

A Generalized ARFIMA Process with Markov-Switching Fractional Differencing Parameter

Wen-Jen Tsay* and Wolfgang Karl Härdle†
(Received 00 Month 200x; In final form 00 Month 200x)

* The Institute of Economics, Academia Sinica, Taiwan

† CASE – Center for Applied Statistics and Economics
Humboldt-Universität zu Berlin,
Spandauer Straße 1, 10178 Berlin, Germany

We propose a general class of Markov-switching-ARFIMA processes in order to combine strands of long memory and Markov-switching literature. Although the coverage of this class of models is broad, we show that these models can be easily estimated with the DLV algorithm proposed. This algorithm combines the Durbin-Levinson and Viterbi procedures. A Monte Carlo experiment reveals that the finite sample performance of the proposed algorithm for a simple mixture model of Markov-switching mean and ARFIMA(1, d , 1) process is satisfactory. We apply the Markov-switching-ARFIMA models to the U.S. real interest rates and the Nile river level data, respectively. The results are all highly consistent with the conjectures made or empirical results found in the literature. Particularly, we confirm the conjecture in Beran and Terrin [?] that the observations 1 to about 100 of the Nile river data seem to be more independent than the subsequent observations, and the value of differencing parameter is lower for the first 100 observations than for the subsequent data.

Key words: Markov chain; ARFIMA process; Viterbi algorithm; Long memory

AMS 2000 Subject classification: 60G18; 60J22

1 Introduction

It is well known that many time series data exhibit long memory, or long-range dependence, including the Nile river level, *ex post* real interest rate,

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 ‘Economic Risk’. We would like to thank Professor Tim Bollerslev of Duke University, an anonymous referee, an Associate Editor and the Editor for their valuable comments and suggestions. * Corresponding author. The Institute of Economics, Academia Sinica, Taipei, Taiwan, R.O.C. Tel: (886-2) 2782-2791 ext. 296. Fax: (886-2) 2785-3946. E-Mail: wtsay@ieas.econ.sinica.edu.tw

forward premium, and the dynamics of aggregate partisanship and macroideology. Among the many other examples that Beran [?] gives the Nile river data has been known for its long memory behavior since ancient times, and this is one of the time series that led to the discovery of the Hurst effect (Hurst [?]) and motivated Mandelbrot and his co-workers (Mandelbrot and van Ness [?]; Mandelbrot and Wallis [?]) to introduce fractional Gaussian noise to model long memory phenomenon.

Long range dependence also has been observed in financial data. As demonstrated by Ding et al. [?], de Lima and Crato [?] and Bollerslev and Mikkelsen [?] that the volatility of most financial time series exhibits strong persistency and can be well described as a long memory process. Evidence of financial market volatility's strong persistency inspired Breidt et al. [?] to propose a class of long memory stochastic volatility (LMSV) models. Deo et al. [?] also show that the LMSV model is useful for forecasting realized volatility (RV) which is an important quantity in finance.

Figure 1 displays the yearly Nile river minima based on measurements at the Roda gauge near Cairo during the years 622-1284. Beran [?] (p.33) documents that "When one only looks at short time periods, then there seem to be cycles or local trend. However, looking at the whole series, there is no apparent persisting cycle." The changing pattern of the Nile river data leads Bhattacharya et al. [?] to argue that the so-called Hurst effect can also be explained as if the observations are composed as the sum of a weakly dependent stationary process and a deterministic function. As a consequence it is important to distinguish between a long memory time series and a weakly dependent time series with change-points in the mean. This question has been intensively considered in the literature, including Künsch [?] and Heyde and Dai [?]. Berkes et al. [?] presents an overview about this strand of literature. Similarly, Diebold and Inoue [?] shows that long memory also may be easily confused with a Markov-switching mean. Thus, most of the existing literature considers long memory as a competing modeling framework against the structural change and Markov-switching models.

< Insert Figure 1 here >

The Nile river level time series is far more complicated than a pure long memory or a weakly dependent time series with change-points in the mean to describe. Beran and Terrin (BT) [?] suggest therefore that the Hurst parameter characterizing the yearly Nile river might change over time. When estimating the Nile river data with the autoregressive fractionally-integrated moving-average (ARFIMA) or $I(d)$ process introduced by Granger [?], Granger and

Joyeux [?] and Hosking [?], Beran and Terrin [?] (p.629) show that the data can be well fitted with an ARFIMA(0, d , 0) model with $d = 0.4$, where the fractional differencing parameter d of ARFIMA process acts like the Hurst parameter H of fractional Gaussian noise in characterizing the hyperbolic decay of the autocovariance function of a long memory process. BT further claim that the observations 1 to about 100 seem to be more independent than the subsequent observations, and the value of the fractional differencing parameter might be *lower* for the first 100 observations than for the subsequent data. If this claim is right, then there should be a structural change in the long range persistence of the Nile river data around the year 720, and the Nile river data neither can be described with a pure long memory nor a weakly dependent time series with change-points in the mean.

The possible change of the differencing parameter stimulate BT to propose a statistic for testing the stability of the fractional differencing parameter. This testing statistic has been further discussed and extended in Horváth and Shao [?] and Horváth [?]. However, their methods can not identify the change points of the fractional differencing parameter. A Bayesian random persistent-shift (RPS) method for detecting structural change in the differencing parameter and the process level has been considered in Ray and Tsay [?]. Nevertheless, Ray and Tsay [?] did not propose a model which embeds the Markov-switching and long memory literature into a unified framework. Haldrup and Nielsen [?] indeed consider a regime switching long memory model for describing electricity prices. An important feature of the model is that the price processes in the different regimes can have different degrees of long memory. However, the state variable considered in Haldrup and Nielsen [?] (p.350) is observable as opposed to a standard Hamilton [?] Markov-switching model where the state variable is latent. As a consequence, the method proposed in Haldrup and Nielsen [?] cannot deal with the puzzle raised by Diebold and Inoue [?] who show that a mixture model of latent Markov-switching mean can generate the aforementioned long memory dependence.

The above considerations lead us to combine the long memory and Markov-switching literature into a unified framework. We introduce a Markov-switching-ARFIMA (MS-ARFIMA) process by extending the hidden Markov model. The distinguished feature of our model is that the state variable is latent although the process in the different regimes can have different degrees of long memory. Therefore, the MS-ARFIMA model subsumes many interesting models in the literature, including the Markov-switching AR model considered in Hamilton [?]. Moreover, this model also can be used to tackle the puzzle raised in Diebold and Inoue [?] by estimating the differencing parameter allowing for the presence of a Markov-switching mean. Given that the hidden Markov model has become extremely popular in speech recognition as shown in Juang and Rabiner [?] and Qian and Titterton [?], and in econometrics,

finance, genetics, and neurophysiology as outlined in Robert et al. [?], the MS-ARFIMA model provides a flexible modeling framework for many applications to these fields.

The remaining parts of this paper are arranged as follows: Section 2 presents the MS-ARFIMA process and the algorithms for estimating the parameters of interest. In Section 3 we consider the finite sample performance of the proposed algorithm under the simple mixture of a Markov-switching mean and an ARFIMA(1, d , 1) process. We then apply the proposed methodology to the U.S. real interest rates and the Nile river data in Section 4. Section 5 provides a conclusion.

2 Models and Main Results

The objective of this paper is to propose a general class of Markov-switching-ARFIMA processes in order to combine strands of long memory and Markov-switching literature. This class of models offers a rich dynamic mixture of a Markov chain and an $I(d)$ process.

Let $\{s_t\}_{t=1}^T$ be the latent sample path of an N -state Markov chain. At each time s_t can assume only an integer value of $1, 2, \dots, N$, and its transition probability matrix is

$$\mathcal{P} \equiv \begin{bmatrix} p_{11} & p_{21} & \cdots & p_{N1} \\ p_{12} & p_{22} & \cdots & p_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1N} & p_{2N} & \cdots & p_{NN} \end{bmatrix},$$

where $p_{ij} = P(s_t = j \mid s_{t-1} = i)$ and $\sum_{j=1}^N p_{ij} = 1$ for all i .

An $I(d)$ process, x_t , is defined as:

$$(1 - L)^d x_t = h_t,$$

where L is the lag operator ($Lk_t = k_{t-1}$) and h_t is a short memory process. When $d > 0$, the $I(d)$ process is often called the long memory process, because its autocovariance function is not summable so as to capture the long range dependence of a time series. In addition, the $I(d)$ process is nonstationary when $d \geq \frac{1}{2}$, otherwise, it is covariance stationary. The major feature of the $I(d)$ process is that its autocorrelation function declines at a slower hyperbolic

rate (instead of the geometric rate found in the conventional ARMA models):

$$\rho(h) = \frac{\Gamma(h+d)\Gamma(1-d)}{\Gamma(h-d+1)\Gamma(d)},$$

where $\rho(h)$ denotes the autocorrelation function of x_t at lag h when x_t is an ARFIMA(0, d ,0) process and $\Gamma(\cdot)$ is the gamma function. By Stirling's formula, we note that

$$\rho(h) \sim \frac{\Gamma(1-d)}{\Gamma(d)} h^{2d-1}, \quad \text{as } h \rightarrow \infty.$$

Combining the defining feature of a Markov chain and that of an $I(d)$ process, we propose the following MS-ARFIMA(p, d, q) process:

$$w_t = \mu_{s_t} I\{t \geq 1\} + (1-L)^{-d_{s_t}} \sigma_{s_t} z_t I\{t \geq 1\} = \mu_{s_t} I\{t \geq 1\} + y_t, \quad (1)$$

where w_t is the observed time series of interest, $I\{\cdot\}$ is the indicator function and z_t is a stationary process with mean zero and bounded positive spectral density $f_u(\lambda) \sim G_0$ as $\lambda \rightarrow 0$ at each possible regime, thus including stationary and invertible ARMA process as its special case. The role of $I\{\cdot\}$ is to truncate the influence of the infinite past observations of z_t on w_t because we allow d_{s_t} to be greater than or equal to $\frac{1}{2}$.

The most distinguished feature of the process is that the fractional differencing parameter d_{s_t} well known in the long memory literature is allowed to be a Markov chain satisfying the following Assumption A:

Assumption A. s_t is independent of z_τ for all t and τ .

The model in (1) subsumes many interesting models in the literature. When $N = 1$,

$$w_t = \mu_0 + (1-L)^{-d_0} \sigma_0 z_t I\{t \geq 1\} \quad (2)$$

and can be represented as:

$$w_t = \mu_0 + \sum_{k=0}^{t-1} \frac{(d_0)_k}{k!} \sigma_0 z_{t-k}, \quad (3)$$

where

$$(d_0)_k = \frac{\Gamma(d_0+k)}{\Gamma(d_0)} = (d_0)(d_0+1)\dots(d_0+k-1) \quad (4)$$

is Pochhammer's symbol for the forward factorial. Moreover, under the model in (1) and $d_{s_t} = 0$, w_t still includes the Markov-switching AR model considered in Hamilton [?] as one of its special cases.

It is well known that the model in (1) cannot be estimated with the recursive algorithm in Hamilton [?] even though we assume z_t in (1) is a zero mean independently and identically distributed white noise (i.i.d.). The reason is that the possible routes of states running from time 1 to time T expand exponentially to be N^T if we want to extract z_t to conduct maximum likelihood estimation (MLE). Moreover, this model cannot be written in a state-space form due to the presence of fractionally differencing parameter. This implies that we cannot apply the EM algorithm considered in Hamilton [?] for the model in (1), because the non-Markovian nature of the model in (1) prevents us from using the results in (4.2) of Hamilton [?]. We will show that the estimation of the model in (1) can be easily implemented with the algorithm proposed in this paper, even though the parameter estimation from a noisy version of realizations of Markov models is extremely difficult in all but very simple examples as well documented in Qian and Titterton [?].

Let the total sample size be T , and denote $\mathcal{W}_t \equiv (w_1, w_2, \dots, w_t)^\top$ the column vector containing the observations from time 1 to time t , while $\mathcal{S}_t = (s_1, s_2, \dots, s_t)^\top$ represents the corresponding states, and $\mathcal{Y}_t = (y_1, y_2, \dots, y_t)^\top$ in (1) is similarly defined. The column vector $\alpha = (\mu_1, \dots, \mu_N, \sigma_1, \dots, \sigma_N, \phi_{11}, \dots, \phi_{1p}, \phi_{21}, \dots, \phi_{Np}, d_1, \dots, d_N, \theta_{11}, \dots, \theta_{Nq})^\top$ and p_{ij} (transition probabilities) consist of the parameters characterizing the conditional density function (cdf) of w_t . After stacking the parameter vector α and the transition probabilities p_{ij} into one column vector ξ , we can represent the cdf of w_t as $f(w_t | \mathcal{S}_t, \mathcal{W}_{t-1}; \xi)$, clearly showing that the cdf of w_t depends on the entire past routes of states (in general). Indeed, there are N^T possible paths of states running throughout the observations \mathcal{W}_T .

To illustrate the proposed algorithm for the model in (1), we first consider the simplest case where w_t in (1) is generated as:

$$w_t = \mu_{s_t} I\{t \geq 1\} + (1 - L)^{-d_0} \sigma_0 \varepsilon_t I\{t \geq 1\} = \mu_{s_t} I\{t \geq 1\} + y_t, \quad (5)$$

where $d < \frac{1}{2}$ and ε_t is a zero mean normally i.i.d. process with $E(\varepsilon_t^2) = 1$. That is, w_t in (5) is a special type of MS-ARFIMA(0, d , 0) process whose differencing parameter is fixed across different regimes. Under Assumption A and $\varepsilon_t \sim N(0, 1)$ i.i.d. process, the likelihood function of \mathcal{W}_T , $L(\mathcal{S}_T, \mathcal{W}_T; \xi)$

hereafter, for the hidden Markov model in (5) equals

$$L(\mathcal{S}_T, \mathcal{W}_T; \xi) = (2\pi)^{-T/2} |\Lambda|^{-1/2} \exp\left(-\frac{1}{2} \mathcal{Y}_T^\top \Lambda^{-1} \mathcal{Y}_T\right) \prod_{t=1}^T P(s_t | s_{t-1}), \quad (6)$$

where $\Lambda = E(\mathcal{Y}_T \mathcal{Y}_T^\top)$, and $P(s_1 | s_0)$ is evaluated with the unconditional probability that the process will be in regime s_1 . Given that y_t in (5) is a simple ARFIMA(0, d , 0) process, we can use the Durbin-Levinson algorithm to derive

$$(2\pi)^{-T/2} |\Lambda|^{-1/2} \exp\left(-\frac{1}{2} \mathcal{Y}_T^\top \Lambda^{-1} \mathcal{Y}_T\right) = \prod_{t=1}^T (2\pi)^{-1/2} v_{t-1}^{-1/2} \exp\left\{-\frac{(y_t - \hat{y}_t)^2}{2v_{t-1}}\right\}, \quad (7)$$

where \hat{y}_t denotes the one-step ahead predictor of y_t with the observation \mathcal{Y}_{t-1} as $j \geq 2$, and v_{t-1} is the corresponding one-step ahead prediction variance. Deriche and Tewfik [?] also have employed the Durbin-Levinson algorithm to estimate a univariate ARFIMA(0, d , 0) processes without Markov-switching characteristic. As a result, the likelihood function in (6) can be rewritten as:

$$L(\mathcal{S}_T, \mathcal{W}_T; \xi) = \prod_{t=1}^T (2\pi)^{-1/2} v_{t-1}^{-1/2} \exp\left\{-\frac{(y_t - \hat{y}_t)^2}{2v_{t-1}}\right\} P(s_t | s_{t-1}), \quad (8)$$

indicating that the *unconditional* likelihood function of the mixture model in (5) can be exactly and recursively evaluated provided that we can identify the true path of s_t , \mathcal{S}_T^* .

We do not know in reality the value of \mathcal{S}_T^* . However, the recursive structure shown in (8) is especially suitable for implementing the Viterbi [?] algorithm well-known in the digital communication literature to identify the most likely path of states among the N^T possible routes within \mathcal{W}_T . See Forney [?] for an interesting overview of the Viterbi algorithm. Basically, the Viterbi algorithm is the standard forward dynamic programming solution to maximum-likelihood decoding of a discrete-time, finite-state dynamic system observed in i.i.d. noise as documented in Omura [?]. Our strategy is to combine the Durbin-Levinson algorithm and the Viterbi algorithm by suggesting a *Durbin-Levinson-Viterbi* (DLV) algorithm for the model in (5). When compared to the original Viterbi algorithm designed for solving the problem of maximum a posteriori probability estimate of the state sequence of a finite-state discrete-time Markov process observed in white noise, the DLV algorithm proposed in this paper is concerned with the hidden Markov process observed in a much more general ARFIMA noise. As a consequence, the DLV algorithm can be used to

tackle the puzzle raised by Diebold and Inoue [?] that long memory can be easily confused with a Markov-switching mean, because the DLV algorithm can estimate the differencing parameter of a time series allowing for the presence of a Markov-switching mean.

To locate the most likely path running through the data \mathcal{W}_T with the idea of Viterbi [?], we note first that, for each time t , there are N possible states ending at time t , i.e., $(s_t = i)$, $i = 1, \dots, N$. For a particular node of these N end points at time t , say $(s_t = j)$, there exists a corresponding most likely path:

$$(\mathcal{S}_{t-1}(s_t = j), s_t = j) = (s_1(s_t = j), s_2(s_t = j), \dots, s_{t-1}(s_t = j), s_t = j), \quad (9)$$

which ends at this particular node $(s_t = j)$. We refer to the path $(\mathcal{S}_{t-1}(s_t = j), s_t = j)$ in (9) as the *survivor* associated with the node $(s_t = j)$. Note that, with little loss of clarity, we do not explicitly specify that the path depends on the parameter ξ and the observations \mathcal{W}_t in order to simplify the notation. The likelihood function generated from this survivor $(\mathcal{S}_{t-1}(s_t = j), s_t = j)$ and the formula in (8) is recorded as $L(\mathcal{S}_{t-1}(s_t = j), s_t = j, \mathcal{W}_t; \xi)$ and is crucial for locating the most likely path running from time 1 to time T . In short, for each node $(s_t = j)$ at time t , there exists a most likely path, survivor $(\mathcal{S}_{t-1}(s_t = j), s_t = j)$, and its associated likelihood function $L(\mathcal{S}_{t-1}(s_t = j), s_t = j, \mathcal{W}_t; \xi)$. Most importantly, the number of survivors at each time t is always equal to N .

Given the N survivors at time t and in order to locate the survivor $(\mathcal{S}_t(s_{t+1} = i), s_{t+1} = i)$ for a particular node $(s_{t+1} = i)$ at time $t + 1$, among the N segments connecting the node $(s_{t+1} = i)$ and the N time- t survivors $(\mathcal{S}_{t-1}(s_t = j), s_t = j)$ recorded at time t , we select the one producing the largest likelihood function $L(\mathcal{S}_t(s_{t+1} = i), s_{t+1} = i, \mathcal{W}_{t+1}; \xi)$ among these N possible candidates, and name it as the survivor $(\mathcal{S}_t(s_{t+1} = i), s_{t+1} = i)$ for this particular node $(s_{t+1} = i)$. The computation of the aforementioned likelihoods is simple, because we record the likelihood functions of the N time- t survivors at each time t .

This recursive updating process proceeds from time 1 to time T and results in N time- T survivors $(\mathcal{S}_{T-1}(s_T = i), s_T = i)$ and their associated likelihood function $L(\mathcal{S}_{T-1}(s_T = i), s_T = i, \mathcal{W}_T; \xi)$, for each $i = 1, \dots, N$. From these N time- T survivors we select the one producing the largest likelihood function, say $L(\mathcal{S}_{T-1}(s_T = g), s_T = g, \mathcal{W}_T; \xi)$, as the most likely path running from time 1 to time T . Combining a numerical optimization procedure and this chosen likelihood function $L(\mathcal{S}_{T-1}(s_T = g), s_T = g, \mathcal{W}_T; \xi)$ generated from the Viterbi algorithm and the Durbin-Levinson algorithm displayed in (7), we can estimate the parameters ξ and identify the states \mathcal{S}_T hidden in the observations \mathcal{W}_T .

We now consider another special type of MS-ARFIMA(p, d, q) process:

$$w_t = \mu_{s_t} I\{t \geq 1\} + y_t = \mu_{s_t} I\{t \geq 1\} + (1-L)^{-d_0} \sigma_0 z_t I\{t \geq 1\}, \phi(L)z_t = \theta(L)\varepsilon_t, \quad (10)$$

where

$$\phi(L) = 1 - \phi_1 L - \dots - \phi_p L^p, \quad \theta(L) = 1 + \theta_1 L + \dots + \theta_q L^q, \quad (11)$$

and the roots of the polynomial $\phi(L)$ and those of $\theta(L)$ in (11) are all outside the unit circle and share no common roots. The model in (10) is much more general than that in (5), but still can be estimated with the preceding Viterbi algorithm after some modifications. Please note that the value of fractional differencing parameter is unchanged across different regimes as that imposed in (5).

Note that the term y_t in (10) can be rearranged as

$$y_t = (1-L)^{-d_0} \sigma_0 \phi(L)^{-1} \theta(L) \varepsilon_t, \quad t = 1, 2, \dots \quad (12)$$

We then have

$$\phi(L)y_t = (1-L)^{-d_0} \sigma_0 \theta(L) \varepsilon_t = \sigma_0 \theta(L) (1-L)^{-d_0} \varepsilon_t = \sigma_0 \theta(L) \tilde{y}_t, \quad t = 1, 2, \dots, \quad (13)$$

where $\tilde{y}_t = (1-L)^{-d_0} \varepsilon_t$ is an ARFIMA(0, d , 0) process. Dueker and Serletis [?] use the same transformation method for estimating an ARFIMA(p, d, q) process. Conditional on a set of $\phi(L)$ and $\theta(L)$ and a suitable starting value, the *conditional* likelihood function of y_t in (12) can still be evaluated exactly with the transformed ARFIMA(0, d , 0) \tilde{y}_t in (13) and the Durbin-Levinson algorithm defined in (7). For example, conditional on y_0 being equal to 0, we can extract an ARFIMA(0, d , 0) process from an ARFIMA(1, d , 1) process as follows:

$$\sigma_0 \tilde{y}_t = y_t - \phi_1 y_{t-1} - \sigma_0 \theta_1 \tilde{y}_{t-1}, \quad t = 1, \dots, T. \quad (14)$$

Conditional on a set of $\phi(L)$ and $\theta(L)$ and a suitable starting value for the parameter ξ , we can recursively and exactly evaluate the conditional likelihood function of the hidden Markov model using the DLV algorithm proposed previously.

The same idea also applies to the class of MS-ARFIMA(p, d, q) processes in (1) where d can be Markov-switching. However, we cannot use the Durbin-Levinson algorithm when the fractional differencing parameter is allowed to be Markov-switching. Nevertheless, the Viterbi algorithm is still powerful enough to locate the most likely path under this circumstance. That is, conditional on

a suitable starting value for the parameter ξ , we employ the recursive structure inherent in Viterbi algorithms to identify the most likely path running through the data set.

3 Monte Carlo Experiment

In this section we consider a Monte Carlo experiment to demonstrate the finite sample performance of the proposed DLV algorithm on a special version of the model in (1):

$$w_t = \mu_{s_t} I\{t \geq 1\} + (1 - L)^{-d_0} \sigma_0 (1 - \phi_1 L)^{-1} (1 + \theta_1 L) \varepsilon_t I\{t \geq 1\}, \quad (15)$$

where $\varepsilon_t \sim N(0, 1)$ i.i.d. process. We employ three different values of the fractional differencing parameter:

$$d_0 = \{0.2, 0.3, 0.4\}, \quad (16)$$

along with the following parameters:

$$\mu_1 = 4, \quad \mu_2 = 1, \quad \phi_1 = 0.5, \quad \theta_1 = 0.5, \quad p_{11} = p_{22} = 0.95, \quad (17)$$

and σ_0 is chosen to ensure that the variance of the ARFIMA(1, d , 1) noise in (15) is equal to 1 across different configurations. Note that the positive values of d_0 in (16) are chosen to reflect the variations used in the long memory literature. As a consequence, there are eight parameters characterizing the behavior of w_t in (15):

$$\xi = (\mu_1, \mu_2, p_{11}, p_{22}, \sigma_0, d_0, \phi_1, \theta_1)^\top.$$

We estimate ξ by the following transformation function:

$$\xi = \left(\mu_1, \mu_2, \frac{e^{\tilde{p}_{11}}}{1 + e^{\tilde{p}_{11}}}, \frac{e^{\tilde{p}_{22}}}{1 + e^{\tilde{p}_{22}}}, e^{\tilde{\sigma}_0}, \frac{\tilde{d}_0}{1 + 2|\tilde{d}_0|}, \frac{e^{\tilde{\phi}_1} - 1}{e^{\tilde{\phi}_1} + 1}, \frac{e^{\tilde{\theta}_1} - 1}{e^{\tilde{\theta}_1} + 1} \right)^\top = \kappa(\tilde{\xi}),$$

where

$$\tilde{\xi} = (\mu_1, \mu_2, \tilde{p}_{11}, \tilde{p}_{22}, \tilde{\sigma}_0, \tilde{d}_0, \tilde{\phi}_1, \tilde{\theta}_1)^\top,$$

are the parameters really estimated with the DLV algorithm. In order to create a realistic simulation scheme, the inverse function of the preceding transformation function calculated at the true parameter value plus an extra (8×1)

random vector generated from a standard normal distribution is used as the initial values for estimation procedure, i.e., the initial values for $\tilde{\xi}$ is:

$$\tilde{\xi}_0 = \kappa(\xi)^{-1} + N(0, 1),$$

where ξ is set at its true parameter value. Furthermore, all the computations are performed with GAUSS at 3 different sample sizes, $T = 400, 800, 1600$. The choice of these relatively large sample sizes reveal that the computational burden of implementing our algorithm is extremely mild.

For each sample size T , 200 additional values are generated in order to alleviate the impacts of starting values on estimation results. The optimization algorithm used to implement the DLV algorithm is the quasi-Newton algorithm of Broyden, Fletcher, Goldfarb, and Shanno (BFGS) contained in the GAUSS MAXLIK library. The maximum number of iterations for each replication is 100. The first two hundred replications of normal convergence are recorded for the subsequent data analysis.

< Insert Table 1 here >

The simulation results contained in Table 1 reveal that the bias performance from the DLV algorithm is satisfactory (especially when the sample size is larger) for all configurations considered. Moreover, the associated root-mean-squared error (RMSE) mostly decreases with the increasing sample size. The two exceptional cases both occur when $d = 0.4$, i.e., the RMSE of estimating the parameter μ_1 and that of σ_0 as $T = 1600$ is found to be higher than that of μ_1 and that of σ_0 as $T = 800$, respectively. These two observations demonstrate the ability of the DLV algorithm to deal with the mixture model considered in this section. The performance of DLV algorithm for estimating the fractional differencing parameter is particularly displayed with the box-plots in Figure 2. The above-mentioned observations are clearly borne out in this figure.

< Insert Figure 2 here >

4 Empirical Applications

The methodology developed in this paper is motivated by the dynamic pattern of long memory behavior. Evidence has been given by many methods for such a changing covariance behavior of the Nile river. The applications of the proposed MS-ARFIMA model to actual data are far reaching. For that reason, we consider two data set. The first one is the U.S. real interest rates, while the second one is the Nile river data.

4.1 Example with real interest rates

In this subsection we first consider the U.S. ex post monthly real interest rate constructed from monthly inflation and Treasury bill rates from January 1953 to December 1990 in Mishkin [?]. The reason we use the original dataset of Mishkin [?] is to employ it as a benchmark for a clear comparison between the results from the MS-ARFIMA model and those generated from the methodology employed in earlier papers.

The main feature of the real interest rate is that the whole dataset can be split into three subperiods, January 1953-October 1979, November 1979-October 1982, and November 1982-December 1990, because the operating procedure of the monetary authority changed in October 1979 and October 1982 as argued in Mishkin [?]. Another interesting feature of the real interest rate is that the data of these three subperiods can be well described with the ARFIMA models as shown in Tsay [?]. The simultaneous presence of structural break and long memory within the real interest rate allows itself to be an ideal subject to be investigated with the MS-ARFIMA model.

Allowing the break points to be endogenously determined, Table 2 contains the parameter estimates from the following mixture model with a 2-state Markov chain and a normally distributed ARFIMA(1, d , 1) noise:

$$w_t = \mu_{s_t} I\{t \geq 1\} + (1 - L)^{-d_0} \sigma_0 z_t I\{t \geq 1\}, \quad (1 - \phi_1 L) z_t = (1 + \theta_1 L) \varepsilon_t, \quad (18)$$

where ϕ_1 or θ_1 is assumed to be zero depending on the noise specification. Following Hamilton [?], asymptotic standard errors are calculated numerically. These calculated standard errors may not be valid, because the DLV algorithm is not equivalent to the usual full complete-data log-likelihood estimation. Nevertheless, the DLV algorithm provides us a convenient way to identify the most likely regime path beneath the observed data under the model in (18).

< Insert Table 2 here >

Table 2 shows that the estimates of μ_1 , μ_2 , p_{11} , p_{22} , σ_0 , and d_0 from the DLV algorithm are quite robust across all 4 different configurations. Interestingly, the estimated values of p_{11} and p_{22} remain intact regardless which model specification is used. This implies that the transition of states within the real interest rates can be successfully captured with the model in (18) which is designed to characterize the joint presence of long memory and Markov-switching phenomena. This finding also explains why there are two identical break points identified with these four models, i.e., November 1980 and May 1986, respectively. As a consequence, using the model in (18), we endogenously divide the whole real interest rates data into three subperiods as suggested in Mishkin [?].

Figure 3 displays the U.S. monthly ex post real interest rates and the path of estimated switching means generated from the DLV algorithm. Without loss of generality, only the path of the estimated switching means from the specification ARFIMA(0, d , 0) in Table 2 is reported. Figure 3 shows that the model in (18) provides a satisfactory fitting of the U.S. monthly real interest rates. Although the endogenously identified break points are later than the well-known monetary operating procedure change points (October 1979 and October 1982), this finding is quite reasonable, because it takes some time for the ex post real interest rate to adjust its path after new information arrives. This argument is buttressed with the findings in Figure 3 that the endogenously identified break points are more closely connected to the observed path of the U.S. monthly ex post real interest rates than the monetary operating procedure change points are.

< Insert Figure 3 here >

Table 2 also shows that a long memory phenomenon is found in the real interest rate as has been documented in Tsay [?]. Nevertheless, the estimate of the fractional differencing parameter in Table 2 is much lower than that of 0.666 in Table 3 of Tsay [?] where the change points are exogenously determined, and it is more in line with the estimates of 0.204, 0.275, and 0.193 from the individual subperiod data presented in Table 3 of Tsay [?]. This implies that the persistence of long memory in the real interest rate is much more mitigated, once we take the potentially switching mean of the data into account, thus confirming the arguments of Diebold and Inoue [?] that the presence of Markov-switching level might increase the persistence of the data under investigation.

4.2 Example with Nile river data

In this subsection we apply the Viterbi algorithm to the Nile river data with the following model:

$$w_t = \mu_{s_t} I\{t \geq 1\} + (1 - L)^{-d_{s_t}} \sigma_{s_t} \varepsilon_t I\{t \geq 1\}, \quad (19)$$

where N is assumed to be 2, and ε_t in (19) is normally distributed. For the purpose of comparison, we estimate a fixed regime ARFIMA(0, d , 0) model for the Nile river data, i.e., $N = 1$ is imposed on this model. The estimated value of d from such a fixed regime ARFIMA(0, d , 0) model is 0.3986 and is almost identical to the finding in Beran and Terrin [?].

< Insert Table 3 here >

When estimating the model in (19) with the Viterbi algorithm, we find that the value of the differencing parameter in Table 3 is 0.5770 (nonstationary) for one state, and is 0.2143 (stationary) for the other one. In addition, we identify 5 transitions within the Nile river data in the year 720, 805, 815, 878, and 1070. The estimated path of d_{s_t} from the MS-ARFIMA(0, d , 0) model in Table 3 is graphed in Figure 4.

< Insert Figure 4 here >

Most impressively, the first transition date occurs in the year of 720, and the associated estimated value of d_{s_t} within the period 622 to 719 is 0.2143 which is lower than the 0.5770 observed in the other regime. These two findings correspond closely to the conjectures in Beran and Terrin [?] that the observations 1 to about 100 seem to be more independent than the subsequent observations and the value of differencing parameter might be lower for the first 100 observations than for the subsequent data. Moreover, the first estimated transition date is also very close to the time of the construction of the new measuring device in the year 715 noted in Percival and Walden [?] (p.387).

In Figures 5 and 6 we present the observations and the fitted values generated from the estimated parameters displayed in Table 3. It is clear that the fitted value from the MS-ARFIMA(0, d , 0) model is much closer to the real data than that generated from the model whose differencing parameter is not Markov switching. Combining the findings of the likelihood values in Table 3, we find that the MS-ARFIMA(0, d , 0) model is a promising alternative to describe the Nile river data.

< Insert Figure 5 here >

< Insert Figure 6 here >

5 Conclusions

A general class of MS-ARFIMA processes is suggested to combine long memory and Markov-switching models into one unified framework. The coverage of this class of MS-ARFIMA models is far-reaching, but we show that they still can be easily estimated with the original Viterbi algorithm or the DLV algorithm proposed in this paper. In addition, the simulation reveals that the finite sample performance of the DLV algorithm for a simple mixture model of Markov-switching mean and ARFIMA(1, d , 1) process is satisfactory. When applying the MS-ARFIMA models to the U.S. real interest rates and the Nile river level, the estimation results are all highly compatible with the conjectures made in the literature. Accordingly, the MS-ARFIMA model considered in this paper not only can be used to tackle the puzzle raised by Diebold and Inoue [?], but can also find many potential applications in several scientific research fields.

Reference

- [1] Beran, J., 1994, *Statistics for Long-Memory Processes*. Chapman and Hall, New York.
- [2] Hurst, H. E., 1951, Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, **116**, 770–799.
- [3] Mandelbrot, B. B. and van Ness, J. W., 1968, Fractional Brownian motions, fractional noises and applications. *SIAM Review*, **10**, 422–437.
- [4] Mandelbrot, B. B. and Wallis, J. R., 1969, Some long-run properties of geophysical records. *Water Resources Research*, **5**, 321–340.
- [5] Ding, Z., Granger, C. W. J. and Engle, R. F., 1993, A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, **1**, 83–106.
- [6] de Lima, P. and Crato, N., 1993, Long-range dependence in the conditional variance of stock returns. *Proceedings of the Business and Economic Statistics Section*, August 1993 Joint Statistical Meetings, San Francisco.
- [7] Bollerslev, T. and Mikkelsen, H. O. A., 1996, Modeling and pricing long-memory in stock market volatility. *Journal of Econometrics*, **73**, 151–184.
- [8] Breidt, F. J., Crato, N. and de Lima, P., 1998, The detection and estimation of long memory in stochastic volatility. *Journal of Econometrics*, **83**, 325–348.
- [9] Deo, R., Hurvich, C. and Lu, Y., 2006, Forecasting realized volatility using a long-memory stochastic volatility model: Estimation, prediction and seasonal adjustment. *Journal of Econometrics*, **131**, 29–58.
- [10] Bhattacharya, R. N., Gupta, V. K. and Waymire, E., 1983, The Hurst effect under trends. *Journal of Applied Probability*, **20**, 649–662.
- [11] Künsch, H., 1986, Discrimination between monotonic trends and long-range dependence. *Journal of Applied Probability*, **23**, 1025–1030.

- [12] Heyde, C. C. and Dai, W., 1996, On the robustness to small trends of estimation based on the smoothed periodogram. *Journal of Time Series Analysis*, **17**, 141–150.
- [13] Berkes I., Horváth L., Kokoszka, P. and Shao, Q. M., 2006, On discriminating between long-range dependence and changes in mean. *The Annals of Statistics*, **34**, 1140–1165.
- [14] Diebold, F. X. and Inoue, A., 2001, Long memory and regime switching. *Journal of Econometrics*, **105**, 131–159.
- [15] Beran J. and Terrin N., 1996, Testing for a change of the long-memory parameter. *Biometrika*, **83**, 627–638.
- [16] Granger, C. W. J., 1980, Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics*, **14**, 227–238.
- [17] Granger, C. W. J. and Joyeux, R., 1980, An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, **1**, 15–29.
- [18] Hosking, J. R. M., 1981, Fractional differencing. *Biometrika*, **68**, 165–176.
- [19] Horváth, L. and Shao, Q. M., 1999, Limit theorems for the union-intersection test. *Journal of Statistical Planning and Inference*, **44**, 133–148.
- [20] Horváth, L., 2001, Change-point detection in long-memory processes. *Journal of Multivariate Analysis*, **78**, 218–234.
- [21] Ray, B. K. and Tsay, R. S., 2002, Bayesian methods for change-point detection in long-range dependent process. *Journal of Time Series Analysis*, **23**, 687–705.
- [22] Haldrup, N. and Nielsen, M. Ø., 2006, A regime switching long memory model for electricity prices. *Journal of Econometrics*, **135**, 349–376.
- [23] Hamilton, J. D., 1989, A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, **57**, 357–384.
- [24] Juang, B. H. and Rabiner, L. R., 1991, Hidden Markov models for speech recognition. *Technometrics*, **33**, 251–272.
- [25] Qian, W. and Titterton, D. M., 1991, Estimation of parameters in hidden Markov models. *Philosophical Transactions: Physical Sciences and Engineering*, **337**, 407–428.
- [26] Robert, C. P., Rydén, R. and Titterton, D. M., 2000, Bayesian inference in hidden Markov models through the reversible jump Markov chain Monte Carlo method. *Journal of the Royal Statistical Society B*, **62**, 57–75.
- [27] Hamilton, J. D., 1990, Analysis of time series subject to changes in regime. *Journal of Econometrics*, **45**, 39–70.
- [28] Deriche, J. A. and Tewfik, A. H., 1993, Maximum likelihood estimation of the parameters of discrete fractionally differenced Gaussian noise process. *IEEE Transactions on Signal Processing*, **41**, 2977–2989.
- [29] Viterbi, A. J., 1967, Error bounds for convolutional codes and an asymptotic optimum decoding algorithm. *IEEE Transactions on Signal Processing*, **IT-13**, 260–269.
- [30] Forney, Jr. G. D., 1973, The Viterbi algorithm. *Proceedings of the IEEE*, **61**, 268–278.
- [31] Omura, J. K., 1969, On the Viterbi decoding algorithm. *IEEE transactions on Information Theory*, **IT-15**, 177–179.
- [32] Dueker, M. and Serletis, A., 2000, Do real exchange rates have autoregressive unit roots? A test under the alternative of long memory and breaks. Working Paper 2000-016A, Federal Reserve Bank of St. Louis.
- [33] Mishkin, F. S., 1990, What does the term structure of interest rate tell us about future inflation? *Journal of Monetary Economics*, **25**, 77–95.
- [34] Tsay, W. J., 2000, Long memory story of the real interest rate. *Economic Letters*, **67**, 325–330.
- [35] Percival, D. B. and Walden, A. T., 2000, *Wavelet Methods for Time Series Analysis*. Cambridge University Press, Cambridge.

Table 1. Finite sample performance of the DLV algorithm

Parameter		μ_1	μ_2	p_{11}	p_{22}	σ_0	d_0	ϕ_1	θ_1
$d_0 = 0.4$									
$T = 400$	Bias	-0.024	-0.153	0.005	0.003	-0.006	0.093	-0.070	-0.007
	RMSE	1.059	1.165	0.023	0.019	0.061	0.188	0.162	0.081
$T = 800$	Bias	-0.058	-0.205	0.003	0.002	-0.002	0.050	-0.036	-0.008
	RMSE	0.990	1.124	0.013	0.013	0.034	0.137	0.122	0.050
$T = 1600$	Bias	-0.041	-0.209	0.001	0.001	-0.005	0.018	-0.014	0.000
	RMSE	1.015	1.115	0.009	0.010	0.046	0.093	0.088	0.049
$d_0 = 0.3$									
$T = 400$	Bias	-0.056	-0.174	0.006	0.002	-0.005	0.112	-0.079	-0.015
	RMSE	0.960	1.026	0.028	0.018	0.053	0.239	0.207	0.075
$T = 800$	Bias	0.036	-0.115	0.002	0.002	-0.002	0.046	-0.029	-0.010
	RMSE	0.877	0.973	0.013	0.013	0.030	0.159	0.145	0.048
$T = 1600$	Bias	0.055	-0.096	0.001	0.001	0.000	0.023	-0.015	-0.006
	RMSE	0.789	0.876	0.009	0.008	0.007	0.120	0.110	0.031
$d_0 = 0.2$									
$T = 400$	Bias	0.016	-0.144	0.006	0.002	-0.006	0.157	-0.117	-0.017
	RMSE	0.762	0.921	0.029	0.018	0.058	0.279	0.233	0.078
$T = 800$	Bias	0.050	-0.136	0.003	0.002	-0.003	0.084	-0.063	-0.011
	RMSE	0.725	0.803	0.015	0.013	0.037	0.199	0.171	0.057
$T = 1600$	Bias	0.094	-0.118	0.001	0.001	-0.004	0.045	-0.037	-0.003
	RMSE	0.678	0.761	0.010	0.009	0.035	0.146	0.125	0.050

Notes: Simulations are based on 200 replications. The data is generated from the mixture model defined in (15), (16) and (17). DLV algorithm is the Durbin-Levinson-Viterbi algorithm proposed in this paper. Bias is computed as the true parameter minus the corresponding average estimated values.

Table 2. Estimates of Parameters Based on Data for U.S. Monthly Real Interest Rate and the DLV Algorithm

	ARFIMA(0, d , 0)		ARFIMA(0, d , 1)		ARFIMA(1, d , 0)		ARFIMA(1, d , 1)	
	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
μ_1	5.3455	0.7494	5.3168	0.7162	5.3116	0.7124	5.3626	0.7706
μ_2	0.7226	0.4814	0.7194	0.4383	0.7184	0.4322	0.7352	0.4958
p_{11}	0.9833	0.0150	0.9833	0.0150	0.9833	0.0150	0.9833	0.0150
p_{22}	0.9977	0.0023	0.9977	0.0023	0.9977	0.0023	0.9977	0.0023
σ_0	2.5094	0.0831	2.5091	0.0831	2.5091	0.0831	2.4979	0.0827
d_0	0.2225	0.0367	0.2062	0.0520	0.2034	0.0653	0.2337	0.0376
ϕ_1	-	-	-	-	0.0324	0.0946	-0.9847	0.0155
θ_1	-	-	0.0279	0.0663	-	-	0.9675	0.0200
L^*	1079.0875		1079.0009		1078.9918		1077.0173	

Notes: The results are based on the MS-ARFIMA model defined in (18). S.E. stands for the standard error of the estimate. L^* represents the negative of the log-likelihood function of the switching model. DLV algorithm is the Durbin-Levinson-Viterbi algorithm proposed in this paper.

Table 3. Estimates of MS-ARFIMA(0, d , 0) Model based on the Nile River Data

	MS-ARFIMA(0, d , 0)		ARFIMA(0, d , 0)	
	Estimate	S.E.	Estimate	S.E.
μ_1	10.8593	0.6903	11.4847	0.2607
μ_2	11.4939	0.0917	-	-
p_{11}	0.9930	0.0042	-	-
p_{22}	0.9918	0.0050	-	-
σ_1	0.5430	0.0202	0.6995	0.0192
σ_2	0.8143	0.0332	-	-
d_1	0.5770	0.0430	0.3986	0.0309
d_2	0.2143	0.0510	-	-
L^*	687.5642		703.8541	

Notes: The MS-ARFIMA(0, d , 0) model is defined in (19). S.E. stands for the standard error of the estimate based on numerical derivative. L^* represents the negative of the log-likelihood function of the estimated model.

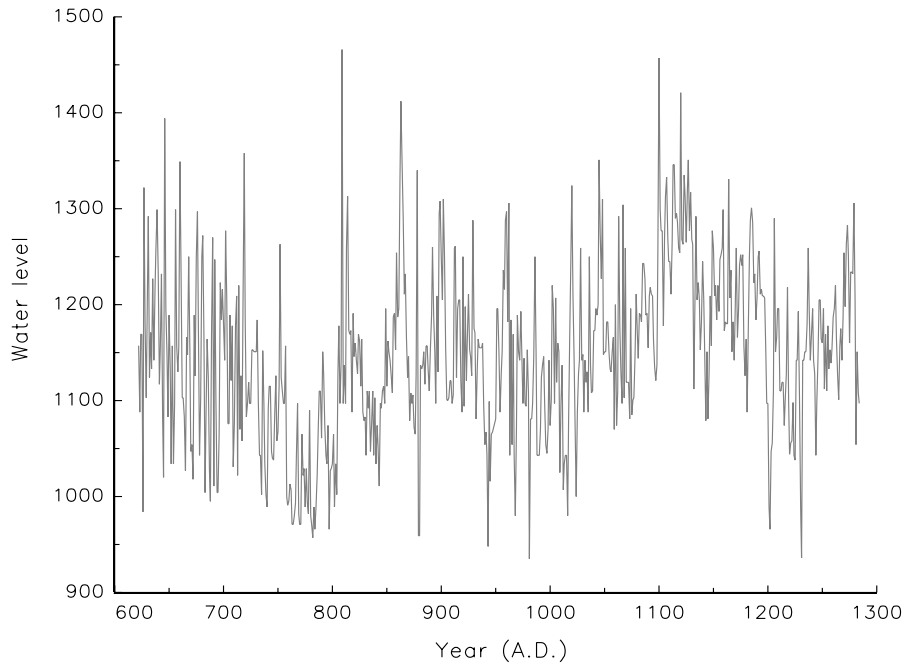


Figure 1. Yearly Nile river minima based on measurements at the Roda gauge near Cairo.

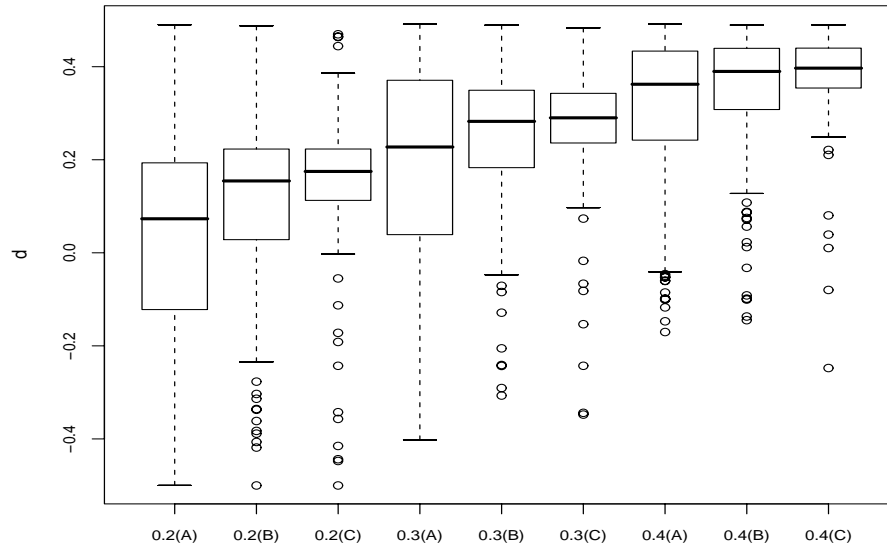


Figure 2. Box-plots of the estimated d from the model defined in (15), (16) and (17) with 200 realizations. The value $f(g)$ denotes the model specification where $f = d$, g denotes the sample size, such that $g = A = 400$, $g = B = 800$, and $g = C = 1600$.

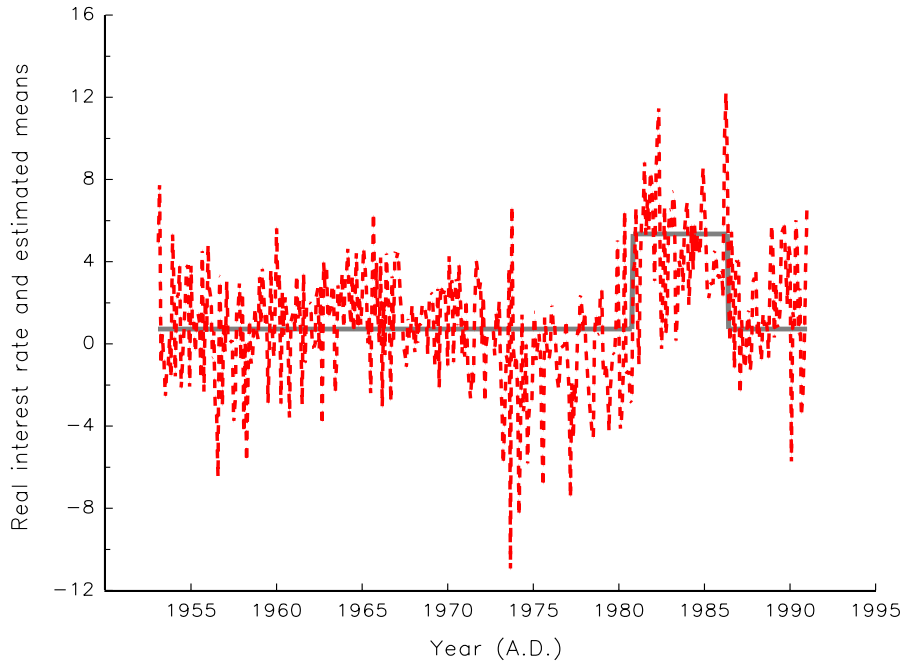


Figure 3. US monthly ex post real interest rates, January 1953-December 1990. Solid line denotes the path of estimated switching means from the specification ARFIMA(0, d , 0) in Table 2, while dotted line denotes the observed monthly ex post real interest rates.

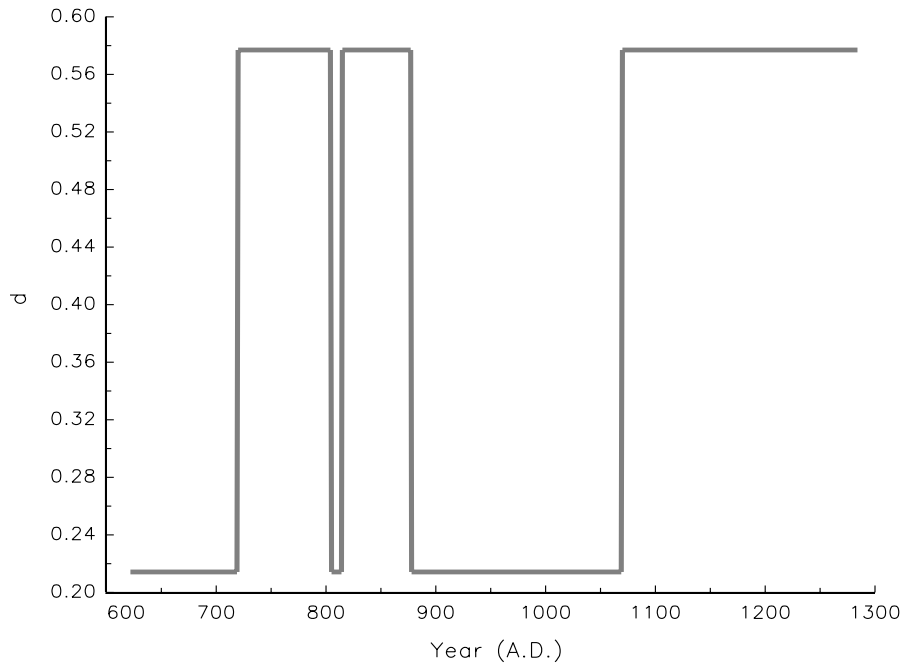


Figure 4. Estimated d_{s_t} from the MS-ARFIMA(0, d , 0) model in Table 3.

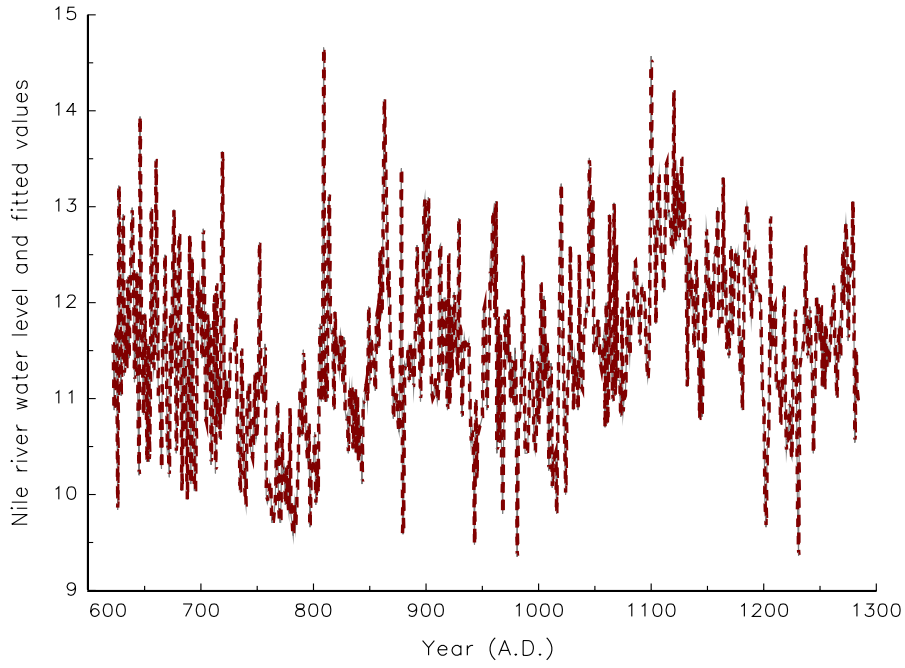


Figure 5. Solid line denotes the Nile river water level divided by 100, while dotted line denotes the corresponding fitted values from the MS-ARFIMA(0, d , 0) model in Table 3.

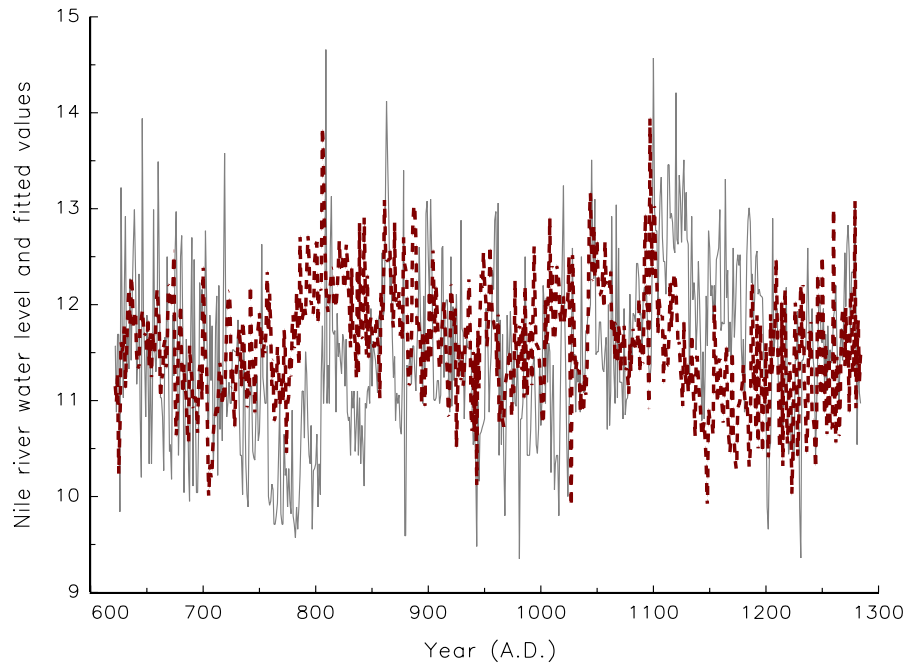


Figure 6. Solid line denotes the Nile river water level divided by 100, while dotted line denotes the corresponding fitted values from the ARFIMA(0, d , 0) model in Table 3.