, 1988):

$$\begin{cases} X_{t+1} = FX_t + H\varepsilon_t & t = 1, \ldots, n \\ y_t = GX_t + \varepsilon_t & t = 1, \ldots, n \end{cases} \tag{3}$$

where $X_t = [y(t \mid t-1)\ y(t+1 \mid t-1)\ y(t+2 \mid t-1) \ldots]'$, $y(t \mid j) = E[y_t \mid y_j, y_{j-1}, \ldots]$

$$F = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & \ldots \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & \ldots \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \ddots & \end{bmatrix} \qquad H = \begin{bmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \vdots \end{bmatrix} \qquad G = [1\ 0\ 0\ 0 \ldots] \tag{4}$$

Assuming the noise sequence $\{\varepsilon_t\}$ to be normally distributed, the log-likelihood function (omitting a constant) of model (1) can be expressed as $l(\boldsymbol{\theta}) = -(1/2n)\log\det T_n(\boldsymbol{\theta}) - (1/2n) Y_n' T_n^{-1}(\boldsymbol{\theta}) Y_n$, where $T_n(\boldsymbol{\theta})$ is the covariance matrix of the observations $Y_n = (y_1, \ldots, y_n)'$ and $\boldsymbol{\theta}$ is the vector of parameters, $\boldsymbol{\theta} = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, d, \sigma)'$.

If expansion (2) is truncated after m components, we can write an approximate model as $y_t = \sum_{j=0}^{m} \psi_j \varepsilon_{t-j}$. An m-dimensional state-space representation of this truncated model is given by:

$$X_{t+1} = \begin{bmatrix} 0 & 1 & 0 & 0 & . & . & . & 0 \\ 0 & 0 & 1 & 0 & . & . & . & 0 \\ . & . & . & . & . & . & . & . \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} X_t + \begin{bmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_m \end{bmatrix} \varepsilon_t \tag{5}$$

$$y_t = [\,1 \quad 0 \quad 0 \quad . \quad . \quad . \quad 0] X_t + \varepsilon_t \tag{6}$$

The likelihood function can be evaluated by means of the Kalman filter. The *exact* ML estimates can be obtained by applying the Kalman filter to the state-space system (3) and (4). On the other hand, *approximate* ML estimates can be obtained by using the truncated m-dimensional state-space representation (5) and (6).

Given the state-space model (3), the one-step prediction of the state is given by $\hat{X}_t = P_{t-1} X_t$, where $P_{t-1} X_t$ represents the projection of X_t onto the closed linear space generated by $y_1, \ldots, y_{t-1}$. The error covariance matrix of the one-step prediction is $\Omega_t = E[X_t X_t' - \hat{X}_t \hat{X}_t']$. The Kalman filter produces the one-step predictions recursively as follows. If the initial values are taken to be $\hat{X}_1 = E[X_1]$ and $\Omega_1 = E[X_1 X_1'] - E[\hat{X}_1 \hat{X}_1']$, then the recursive equations are

$$\Delta_t = G \Omega_t G' + R \tag{7}$$

$$\Theta_t = F \Omega_t G' + S \tag{8}$$

$$\Omega_{t+1} = F \Omega_t F' + Q - \Theta_t \Delta_t^{-1} \Theta_t' \tag{9}$$

$$\hat{X}_{t+1} = F \hat{X}_t + \Theta_t \Delta_t^{-1} [y_t - G_t \hat{X}_t] \tag{10}$$

$$\hat{y}_t = G \hat{X}_t \tag{11}$$

where $R = \mathrm{Var}(\varepsilon_t) = \sigma^2$, $Q = \mathrm{Var}(H\varepsilon_t) = \sigma^2 HH'$ and $S = \mathrm{cov}(\varepsilon_t, H\varepsilon_t) = \sigma^2 H$. As usual, the initial state is estimated by $\hat{X}_1 = E[X_1] = 0$, and the initial prediction error variance covariance matrix is estimated by $\Omega_1 = E[X_1 X_1'] = (\omega_{ij}(\boldsymbol{\theta}))_{i,j=1,2,\ldots,}$, where $\omega_{ij}(\boldsymbol{\theta}) = \sum_{k=0}^{\infty} \psi_{i+k}(\boldsymbol{\theta}) \psi_{j+k}(\boldsymbol{\theta})$. (The coefficients $\psi_j(\boldsymbol{\theta})$ are given in the Appendix.) In the optimization process, it is necessary to choose an initial value for the parameter $\boldsymbol{\theta}$. In our case, the initial parameter is taken to be $\boldsymbol{\theta}_0 = (-0.5, 0.1, -0.6)$. Even though the algorithm is not highly sensitive to the initial value, $\boldsymbol{\theta}_0$ must correspond to a stationary and invertible ARFIMA model.

Based on the Kalman recursive equations (7)–(11), the likelihood function can be written as

$$L(\boldsymbol{\theta}) = (2\pi)^{-n/2}\left(\prod_{j=1}^{n}\sigma_j^2\right)^{-1/2}\exp\left\{-\frac{1}{2}\sum_{j=1}^{n}\frac{(y_j-\hat{y}_j)^2}{\sigma_j^2}\right\} \tag{12}$$

where $\hat{y}_j$ is the one-step prediction of y_j and $\sigma_j^2 = \mathrm{Var}(y_j - \hat{y}_j) = \Delta_j$, the one-step prediction error variance.

The exact likelihood function can be calculated by applying the Kalman recursion (7)–(11) directly to the infinite-dimensional system. Even though the exact likelihood function can be computed in a finite number of steps, such a computation may be cumbersome. Thus, for moderate to large sample sizes, it may be prudent to use the truncated representation to obtain an approximate ML estimate for $\boldsymbol{\theta}$.

Specifically, according to the Kalman filter equations, the evaluation of the likelihood function consists of n iterations (sample size) and each iteration consists of a number of matrix evaluations. For the exact method, these matrices are of dimensions $n \times n$, so there are n^2 evaluations. The resulting algorithm is then of order n^3 ($n \times n \times n$). On the other hand, the matrices involved in the approximate approach are of dimensions $m \times m$. Therefore, m^2 evaluations are required for each iteration. In this case, the algorithm is of order $n \times m^2$. Hence, for a fixed truncation parameter m, the calculation of the likelihood function is of order n for the approximate ML method. The truncated state-space approach is n^2/m times faster than the exact ML estimation. For moderate to large samples, it may be desirable to consider truncating the Kalman recursive equations after m components. With the truncation, the number of operations required for a single evaluation of the log-likelihood function is reduced to an order of n. Asymptotic behaviours of the estimates based on the approximate state-space are given in Chan and Palma (1996). In the next section, we discuss a modification to the Kalman filter method to handle the incomplete data problem for a long-memory time series.

THE PROBLEM OF MISSING DATA

Incomplete data present a serious problem for time-series analysts. However, an appropriate likelihood function for time series with missing data can be obtained by modifying the Kalman filter equations. Previous work along this line can be found in Jones (1980) who develops a Kalman filter approach to deal with missing values in ARMA models. Harvey and Pierse (1984), Ansley and Kohn (1985) and Kohn and Ansley (1986) extend Jones' results to ARIMA processes. An alternate approach using the Expectation-Maximization (EM) algorithm can be found in Shumway and Stoffer (1982). In this section, Kalman filter techniques are developed to compute the ML estimates of an ARFIMA process with missing observations.

A modification to the Kalman filter equations (8)–(11) to incorporate missing values can be proceeded as follows. Consider the state-space system

$$\begin{cases} X_{t+1} = FX_t + V_t \\ y_t = GX_t + W_t \end{cases}$$

where X_t is the state, y_t is the observation, F and G are the system matrices, $\mathrm{Var}(V_t) = Q$, $\mathrm{Var}(W_t) = R$, and $\mathrm{Cov}(V_t, W_t) = S$. The likelihood function can be evaluated recursively by means of the modified Kalman filter equations. Let $\Delta_t = \mathrm{Var}(y_t - \hat{y}_t) = \sigma_t^2$ and $\Omega_t = \mathrm{Var}(X_t - \hat{X}_t)$. The modified Kalman equations are:

$$\Delta_t = G\Omega_t G^T + R$$

$$\Theta_t = F\Omega_t G^T + S$$

$$\Omega_{t+1} = \begin{cases} F\Omega_t F^T + Q - \Theta_t \Delta_t^{-1} \Theta_t^T & \text{if observation } t \text{ is known} \\ F\Omega_t F^T + Q & \text{if observation } t \text{ is missing} \end{cases}$$

$$\hat{X}_{t+1} = \begin{cases} F\hat{X}_t + \Theta_t \Delta_t^{-1}(y_t - G\hat{X}_t) & \text{if observation } t \text{ is known} \\ F\hat{X}_t & \text{if observation } t \text{ is missing} \end{cases}$$

Modifications of the state covariance matrix equation and the state prediction are obtained by observing that if y_t is missing, then $K_t = K_{t-1}$, where K_t denotes the closed linear space generated by the *observed* values through time t, and $\hat{X}_{t+1} = E[X_{t+1} \mid K_t] = E[FX_t + V_t \mid K_t] = FE[X_t \mid K_t] + E[V_t \mid K_t] = FE[X_t \mid K_{t-1}] = F\hat{X}_t$. So $X_{t+1} - \hat{X}_{t+1} = FX_t + V_t - F\hat{X}_t = F(X_t - \hat{X}_t) + V_t$. Thus, the covariance matrix of the state estimation error becomes

$$\begin{aligned} \Omega_{t+1} &= \mathrm{Var}[X_{t+1} - \hat{X}_{t+1}] \\ &= \mathrm{Var}[F(X_t - \hat{X}_t) + V_t] \\ &= \mathrm{Var}[F(X_t - \hat{X}_t)] + \mathrm{Var}[V_t] + 2\,\mathrm{Cov}[F(X_t - \hat{X}_t), V_t] \end{aligned}$$

Since V_t is uncorrelated with $X_t - \hat{X}_t$,

$$\Omega_{t+1} = \mathrm{Var}[F(X_t - \hat{X}_t)] + \mathrm{Var}[V_t] = F\Omega_t F^T + Q$$

The missing values can be estimated as follows. Let the observation y_t be missing. By definition, $\hat{y}_t = E[y_t \mid K_{t-1}]$, where $K_{t-1} = \bar{sp}[y_{k_1}, \ldots, y_{k_{t-1}}]$ and the integers $k_1, \ldots, k_{t-1}$ denote the locations of the *observed* values up to time $t-1$. From the state-space system equation: $y_t = GX_t + W_t$, hence, $\hat{y}_t = E[GX_t + W_t \mid K_{t-1}] = G\hat{X}_t + E[W_t \mid K_{t-1}] = G\hat{X}_t$. Thus, the equation linking the estimation of the state and the observation is the same regardless of whether y_t is missing or not. The absence of y_t, however, affects the forecast of future observations. The variance of the prediction error is given by:

$$\mathrm{Var}[y_t - \hat{y}_t] = \mathrm{Var}[G(X_t - \hat{X}_t)] + \mathrm{Var}[W_t] + 2\,\mathrm{Cov}[G(X_t - \hat{X}_t), W_t]$$

Since $G(X_t - \hat{X}_t)$ and W_t are uncorrelated: $\mathrm{Var}[y_t - \hat{y}_t] = G\,\mathrm{Var}(X_t - \hat{X}_t)G^T + R = G\Omega_t G^T + R$. Summing up, the Kalman filter *recursive* likelihood with missing observation is given by

$$L(\boldsymbol{\theta}) = (2\pi)^{-(n-r)/2} \left(\prod_{t \in K_n} \sigma_t^2 \right)^{-1/2} \mathrm{e}^{-1/2[\Sigma_{t \in K_n}((y_t - \hat{y}_t)^2/\sigma_t^2)]} \tag{13}$$

where r is the number of missing values. If there are no missing data, formula (13) corresponds to the likelihood function (12). The *recursive* likelihood function for both complete and incomplete time series is used in the next section to calculate ML estimates for the exchange rate data.

FOREIGN EXCHANGE DATA ANALYSIS

In order to illustrate the application and performance of the Kalman filter techniques described in the previous section, we analyse a data set consisting of monthly observations of the exchange rate of French francs per US dollar. The first step in our analysis consists of a preliminary examination of the data. Second, an ARFIMA model is fitted to the foreign exchange series. Third, a study of the effects of missing values on the estimation and forecasting of the fitted model is conducted.

Figure 1(a) displays a time-series plot, from January 1971 to August 1994 with a total of $n = 284$ observations. The series presents a peak around February 1985. Due to increasing variability and trends in the data, the first difference of the log of the exchange rate is considered. The resulting series is plotted in Figure 1(b). The differenced series seems to be stationary, even though a small heterocedasticity is observed. The sample autocorrelation function (ACF) is shown in Figure 2(a). Both the autocorrelations and the partial autocorrelation function (PACF) decay slowly (see Figure 2(b)). The spectral density is depicted in Figure 2(c). The spectrum contains high peaks for small to medium frequencies. This seems to indicate the presence of both long- and short-memory component in the data.

The variance plot in Figure 3 is a useful tool to detect the presence of long-memory behaviour in the data. As discussed in Beran (1994, pp. 92–4), for a long-range dependent time series x_t, the variance of its mean values $\bar{x}_k$ satisfies $\mathrm{Var}(\bar{x}_k) \sim k^{2d-1}$. Therefore by plotting $\log(\mathrm{Var}(\bar{x}_k))$ versus $\log(k)$ for different values of k, a straight line with slope $2d - 1$ should be found. Since $d = 0$ for a short-memory process, the slope would be -1. Plots with slopes greater than -1 would indicate the presence of long-memory behaviour. For the foreign exchange data, the estimated slope is -0.64 (estimated through least squares), suggesting a crude estimate of the long-memory parameter $\hat{d} = 0.18$. For comparison, a straight line with slope -1 is plotted in Figure 3.

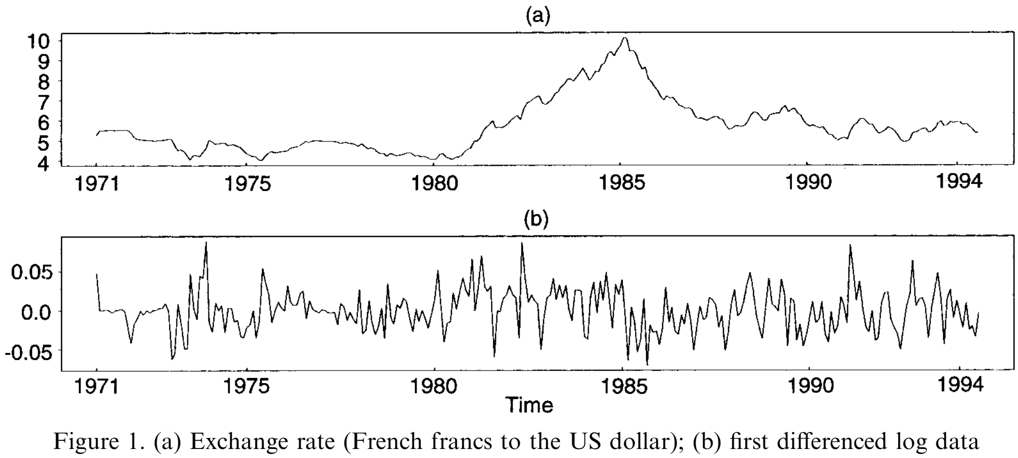

Figure 1. (a) Exchange rate (French francs to the US dollar); (b) first differenced log data

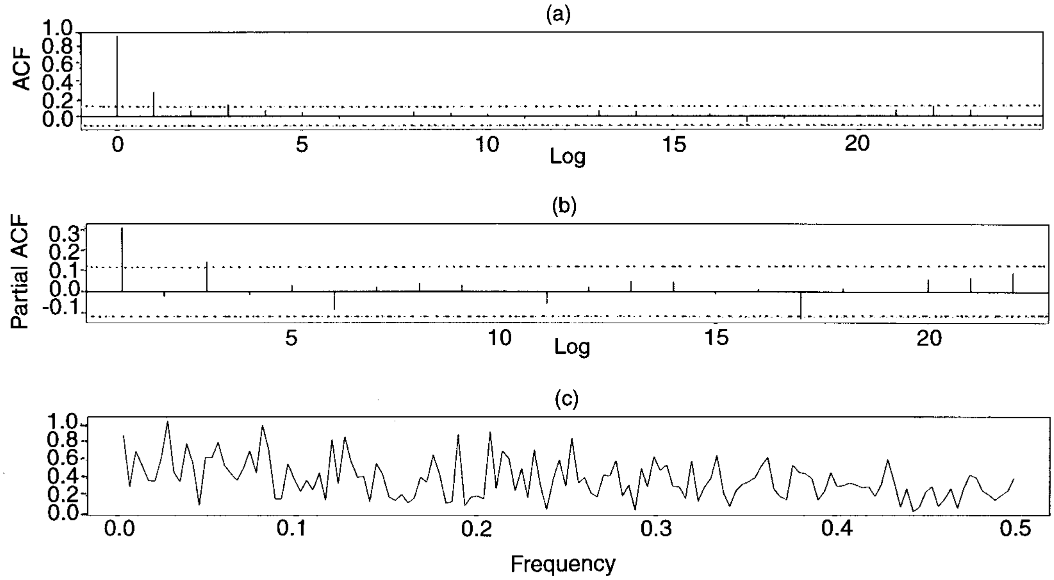

Figure 2. (a) Sample ACF of the differenced log data; (b) sample PACF of the differenced log data; (c) spectral density

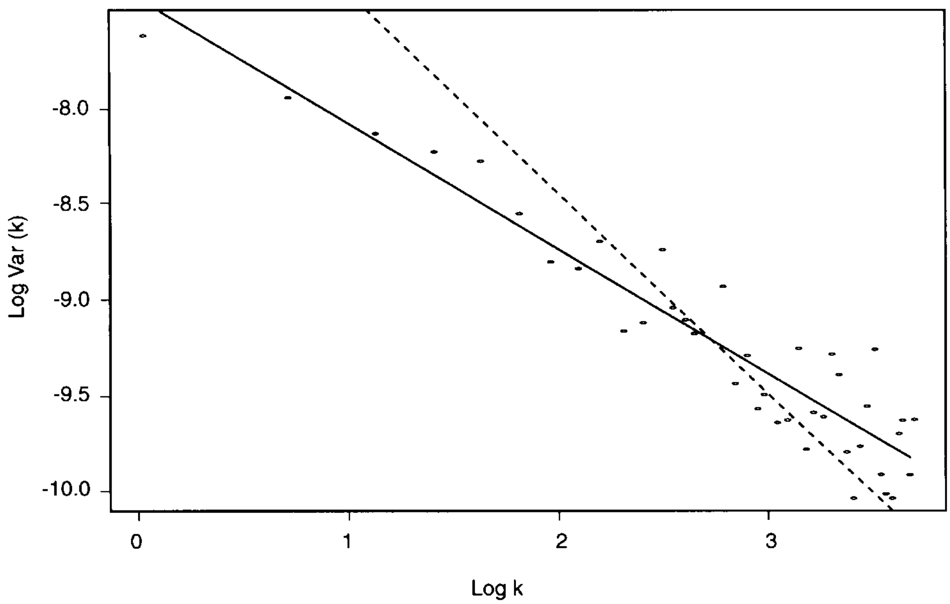

Figure 3. Variance plot. Heavy line: fitted straight line with slope -0.63; dotted line: fitted straight line with slope -1

Accordingly, a class of ARFIMA(p, d, q) models, with p and q less than or equal to 2 are considered. This set of models includes the short-memory ARMA(p, q) process as a particular case when $d = 0$. Parameter estimations are carried out by using the truncated state-space approach, with $m = 30$ for the ARFIMA(p, d, q) case and $m = \max\{p, q + 1\}$ for the ARMA(p, q) process. The optimizations are carried out by using the subroutine DUMINF in IMSL with numerical derivatives. The Hessian matrix and its inverse are computed by using the subroutines DFDHES and DLINDS in IMLS, respectively.

It is worth noting that by using the truncated approach, the model fitting process is speeded up considerably. For example, with a sample size $n = 284$ and a truncation parameter $m = 30$, the approximate ML algorithm is about 90 $(284^2/30^2)$ times faster than the exact ML method.

The model selection is based on the Akaike's Information Criterion (AIC) (see Hosking, 1984; Crato and Ray, 1996). Table I gives the parameter estimates of the selected ARFIMA$(1, d, 1)$ model. The estimated standard deviation of the noise is $\hat{\sigma} = 0.0252$. The covariance matrix of the coefficients is displayed in Table II.

Table I. Exchange-rate data: ARFIMA approximate ML parameter estimation

Parameter	MLE	t-statistic
d	0.133	2.3
ϕ_1	−0.490	−2.8
θ_1	−0.677	−5.1

Table II. Exchange-rate data: Parameter covariance matrix

	d	ϕ_1	θ_1
d	0.0033	−0.0045	−0.0020
ϕ_1		0.0296	0.0212
θ_1			0.0176

For comparison, the exact ML estimates of the selected ARFIMA$(1, d, 1)$ model are also calculated. The exact ML estimation is based on the state-space representation (3) and (4) and the Kalman filter equations (7)–(11). Table III displays the exact ML estimates for d, ϕ_1 and θ_1 and their t-statistics. Table IV shows the covariance matrix of the estimated parameters. From these two tables, it can be seen that exact and approximate ML estimates are in close agreement, indicating that the truncated approach provides reasonable ML estimates. This is not surprising since a Monte Carlo study conducted in Chan and Palma (1996) suggests that approximate ML performs very well under different situations.

Table III. Exchange-rate data: ARFIMA exact ML parameter estimation

Parameter	MLE	t-statistic
d	0.137	2.2
ϕ_1	−0.498	−2.7
θ_1	−0.685	−5.0

Table IV. Exchange-rate data: exact parameter covariance matrix

	d	ϕ_1	θ_1
d	0.0039	−0.0059	−0.0031
ϕ_1		0.0333	0.0237
θ_1			0.0191

Figure 4 shows a residual analysis of the fitted model. Based on Figure 4(a)–(c), the residuals seem to be white noise. The modified portmanteau test (Ljung and Box, 1978; Li and McLeod, 1986)

$$Q_M = n(n+2)\sum_{k=1}^{M}\frac{r_k^2}{n-k}$$

is distributed approximately as a $\chi^2(M-p-q-1)$ random variable, where r_k is the sample autocorrelation at lag k, n is the sample size, and M is a prespecified integer depending on n. In the present context, M is chosen as 20 and $Q_{20} = 10.04$ with a p-value of 0.902. This indicates that the ARFIMA$(1, d, 1)$ model fits the data reasonably well.

One-step predictions based on past observations of the exchange rate data are shown in Figure 5. The forecasts are very close to the actual observations. Figure 6 displays a scatterplot of the exchange rate and forecasts (the correlation coefficient is 0.98).

Having fitted a reasonable long-memory model, it will be interesting to find out if this statistically viable model produces useful forecasts. To this end, consider the following scheme. Suppose that on the first day of January of each year, 100 US dollars (or an equivalent amount of

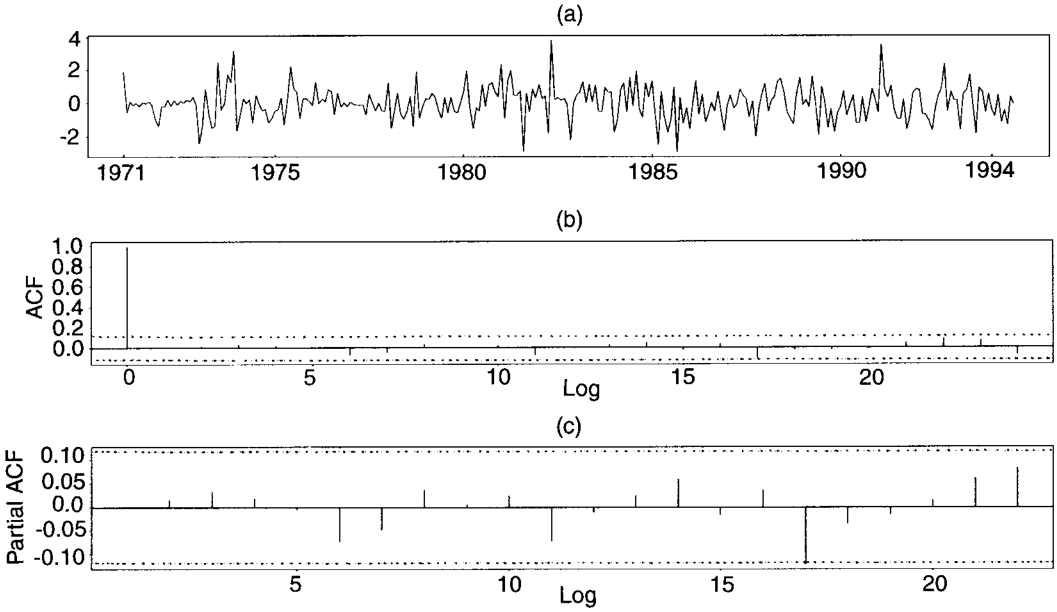

Figure 4. (a) Standardized residuals; (b) ACF of the residuals; (c) PACF of the residuals

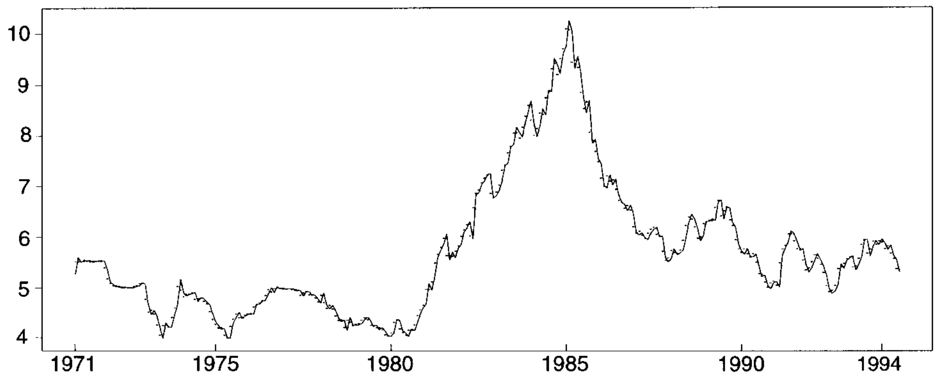

Figure 5. Forecasted and actual exchange rates (French francs to the US dollar). Dots: data; heavy line: forecasts

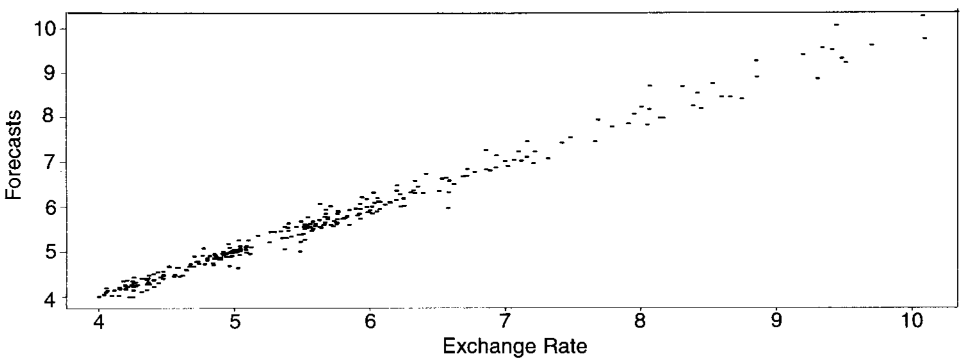

Figure 6. Scatterplot of forecasted and actual exchange rates (French francs to the US dollar). Correlation coefficient: $\hat{\rho} = 0.98$

French francs) are available for investment. If, according to the forecast for January, the exchange rate is going to decline at the end of the month, then $100r_1$ francs are bought, where r_1 is the exchange rate for January. These $100r_1$ francs are invested in a French money market through the end of the month. On the other hand, if the exchange rate is expected to increase, then 100 US dollars are bought and they are invested in the US money market for the current month. At the end of January, a new forecast for February is obtained and all available capital is reinvested. This procedure is repeated month to month throughout the year. Table V presents the annual percentage returns of the strategy, assuming a uniform 8% annual return from the money market. The average annual return is 14.45 with a standard deviation of 9.89. The high volatility of the returns is reflected in Table V: the returns are very high for 1973, 1974, 1980, 1983, 1988 and 1992; the returns for 1975, 1978, 1982, 1984, 1987, 1989 and 1990 are below the average 8% annual return from the money market. Figure 7 displays a time-series plot of the annual percentage returns. The 8% benchmark is also plotted. From this plot, it seems that a year of high return is followed by a year of low return. This is confirmed by the first component of the autocorrelation function of the returns, which is -0.19. Overall, the strategy based on the forecasts outperforms the 8% benchmark in 16 out of 23 years.

Table V. Annual percentage returns: ARFIMA(1, d, 1) model

Year	Return (%)	Year	Return (%)	Year	Return (%)
1971	14.38	1979	12.91	1987	2.72
1972	8.97	1980	26.38	1988	30.38
1973	37.60	1981	13.90	1989	5.60
1974	17.44	1982	3.98	1990	4.47
1975	7.39	1983	28.94	1991	13.08
1976	8.28	1984	7.84	1992	30.62
1977	12.35	1985	14.12	1993	16.95
1978	3.10	1986	11.09		

Table VI. Annual percentage returns: ARMA(0,1) model

Year	Return (%)	Year	Return (%)	Year	Return (%)
1971	14.86	1979	9.71	1987	2.72
1972	8.97	1980	26.38	1988	30.38
1973	26.71	1981	13.90	1989	5.60
1974	16.05	1982	3.98	1990	4.47
1975	7.39	1983	28.94	1991	13.08
1976	9.60	1984	7.84	1992	30.62
1977	9.72	1985	14.12	1993	7.49
1978	3.10	1986	7.37		

As Figure 2 could suggest an MA(1) model, a short-memory ARMA(0,1) model, $y_t = (1 - \theta B)\varepsilon_t, t = 1, \ldots, n$, is also fitted to the foreign exchange data for comparison. This comparison is meaningful since adaptive techniques proposed in Tiao and Tsay (1994) based on short-memory ARMA(1,1) models could provide reasonable forecasts for some long-memory data.

The estimated value of θ is $-0.329(t_\theta = -5.87)$ with the standard deviation of the estimated residual $\hat{\sigma} = 0.0257$. The modified portmanteau statistics is $Q_{20} = 13.34$ with a p-value of 0.71. The forecasting performance of this model is displayed in Table VI. The average

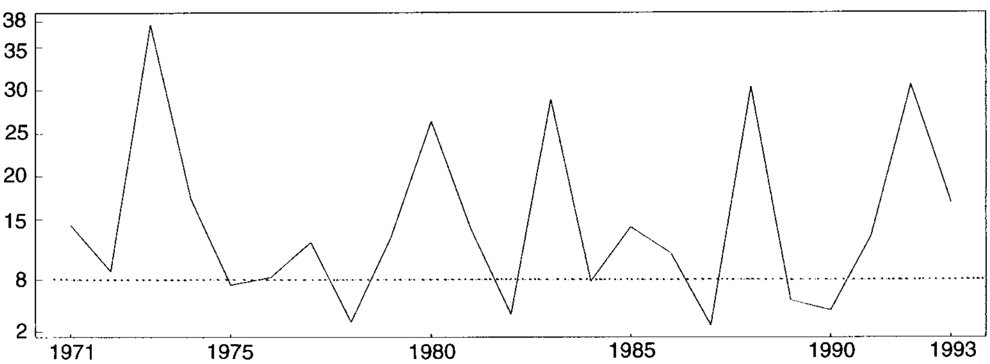

Figure 7. Annual percentage returns. Heavy line: strategy based on ARFIMA forecasts; dotted line: 8% benchmark

Table VII. Parameter estimation with missing observations

Parameter	MLE	t-statistic
d	0.130	2.0
ϕ_1	−0.490	−2.5
θ_1	−0.671	−4.4

Table VIII. Covariance matrix of parameters

	d	ϕ_1	θ_1
d	0.0041	−0.0063	−0.0033
ϕ_1		0.0379	0.0281
θ_1			0.0238

return is 13.17 with a standard deviation of 9.12. Although the returns of this model are similar to the ARFIMA model, in 1973 and 1993 it underperforms the long-memory counterpart for about 10%.

Since economic time series are often incomplete, we also analyse the influence of missing values on the quality of forecasts for the fitted model. For this purpose, the modified Kalman filter equations are used to calculate ML estimates of the missing data for the foreign exchange model. Using the subroutine RNSRI in IMSL, a random sample of 10 locations is selected from 284 observations. The locations of missing observations selected by this procedure are: Oct. 74, Nov. 78, May 79, Nov. 80, Jun. 81, Sep. 84, Jan. 85, Jun. 90, Feb. 92 and Oct. 92. Table VII displays the results from the maximum likelihood estimation with incomplete data.

The residual standard deviation is $\hat{\sigma} = 0.0258$. As displayed in Table VII, all the parameters are significant at the 95% level. However, the fractional parameter d for the incomplete data is slightly smaller than the estimated value for the original data set. Moreover, the variance of the parameter estimates given in Table VIII are slightly greater than those shown in Table II.

Figure 8 shows the time-series plots for the foreign exchange data. In order to achieve a higher resolution, the full period 1971–94 is broken down into four subperiods. The figure also includes the one-step predictions and 95% confidence bands. The gaps in the time series plots correspond to the locations of missing values.

Figure 9 plots the root mean square prediction error (RMSPE), $r_t = \sqrt{E[(Y_t - \hat{Y}_t)^2]}$. The jumps in the sequence $\{r_t\}$ introduced by the ten missing observations are clear from those plots. In most cases, the change on the RMSPE persists for about 8 months.

The percentage changes between the forecasts with or without missing values are displayed in Figure 10 and most of those changes are less than 1%. Thus, missing observations do not seem to affect the forecasting power of this ARFIMA model substantially.

To assess the effect of the missing data on the performance of the business strategy discussed earlier, the average of annual returns are plotted in Figure 11. For these randomly selected locations, the performance of the new forecasts is comparable to the original predictions.

A similar study is also carried out for 84 other missing at random samples each of which consists of ten missing observations. Their average annualized returns are displayed in Figure 12. Also, 95% confidence bands are plotted. As shown, most of the time the returns are above the 8% benchmark.

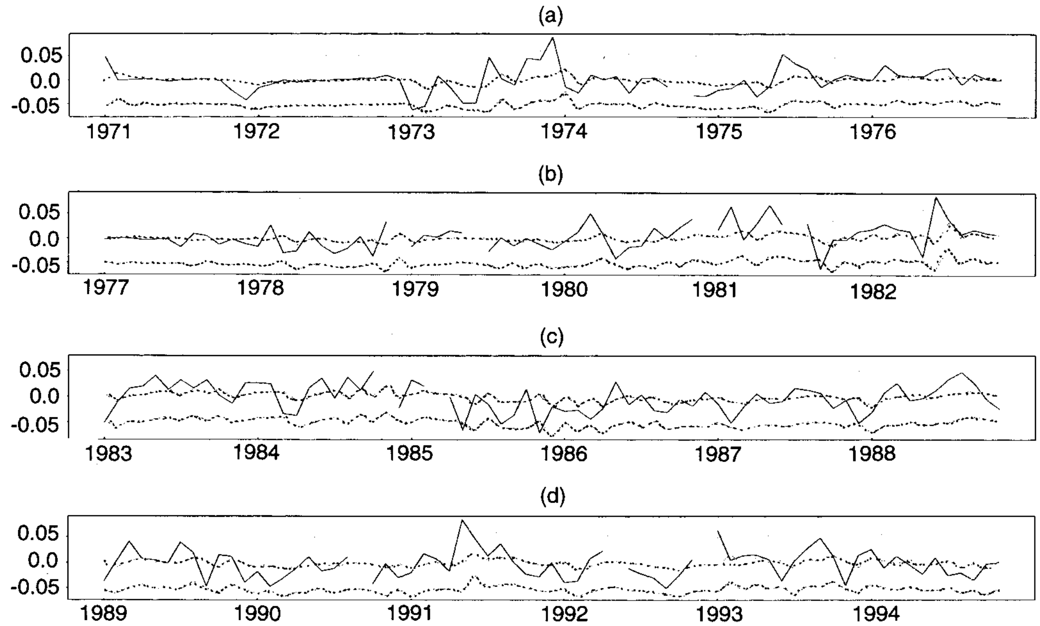

Figure 8. Time-series plots of foreign exchange data. (a) 1971–6 period; (b) 1977–82 period; (c) 1983–8 period; (d) 1989–94 period. Heavy lines: data; broken lines: one-step predictions; dotted lines: upper and lower 95% confidence bands

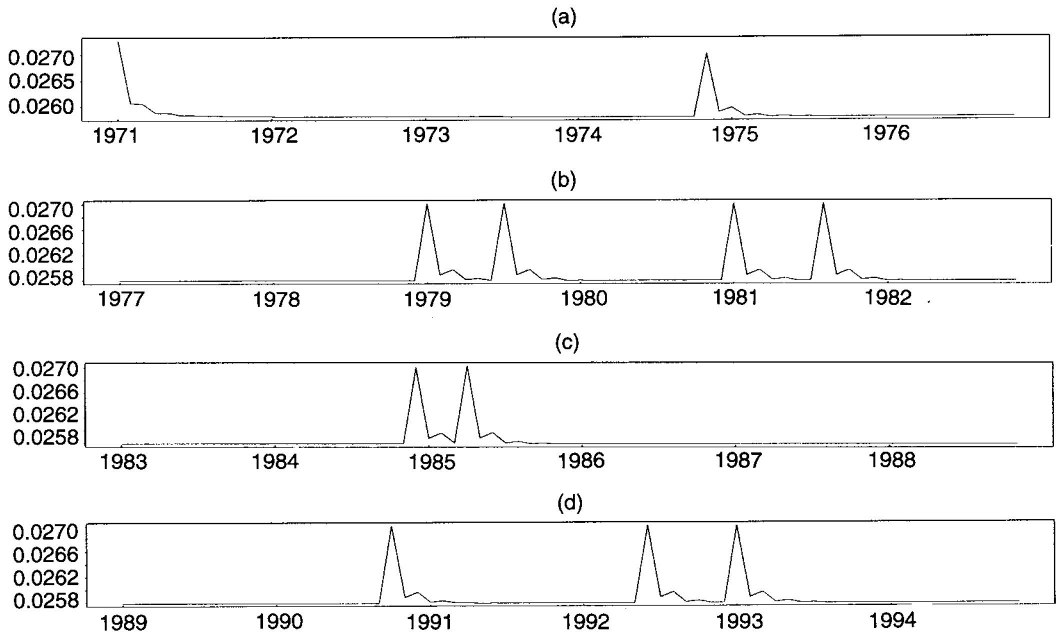

Figure 9. RMSPE. (a) 1971–6 period; (b) 1977–82 period; (c) 1983–8 period; (d) 1989–94 period

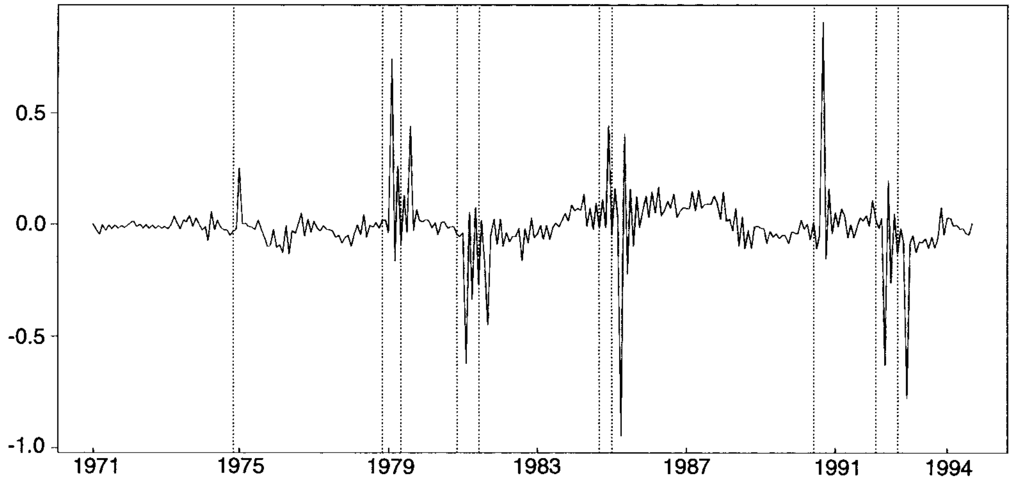

Figure 10. Percentage change of forecasts. Heavy lines: percentage; vertical dotted lines: locations of missing values

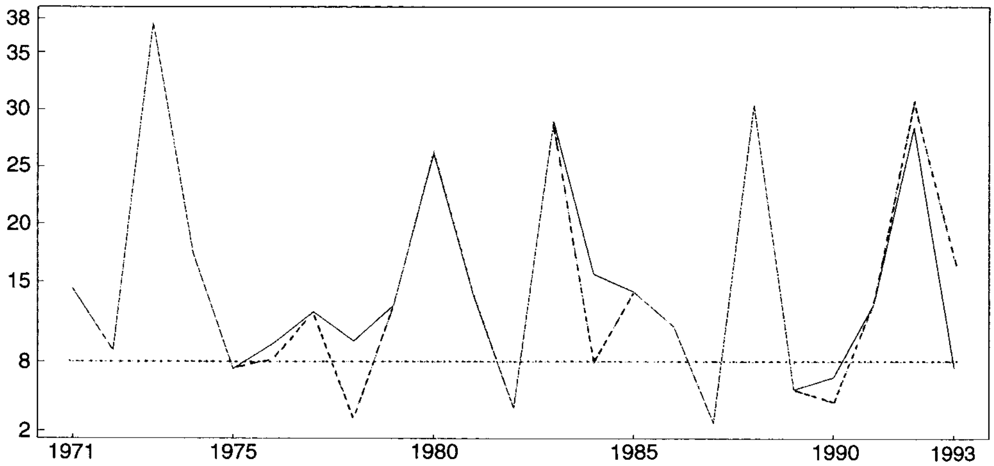

Figure 11. Annualized returns. Heavy line: missing value estimation; broken line: complete data estimation; horizontal dotted line: 8% benchmark

According to the previous analysis, French exchange rates seem to exhibit long-memory behaviour. This conclusion agrees with the study of international foreign exchange data by Cheung (1993). Furthermore, the fitted ARFIMA$(1, d, 1)$ model provides forecasts that are not greatly affected by missing observations.

CONCLUSIONS

The problems of maximum likelihood estimation and forecasting of long-memory processes are studied by means of the state-space models and the Kalman filter. Exact and approximate ML estimates are considered. Since computations of exact ML estimates are cumbersome for data

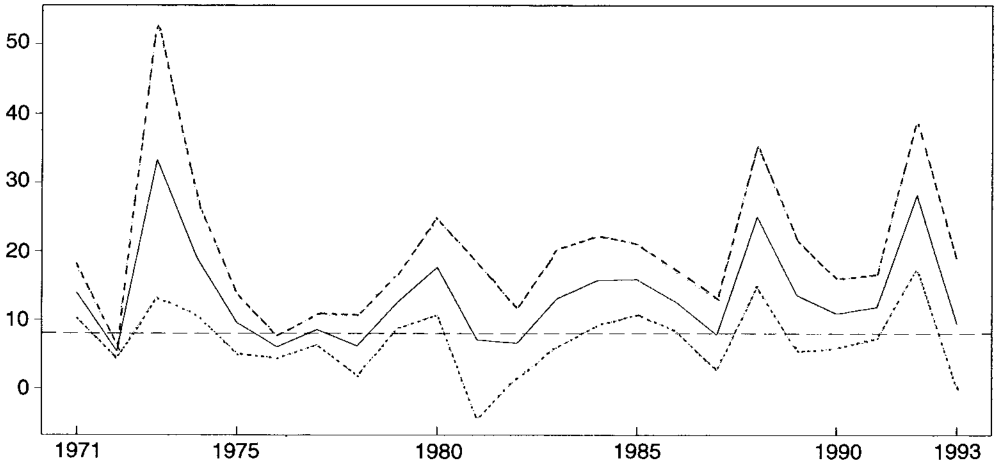

Figure 12. Average annualized returns. Heavy line: missing value estimations; horizontal broken line: 8% benchmark; dotted lines: 95% confidence bands

sets with moderate to large samples, approximate ML estimates obtained by truncating the state-space representation are proposed to facilitate the fitting of ARFIMA models. Numerical calculations are reduced from an order of n^3 for the evaluation of the exact likelihood function to an order of n for the approximate likelihood. As demonstrated in this paper, the proposed method performs well with a foreign exchange data set. It is hoped that these techniques will prove to be useful for analyzing and forecasting long memory time series with missing observations.

APPENDIX

The coefficients of expansion (2) can be calculated as follows. By definition,

$$\psi(z) = \frac{\Theta(z)(1-z)^{-d}}{\Phi(z)}.$$

Define $\varphi(z) = \Theta(z)/\Phi(z)$ and $\eta(z) = (1-z)^{-d}$. Thus, φ satisfies: $\Phi(z)\varphi(z) = \Theta(z)$. Therefore, $\sum_{i=0}^{p} \phi_i z^i \sum_{j=0}^{\infty} \varphi_j z^j = \sum_{j=0}^{q} \theta_j z^j$, and then, $\sum_{j=0}^{\infty}(\sum_{i=0}^{p} \phi_i \varphi_{j-i}) z^j = \sum_{j=0}^{q} \theta_j z^j$. Hence, $\sum_{i=0}^{p} \phi_i \varphi_{j-i} = \theta_j$, and then, $\varphi_j = \theta_j - \sum_{i=1}^{p} \phi_i \varphi_{j-i}$, with $\varphi_0 = 1$ and $\theta_j = 0$, for $j > q$.

On the other hand, the coefficients of $\eta(z) = \sum_{j=0}^{\infty} \eta_j z^j$ are given by

$$\eta_j = \frac{\Gamma(1-d)}{\Gamma(j+1)\Gamma(1-d-j)}.$$

Thus, based on the expansions of $\varphi(z)$ and $\eta(z)$, the coefficients ψ_j can be calculated as follows:

$$\psi(z) = \varphi(z)\eta(z) = \left(\sum_{i=0}^{\infty} \varphi_i z^i\right)\left(\sum_{j=0}^{\infty} \eta_j z^j\right) = \sum_{i=0}^{\infty}\left(\sum_{j=0}^{i} \varphi_j \eta_{i-j}\right) z^i$$

hence, $\psi_i = \sum_{j=0}^{i} \varphi_j \eta_{i-j}$, for $i \geqslant 0$. □

ACKNOWLEDGEMENTS

This research is supported in part by the Office of Naval Research Grant N00014-89-J-1851 and an Earmarked Research Grant from the Research Grant Council of Hong Kong. The authors wish to thank a referee for helpful comments which led to an improved version of this paper.

REFERENCES

Ansley, C. and Kohn, R., 'Estimation, filtering and smoothing in state space models with incompletely specified initial conditions', *The Annals of Statistics*, **13** (1985), 1286–1316.

Beran, J., *Statistics for Long-memory Processes*, New York: Chapman and Hall (1994).

Chan, N. H. and Palma, W., 'State-space modeling of long-memory time series', Technical Report, Department of Statistics, Carnegie Mellon University, Pittsburgh (1996).

Cheung, Y. W., 'Long memory in foreign-exchange rates', *Journal of Business & Economics Statistics*, **11** (1993), 93–101.

Crato, N. and Ray, B. K., 'Model selection and forecasting for long-range dependent processes', *Journal of Forecasting*, **15** (1996), 107–25.

Hannan, E. J. and Deistler, M., *Statistical Theory of Linear Systems*, New York: Wiley (1988).

Harvey, A. C. and Pierse, R. G., 'Estimating missing observations in economic time series', *J. Amer. Statist. Assoc.*, **79** (1984), 125–31.

Hosking, J. R. M., 'Modeling persistence in hydrological time series using fractional differencing', *Water Resources Research*, **20** (1984), 1898–1908.

Jones, R. H., 'Maximum likelihood fitting of ARMA models to time series with missing observations', *Technometrics*, **22** (1980), 389–95.

Kohn, R. and Ansley, C. F., 'Estimation, prediction and interpolation of ARIMA models with missing data', *J. Amer. Statist. Assoc.*, **81** (1986), 751–61.

Li, W. K. and McLeod, A. I., 'Fractional time series modeling', *Biometrika*, **73** (1986), 217–21.

Ljung, G. M. and Box, G. E. P., 'On a measure of lack of fit in time series models', *Biometrika*, **65** (1978), 297–303.

Shumway, R. H. and Stoffer, D. S., 'An approach to time series smoothing and forecasting using EM algorithm', *J. Time Series Analysis*, **3** (1982), 253–64.

Tiao, G. C. and Tsay, R. S., 'Some advances in non-linear and adaptive modelling in time series', *Journal of Forecasting*, **14** (1994), 109–31.

Authors' biographies:

Wilfredo Palma received his PhD in Statistics from Carnegie Mellon University. He is currently an Assistant Professor of Statistics at Pontificia Universidad Católica de Chile. His research interests include time series analysis and forecasting.

Ngai Hang Chan is Professor of Statistics, Department of Statistics at Carnegie Mellon University. His research interests include time series analysis, econometrics, and modelling and analysis of environmental and oceanographic data.

Authors' addresses:

Wilfredo Palma, Departamento de Probabilidad y Estadística, Universidad Católica de Chile, Casilla 306, Santiago 22, Chile. E-mail: wilfredo@mat.puc.cl.

Ngai Hang Chan, Department of Statistics, Carnegie Mellon University, Baker Hall 132, Pittsburgh, PA 15213, USA. E-mail: chan@stat.cmu.edu.

