Journal of Forecasting, *J. Forecast.* **32**, 86–96 (2013)
Published online 20 November 2011 in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/for.1259

Estimation and Forecasting of Locally Stationary Processes

WILFREDO PALMA,[1]* RICARDO OLEA[1] AND GUILLERMO FERREIRA[2]
[1] *Department of Statistics, Pontificia Universidad Católica de Chile, Santiago, Chile*
[2] *Department of Statistics, Universidad de Concepción, Chile*

ABSTRACT

This paper develops a state space framework for the statistical analysis of a class of locally stationary processes. The proposed Kalman filter approach provides a numerically efficient methodology for estimating and predicting locally stationary models and allows for the handling of missing values. It provides both exact and approximate maximum likelihood estimates. Furthermore, as suggested by the Monte Carlo simulations reported in this work, the performance of the proposed methodology is very good, even for relatively small sample sizes. Copyright © 2011 John Wiley & Sons, Ltd.

KEY WORDS Kalman filter; state space system; nonstationarity; long-range dependence; local stationarity; time-varying models

INTRODUCTION

A common problem in the statistical analysis of time series is that stationarity is an assumption hard to justify in practice. To deal with nonstationary data, a number of techniques have been proposed. For instance, differentiation, trend removal and regression analysis are well-known approaches for attempting to convert nonstationary time series into stationary ones. Other methods are based on the concept of evolutionary spectra developed by Priestley (1965) and others. More recently, Dahlhaus (1997) introduced a general class of locally stationary (LS) processes. In this approach, the process under study is nonstationary but with parameters slowly varying over time. Thus, under some regularity conditions, the process can be locally approximated by stationary processes. Other locally stationary processes have been discussed, for example by Wang *et al.* (2001), Cavanaugh *et al.* (2003) and Last and Shumway (2008).

Local stationarity is playing a key role among the methodologies for analyzing nonstationary data (see, for example, Dahlhaus, 2000; Jensen and Witcher, 2000; Dahlhaus and Polonik, 2006, 2009; and Palma and Olea, 2010; among others). In this paper we propose a computationally efficient state space method for estimating, predicting and making statistical inferences about a class of LS models. State space systems have been extensively discussed in the literature (see, for instance, Shumway and Stoffer, 2010, for a recent overview). The state state framework proposed in this paper allows for the statistical modeling of LS processes with short and long memory, and the handling of missing values. Thus exact and approximate maximum likelihood estimates (MLE), one-step and multi-step predictors along with their error bands can be obtained by means of the Kalman recursive equations.

The remainder of this article is structured as follows. The next section defines a family of LS processes, while the third section discusses a state space framework for modeling that class of nonstationary time series. The fourth and fifth sections assess the finite sample performance of the Kalman estimates and predictors, respectively. The application of the proposed procedures to tree ring data is discussed in the sixth section.

LOCALLY STATIONARY PROCESSES

Following Dahlhaus (1997), a class of Gaussian LS processes can be defined by the spectral representation

$$Y_{t,T} = \int_{-\pi}^{\pi} A^0_{t,T}(\lambda)\, e^{i\lambda t}\, \mathrm{d}B(\lambda) \tag{1}$$

for $t = 1, \ldots, T$, where $B(\lambda)$ is a Brownian motion on $[-\pi, \pi]$ and there is a positive constant K and a 2π-periodic function $A\colon (0, 1] \times \mathbb{R} \to \mathbb{C}$ with $A(u, -\lambda) = \overline{A(u, \lambda)}$ such that $\sup_{t,\lambda} |A^0_{t,T}(\lambda) - A\left(\frac{t}{T}, \lambda\right)| \leqslant \frac{K}{T}$, for all T. The transfer function $A^0_{t,T}(\lambda)$ of this class of nontstationary processes changes smoothly over time so that they can be locally approximated by stationary processes. Some examples are discussed below.

* Correspondence to: Wilfredo Palma, Department of Statistics, Pontificia Universidad Católica de Chile, Santiago, Chile. E-mail: wilfredo@mat.puc.cl

Copyright © 2011 John Wiley & Sons, Ltd

Example 1. A family of LS processes is given by the moving average expansion

$$Y_{t,T} = \sigma\left(\frac{t}{T}\right)\sum_{j=0}^{\infty}\psi_j\left(\frac{t}{T}\right)\varepsilon_{t-j} \tag{2}$$

$t = 1, \ldots, T$, where $\{\varepsilon_t\}$ is a zero-mean and unit variance Gaussian white noise and $\{\psi_j(u)\}$ are coefficients satisfying $\sum_{j=0}^{\infty}\psi_j(u)^2 < \infty$ for all $u \in [0,1]$. This model will be denoted LSMA(∞) hereafter. The time-varying spectral density of (2) is $f_\theta(u,\lambda) = \sigma^2(u)|\sum_{j=0}^{\infty}\psi_j(u)e^{i\lambda j}|^2$, for $u \in [0,1]$ and $\lambda \in [-\pi,\pi]$. Let K be a positive constant that may change from line to line. For simplicity, if $|\psi_j(u)| \leqslant K\exp(-aj)$ for $j \geqslant 1$ and $u \in [0,1]$ with a a positive constant, model (2) will be called a *short-memory process*. On the other hand, if $|\psi_j(u)| \leqslant Kj^{d-1}$ for $u \in [0,1]$ and some $d \in (0,1/2)$, model (2) will be called a *long-memory process*. Another characterization is based on the spectral density. It is said that an LS process has *short memory* if its spectral density is bounded at $\lambda = 0$ for $u \in [0,1]$. On the other hand, the process has *long memory* if its spectral density is unbounded near the origin for $u \in [0,1]$.

Example 2. A particular case of model (2) is the LSMA(∞) process with coefficients $\psi_j(u) = \phi(u)^j$. The covariance structure for this model is given by

$$\kappa_T(s,t) = \sigma\left(\frac{s}{T}\right)\sigma\left(\frac{t}{T}\right)\phi\left(\frac{s}{T}\right)^{(s-t)}\left[1-\phi\left(\frac{s}{T}\right)\phi\left(\frac{t}{T}\right)\right]^{-1}$$

for $s \geqslant t$ and its spectral density is $f_\theta(u,\lambda) = \sigma(u)^2|1-\phi(u)e^{i\lambda}|^{-2}$. As a consequence, for $|\phi(u)| < 1$, $u \in [0,1]$, $f_\theta(u,0) = \sigma(u)^2|1-\phi(u)|^{-2} < \infty$, indicating that this LSMA model has short memory. A similar conclusion holds for the LSMA(q) model defined by $Y_{t,T} = \sigma\left(\frac{t}{T}\right)\sum_{j=0}^{q}\psi_j\left(\frac{t}{T}\right)\varepsilon_{t-j}$.

Example 3. Consider the LS autoregressive process LSAR(p) defined by equation (4.1) of Dahlhaus (1997), $Y_{t,T} = \sum_{j=1}^{p}a_j(\frac{t}{T})Y_{t-j,T} + \varepsilon_t$, for $T = 1, \ldots, T$. The spectral density of the limiting process is $f_\theta(u,\lambda) = \sigma(u)^2|1 - \sum_{j=1}^{p}a_j(u)e^{i\lambda j}|^{-2}$. This process satisfies definition (1). In this case, the spectral density is bounded at the origin under some regularity conditions on the roots of the polynomial $a(B) = 1 - \sum_{j=1}^{p}a_jB^j$. Thus these LSAR($p$) processes have short memory.

Example 4. A generalization of the fractional noise model is the LS fractional noise process (LSFN) with coefficients $\psi_j(u) = \frac{\Gamma[j+d(u)]}{\Gamma(j+1)\Gamma[d(u)]}$, where $\Gamma(\cdot)$ is the Gamma function and $d(\cdot)$ is a smoothly time-varying long-memory parameter. Observe that according to Lemma A.1 of Palma (2010), the covariances of a LSFN process are

$$\kappa_T(s,t) = \sigma\left(\frac{s}{T}\right)\sigma\left(\frac{t}{T}\right)\frac{\Gamma\left[1-d\left(\frac{s}{T}\right)-d\left(\frac{t}{T}\right)\right]\Gamma\left[s-t+d\left(\frac{s}{T}\right)\right]}{\Gamma\left[1-d\left(\frac{s}{T}\right)\right]\Gamma\left[d\left(\frac{s}{T}\right)\right]\Gamma\left[s-t+1-d\left(\frac{t}{T}\right)\right]}$$

for $s,t = 1, \ldots, T$, $s \geqslant t$ and the spectral density of this process is given by

$$f_\theta(u,\lambda) = \frac{\sigma^2(u)}{2\pi}\left(2\sin\frac{\lambda}{2}\right)^{-2d_\theta(u)}$$

for $\lambda \in [-\pi,\pi]$. Thus $f_\theta(u,\lambda) \sim \frac{\sigma^2(u)}{2\pi}|\lambda|^{-2d(u)}$, for $|\lambda| \to 0$. Consequently, $f_\theta(u,\lambda)$ has a pole at the origin indicating the long-memory behavior of the LSFN process.

STATE SPACE REPRESENTATIONS

Consider the nonstationary state space system

$$\begin{aligned} X_{t+1,T} &= F_{t,T}X_{t,T} + V_{t,T}, \\ Y_{t,T} &= G_{t,T}X_{t,T} + W_{t,T} \end{aligned} \tag{3}$$

where $X_{t,T}$ is a state vector, $F_{t,T}$ is a state transition operator, V_t is a state noise with variance $Q_{t,T}$, $Y_{t,T}$ is the observation, $G_{t,T}$ is observation operator and W_t is a observation noise with variance $R_{t,T}$. The LS process (2) can be represented by a state space system (3) by generalizing the infinite-dimensional equations given on page 22 of Hannan and Deistler (1988) to the nonstationary case. The following lemma provides a specific representation of (2). Proofs of the results presented in this paper are straightforward and available from the authors.

Copyright © 2011 John Wiley & Sons, Ltd

Lemma 1. The process (2) can be represented by the following infinite-dimensional state space system:

$$\begin{aligned} X_{t+1,T} &= \begin{bmatrix} 0 \\ I_\infty \end{bmatrix} X_{t,T} + \begin{bmatrix} 1 & 0 & 0 & \cdots \end{bmatrix}' \varepsilon_{t+1}, \\ Y_{t,T} &= \sigma\left(\tfrac{t}{T}\right)\begin{bmatrix} 1 & \psi_1\left(\tfrac{t}{T}\right) & \psi_2\left(\tfrac{t}{T}\right) & \psi_3\left(\tfrac{t}{T}\right) & \cdots \end{bmatrix} X_{t,T} \end{aligned} \tag{4}$$

for $t = 1, \ldots, T$, $\mathrm{var}(X_{t,T}) = I_\infty$, where $I_\infty = \mathrm{diag}\{1, 1, \ldots\}$, $R_{t,T} = 0$, $Q_{t,T} = (q_{ij})$ with $q_{ij} = 1$ if $i = j = 1$ and $q_{ij} = 0$ otherwise.

Note that the state space representation provided by Lemma 1 may not be minimal in some cases. For instance, a one-dimensional state space representation for the LSAR(1) model studied in Example 3 is given by (3) with $Y_{t,T} = X_{t,T}$, $F_{t,T} = a\left(\frac{t+1}{T}\right)$ and $V_{t,T} = \varepsilon_t$. More generally, for LSAR(p) processes, the state space is p-dimensional:

$$X_{t+1,T} = \begin{bmatrix} a_1\left(\frac{t}{T}\right) & a_2\left(\frac{t}{T}\right) & \cdots & a_p\left(\frac{t}{T}\right) \\ 0 & & I_{p-1} & \end{bmatrix} X_{t,T} + \varepsilon_{t+1}, \quad Y_{t,T} = [1\ 0\ 0\ \cdots]\, X_{t,T}$$

where I_r denotes the $r \times r$ identity matrix hereafter. For computational efficiency it is sometimes more appropriate to consider the truncated process

$$Y_{t,T} = \sigma\left(\frac{t}{T}\right) \sum_{j=0}^{m} \psi_j\left(\frac{t}{T}\right) \varepsilon_{t-j} \tag{5}$$

for $t = 1, \ldots, T$ and some positive integer m. Naturally, (5) is an exact expansion for an LSMA(p) with $p \leqslant m$ since in that case $\psi_j(u) = 0$ for $j > p$. A finite-dimensional state state system for (5) is given by

$$\begin{aligned} X_{t+1,T} &= \begin{bmatrix} 0 & 0 \\ I_m & 0 \end{bmatrix} X_{t,T} + \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}' \varepsilon_{t+1} \\ Y_{t,T} &= \sigma\left(\frac{t}{T}\right)\left[1\ \psi_1\left(\frac{t}{T}\right)\ \psi_2\left(\frac{t}{T}\right)\ \psi_3\left(\frac{t}{T}\right)\ \cdots \psi_m\left(\frac{t}{T}\right)\right] X_{t,T} \end{aligned} \tag{6}$$

for $t = 1, \ldots, T$. The following result establishes the asymptotic magnitude of the truncation error when approximating (2) by (5).

Lemma 2. Let $r_m = \mathrm{var}\left[\sum_{j=m+1}^{\infty} \psi_j(u)\varepsilon_{t-j}\right]$ be the variance of the truncation error for approximating $\{Y_{t,T}\}$ by the finite moving average expansion (5). Then,

$$r_m \sim \begin{cases} \mathcal{O}(e^{-am}) & \text{for a short-memory process,} \\ \mathcal{O}(m^{2d-1}) & \text{for a long-memory process} \end{cases}$$

for large m, where $a > 0$ and $d = sup_u d(u) < 1/2$.

Estimation

Based on the state space representation (4) of $Y_{t,T}$, the Kalman filter equations can be used for estimating model parameters, state vectors, future observations and missing values. Let $\Delta_{t,T} = \mathrm{var}(Y_{t,T} - \widehat{Y}_{t,T})$ be the prediction error variance and let $\Omega_{t,T} = \mathrm{var}(X_{t,T} - \widehat{X}_{t,T}) = \left(\omega_{i,j}(t,T)\right)$ be the state prediction error variance–covariance matrix. The Kalman recursive equations are given by the following result.

Lemma 3. Given the initial conditions $Y_{0,T} = (0, 0, \ldots)$, $\widehat{X}_1 = \mathrm{E}(X_1) = (0, 0, \ldots)$ and $\Omega_{1,T} = \mathrm{E}(X_1 X_1') = \mathrm{diag}\{1, 1, \ldots\}$, the updating Kalman equations are

$$\begin{aligned} \Delta_{t,T} &= \sigma^2\left(\tfrac{t}{T}\right) \textstyle\sum_{i,j=1}^{\infty} \psi_{i-1}\left(\tfrac{t}{T}\right) \omega_{i,j}(t,T) \psi_{j-1}\left(\tfrac{t}{T}\right), \\ \Theta_{t,T}(i) &= \sigma\left(\tfrac{t}{T}\right) \textstyle\sum_{j=1}^{\infty} \omega_{i-1,j}(t,T) \psi_{j-1}\left(\tfrac{t}{T}\right), \\ \omega_{t+1,T}(i,j) &= \omega_{t,T}(i+1, j+1) + q_{ij} - \delta(t)\Theta_{t,T}(i)\Theta_{t,T}(j)/\Delta_{t,T}, \\ \widehat{Y}_{t,T} &= \sigma\left(\tfrac{t}{T}\right) \textstyle\sum_{j=1}^{\infty} \psi_{j-1}\left(\tfrac{t}{T}\right) \widehat{X}_{t,T}(j), \\ \widehat{X}_{t+1,T}(i) &= \widehat{X}_{t,T}(i-1) + \Theta_{t,T}(i)(Y_{t,T} - \widehat{Y}_{t,T})/\Delta_{t,T} \end{aligned} \tag{7}$$

where $\delta(t) = 1$ if observation $Y_{t,T}$ is available and $\delta(t) = 0$ otherwise.

Let θ be the model parameter, then the log-likelihood function (up to a constant) can be readily obtained from (7), $\mathcal{L}(\theta) = \sum_{t=1}^{T} \log \Delta_{t,T} + \sum_{t=1}^{T} \frac{(Y_{t,T} - \widehat{Y}_{t,T})^2}{\Delta_{t,T}}$. Hence the exact MLE provided by the Kalman equations (7) is given

Copyright © 2011 John Wiley & Sons, Ltd

by $\widehat{\theta} = \arg\max_{\theta\in\Theta} \mathcal{L}(\theta)$, where Θ is a parameter space. Note that the Kalman equations (7) can be applied directly to the general state space representation (3) or to the truncated representation (6), yielding in this case an approximate MLE.

Prediction

Let $\widehat{Y}_{n+j,T} = \mathrm{E}[Y_{n+j,T}|Y_{n,T}, Y_{n-1,T}Y_{n-2,T}, \ldots, Y_{1,T}]$ be the j-step in-sample predictor based on the finite past for $1 \leqslant n + j \leqslant T$. These forecasts are obtained from the Kalman recursive equations. Thus the best linear mean square predictors, $\widehat{Y}_{n+j,T}$ are given by

$$\widehat{Y}_{n+j,T} = G_{n+j,T}\widehat{X}_{n+j} \tag{8}$$

for $j = 1, \ldots, T-n$, where $G_{n+j,T} = [1, \psi_1(\frac{n+j}{T}), \psi_2(\frac{n+j}{T}), \ldots]$ with $\Omega_{n+j,T} = F^j\Omega_{n,T}F'^j + \sum_{k=0}^{j-1} F^k Q F'^k$ and $\Delta_{n+j,T} = G_{n+j,T}\Omega_{n+j,T}G'_{n+j,T}$, for $j = 1, \ldots, T-n$. On the other hand, out-of-sample predictors $\widehat{Y}_{T+h}$ for $h > 0$ can be obtained from the Kalman filter equations (7) by redefining the sample size $T^* = T + h$ and treating observations $T + 1, \ldots, T^*$ as missing data. Similar formulas give the predictors and their error variances for the truncated representation (6).

SIMULATION STUDIES

This section assesses the finite sample performance of the proposed state space estimation methodology, for both complete and incomplete data. All simulation results are based on 1000 replications and the locations of the missing values have been randomly selected. We start assessing the method for a short-memory LS process and then we analyze the long-memory LS case.

Short-memory case

As an illustration, consider the LSMA(∞) model described by (2) with

$$\psi_j(u) = \phi(u)^j, \qquad \phi(u) = \alpha_0 + \alpha_1 u, \qquad \sigma(u) = \beta_0 + \beta_1 u \tag{9}$$

with $|\phi(u)| < 1$ and $\sigma(u) > 0$ for $u \in [0, 1]$. Denote the parameter vector by $\boldsymbol{\alpha} = (\alpha_0, \alpha_1)$ for $\phi(\cdot)$ and $\boldsymbol{\beta} = (\beta_0, \beta_1)$ for the noise scale $\sigma(\cdot)$. The simulated processes are generated by using the innovation algorithm (see Brockwell and Davis, 1991). The optimal theoretical standard deviations (SD) are based on the formulas given by Theorem 3.1 of Dahlhaus (2000) for the MLE of short-memory LS models:

$$\Gamma(\boldsymbol{\alpha})_{ij} = \int_0^1 \frac{u^{i+j-2}}{1-[\phi(u)]^2}\mathrm{d}u, \qquad \Gamma(\boldsymbol{\beta})_{ij} = 2\int_0^1 \frac{u^{i+j-2}}{\sigma^2(u)}\mathrm{d}u \tag{10}$$

for $i, j = 1, 2$. Furthermore, $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\beta}}$ are asymptotically independent. Table I displays the simulation results from the Kalman method for two truncation levels, $m = 40, 80$. The number of missing values are 10% and 20%, which

Table I. Estimation of model (9) with $(\alpha_0, \alpha_1, \beta_0, \beta_1) = (-0.3, 0.8, 0.5, 0.5)$

	$m = 40$			$m = 80$		
% NA	0%	10%	20%	0%	10%	20%
$\widehat{\alpha}_0$	−0.301	−0.303	−0.299	−0.304	−0.303	−0.300
$\widehat{\alpha}_1$	0.802	0.803	0.794	0.804	0.804	0.793
$\widehat{\beta}_0$	0.506	0.481	0.453	0.506	0.481	0.454
$\widehat{\beta}_1$	0.488	0.460	0.437	0.488	0.460	0.433
$\sigma(\widehat{\alpha}_0)$	0.061	0.064	0.068	0.061	0.064	0.068
$\sigma(\widehat{\alpha}_1)$	0.102	0.108	0.114	0.102	0.108	0.114
$\sigma(\widehat{\beta}_0)$	0.027	0.028	0.030	0.027	0.028	0.030
$\sigma(\widehat{\beta}_1)$	0.056	0.059	0.063	0.056	0.059	0.063
$\widehat{\sigma}(\widehat{\alpha}_0)$	0.062	0.069	0.075	0.061	0.069	0.075
$\widehat{\sigma}(\widehat{\alpha}_1)$	0.106	0.117	0.124	0.103	0.114	0.127
$\widehat{\sigma}(\widehat{\beta}_0)$	0.028	0.029	0.030	0.026	0.029	0.031
$\widehat{\sigma}(\widehat{\beta}_1)$	0.058	0.059	0.061	0.056	0.060	0.062

Copyright © 2011 John Wiley & Sons, Ltd

have been randomly selected for each simulation. From this table, note that the estimated parameters and SD are close to their theoretical counterparts, especially as the truncation level m increases. As expected, the precision of the estimates deteriorates as the percentage of missing data increases.

Long-memory case

Consider the LSFN defined in Example 4 with long-memory and scale parameters given by

$$d(u) = \alpha_0 + \alpha_1 f(u), \qquad \sigma(u) = \beta_0 + \beta_1 u \tag{11}$$

respectively, where f is a known function, $0 < d(u) < 1/2$ and $\sigma(u) > 0$ for $u \in [0, 1]$. We consider three models, with combinations of linear and harmonic specifications of f, i.e. $f(u) = u$ and $f(u) = \cos(\omega u)$, respectively. Tables II–IV show the results from these simulations. The optimal SDs for the linear case $f(u) = u$ are based on the matrix Γ provided by Theorem 2.2 of Palma and Olea (2010), $\Gamma(\boldsymbol{\alpha})_{ij} = \pi^2[6(i + j - 1)]^{-1}$, $i, j = 1, 2$, and $\Gamma(\boldsymbol{\beta})$ is given by (10). On the other hand, for the harmonic case $f(u) = \cos(\omega u)$ the optimal SD of $\widehat{\boldsymbol{\alpha}}$ are based on the following matrix given by Palma and Olea (2010):

$$\Gamma(\boldsymbol{\alpha})_{ij} = \frac{\pi^2}{12}\left[\frac{\sin(\lambda_i - \lambda_j)}{\lambda_i - \lambda_j} + \frac{\sin(\lambda_i + \lambda_j)}{\lambda_i + \lambda_j}\right], \qquad i, j = 1, 2 \tag{12}$$

where $(\lambda_1, \lambda_2) = (0, 2\pi)$ and the SDs of $\widehat{\boldsymbol{\beta}}$ are based on the matrix $\Gamma(\boldsymbol{\beta})$ given by (10).

Table II. Estimation of model (11) with $(\alpha_0, \alpha_1, \beta_0, \beta_1, f) = (0.2, 0.25, 0.5, 0.5, u)$

	$m = 40$			$m = 80$		
% NA	0%	10%	20%	0%	10%	20%
$\widehat{\alpha}_0$	0.209	0.213	0.213	0.204	0.208	0.209
$\widehat{\alpha}_1$	0.248	0.240	0.243	0.249	0.244	0.243
$\widehat{\beta}_0$	0.508	0.482	0.456	0.507	0.482	0.455
$\widehat{\beta}_1$	0.489	0.465	0.439	0.490	0.464	0.438
$\sigma(\widehat{\alpha}_0)$	0.049	0.051	0.055	0.049	0.051	0.055
$\sigma(\widehat{\alpha}_1)$	0.084	0.089	0.094	0.084	0.089	0.094
$\sigma(\widehat{\beta}_0)$	0.027	0.028	0.030	0.027	0.028	0.030
$\sigma(\widehat{\beta}_1)$	0.056	0.059	0.063	0.056	0.059	0.063
$\widehat{\sigma}(\widehat{\alpha}_0)$	0.043	0.045	0.048	0.045	0.049	0.049
$\widehat{\sigma}(\widehat{\alpha}_1)$	0.066	0.072	0.075	0.071	0.074	0.075
$\widehat{\sigma}(\widehat{\beta}_0)$	0.026	0.027	0.028	0.025	0.028	0.028
$\widehat{\sigma}(\widehat{\beta}_1)$	0.053	0.059	0.059	0.051	0.056	0.056

Table III. Estimation of model (11) with $(\alpha_0, \alpha_1, \beta_0, \beta_1, f) = (0.35, 0.1, 0, 1, \cos(2\pi u))$

	$m = 40$			$m = 80$		
% NA	0%	10%	20%	0%	10%	20%
$\widehat{\alpha}_0$	0.357	0.360	0.345	0.354	0.357	0.347
$\widehat{\alpha}_1$	0.106	0.105	0.120	0.103	0.104	0.117
$\widehat{\beta}_0$	0.953	0.844	0.902	0.966	0.859	0.911
$\widehat{\beta}_1$	0.037	0.041	0.039	0.036	0.029	0.037
$\sigma(\widehat{\alpha}_0)$	0.024	0.026	0.027	0.024	0.026	0.027
$\sigma(\widehat{\alpha}_1)$	0.034	0.036	0.039	0.034	0.036	0.039
$\sigma(\widehat{\beta}_0)$	0.044	0.047	0.049	0.044	0.047	0.049
$\sigma(\widehat{\beta}_1)$	0.077	0.081	0.086	0.077	0.081	0.086
$\widehat{\sigma}(\widehat{\alpha}_0)$	0.024	0.033	0.034	0.024	0.028	0.029
$\widehat{\sigma}(\widehat{\alpha}_1)$	0.027	0.035	0.033	0.032	0.035	0.037
$\widehat{\sigma}(\widehat{\beta}_0)$	0.053	0.066	0.093	0.046	0.052	0.054
$\widehat{\sigma}(\widehat{\beta}_1)$	0.081	0.092	0.092	0.079	0.083	0.091

Copyright © 2011 John Wiley & Sons, Ltd

Table IV. Estimation of model (11), $(\alpha_0, \alpha_1, \beta_0, \beta_1, f) = (0.25, 0.2, 1.5, 1-0, 5, \cos(2\pi u))$

	$m = 40$			$m = 80$		
% NA	0%	10%	20%	0%	10%	20%
$\widehat{\alpha}_0$	0.253	0.254	0.253	0.252	0.251	0.251
$\widehat{\alpha}_1$	0.207	0.207	0.209	0.203	0.207	0.210
$\widehat{\beta}_0$	1.427	1.359	1.290	1.457	1.449	1.303
$\widehat{\beta}_1$	−0.493	−0.394	−0.379	−0.468	−0.452	−0.389
$\sigma(\widehat{\alpha}_0)$	0.024	0.026	0.027	0.024	0.026	0.027
$\sigma(\widehat{\alpha}_1)$	0.034	0.036	0.039	0.034	0.036	0.039
$\sigma(\widehat{\beta}_0)$	0.060	0.063	0.067	0.060	0.063	0.067
$\sigma(\widehat{\beta}_1)$	0.095	0.100	0.106	0.095	0.100	0.106
$\widehat{\sigma}(\widehat{\alpha}_0)$	0.025	0.028	0.029	0.025	0.027	0.030
$\widehat{\sigma}(\widehat{\alpha}_1)$	0.033	0.034	0.037	0.033	0.035	0.038
$\widehat{\sigma}(\widehat{\beta}_0)$	0.071	0.073	0.081	0.063	0.065	0.071
$\widehat{\sigma}(\widehat{\beta}_1)$	0.099	0.121	0.135	0.097	0.101	0.118

From Tables II–IV, observe that the estimates are close to their true values for the two truncation levels m. On the other hand, the empirical SDs get closer the their optimal theoretical counterparts as m increases. In particular, for $m = 80$ the empirical and the theoretical SDs are very similar for both complete and incomplete data.

SIMULATION STUDIES: PREDICTION

The calculation and finite sample performance of one-step and multi-step predictors of short- and long-memory LS processes is illustrated in this section.

Short-memory case

Consider the time-varying moving average LSMA process discussed in Example 2. We calculate the best linear predictor, $\widehat{Y}_{974+j,T}$ for $j = 1, \ldots, 50$, given by (8) with sample size $T = 1,024$ and time-varying $\phi(u)$ parameter and scale factor $\sigma(u)$ given by specification (9) with $(\alpha_0, \alpha_1, \beta_0, \beta_1) = (-0.4, 0.8, 0.5, 0.5)$. The model is estimated using truncation $m = 80$ with observations $t = 1, \ldots, 974$ and values $t = 975, \ldots, 1024$ are left for out-of-sample forecasting. Figure 1 shows the data (light line for the first 974 observations and dotted line for the last 50 values) and predictors (heavy line), along 95% prediction bands for the out-of-sample forecasts.

The evolution of the prediction error SD is depicted in Figure 2. The broken line represents the SD of the process $\sigma(Y_{t,T}) = \kappa(t,t)^{\frac{1}{2}} = \sigma(t/T)[1-\phi(t/T)^2]^{-\frac{1}{2}}$ with $\phi(u) = 0.8u - 0.4$, the dotted line corresponds to the noise SD, $\sigma(t/T) = 0.5(1+u)$, and the heavy line denotes the estimated SD of the prediction error $\Delta_{t,T}^{\frac{1}{2}}$. From this figure, notice that $\sigma(Y_{512,1024}) = \sigma(0.5)$ and that $\Delta_{t,T}^{\frac{1}{2}}$ gets closer to its upper limit $\sigma(Y_{t,T})$ for $t > 974$ since the last 50 observations are missing, consequently no new information is available to improve the forecasts. Observe that, by definition, $\sigma^2(t/T) \leqslant \Delta_{t,T} \leqslant \sigma^2(Y_{t,T})$ for $t = 1, \ldots, T$.

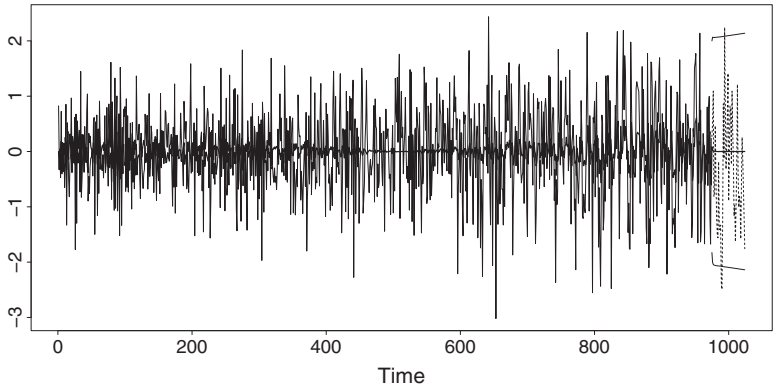

Figure 1. Simulated LSMA process: multi-step out-of-sample forecasts of the last 50 observations and 95% prediction bands

Copyright © 2011 John Wiley & Sons, Ltd

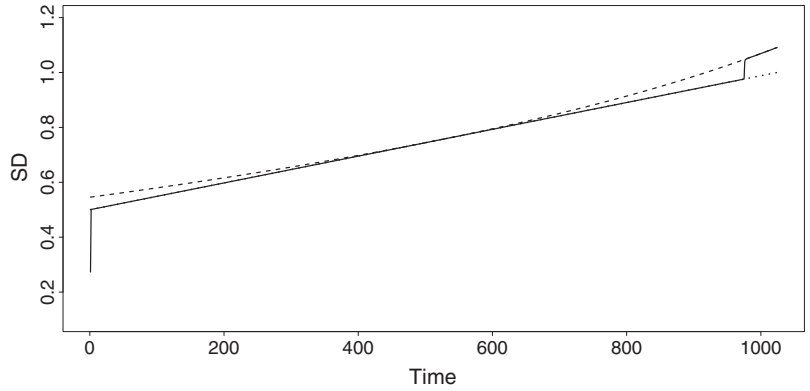

Figure 2. Prediction of a LSMA model. Broken line, standard deviation of $Y_{t,T}$; dotted line, SD of the noise $\sigma(t/T)$; heavy line, empirical prediction error SD, $\Delta_{t,T}^{\frac{1}{2}}$

Long-memory case

We illustrate this case by considering the LSFN model with time-varying long-memory parameter $d(u)$ and scale factor $\sigma(u)$ given by specification (11) with $\alpha_0 = 0.15$, $\alpha_1 = 0.25$, $f(u) = u$, $\beta_0 = 1$, $\beta_1 = -0.7$ and sample size $T = 1024$. As in the previous case, the Kalman recursive equations allow for the prediction of future values $\widehat{Y}_{n+j,T}$ with $j = 1, \ldots, 50$ and $n = 974$, based on the model estimated with observations $Y_1, \ldots, Y_{974}$ and truncation $m = 80$.

Figure 3 displays a simulated LSFN process with model parameters specified above. The series is represented by the light line for $t = 1, \ldots, 974$ and by the dotted line for the last 50 values. The heavy line represents the one-step predictors for $t = 1, \ldots, 974$ and j-step predictors $\widehat{Y}_{974+j}$ for $j = 1, \ldots, 50$. In addition, 95% prediction bands are shown for the last 50 values. Furthermore, Figure 4 shows the evolution of the empirical prediction error SD, $\Delta_{t,T}^{1/2}$,

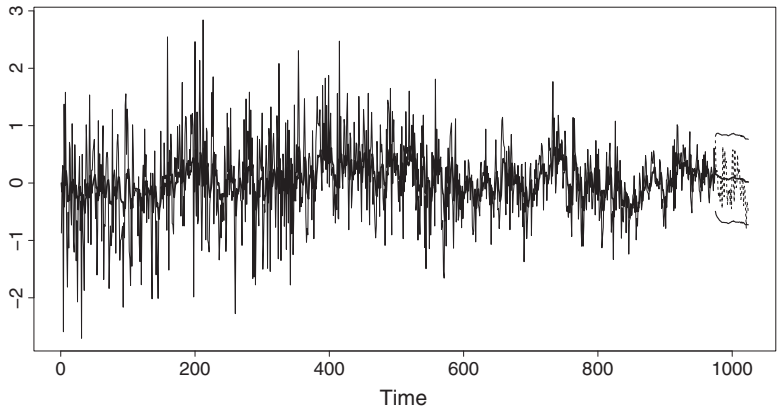

Figure 3. Simulated LSFN process: multi-step out-of-sample forecasts of the last 50 observations predicted and 95% prediction bands

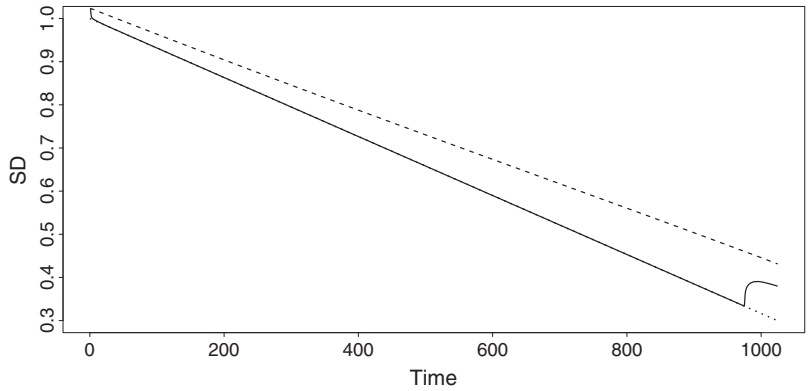

Figure 4. Prediction of an LSFN model. Broken line, SD of $Y_{t,T}$; dotted line, SD of the noise $\sigma(u)$; heavy line, empirical prediction error SD, $\Delta_{t,T}^{\frac{1}{2}}$

Copyright © 2011 John Wiley & Sons, Ltd

represented by the heavy line. These SD are based on the multi-step prediction error variance discussed above for a truncated state space representation with $m = 80$. The broken line represents the evolution of the theoretical SD of $Y_{t,T}$, $\sigma(Y_{t,T})$. Note that the broken line is an upper limit for the prediction error SD. In addition, the dotted line corresponds to the theoretical noise SD, $\sigma(t/T)$. Observe the increase of $\Delta_{t,T}^{1/2}$ at the end of the series due to the missing data.

Estimation of missing values can be readily carried out by means of the Kalman methodology. As an illustration, consider a simulated LSFN process of length $T = 512$ with two blocks of 50 observations each, removed from the sample. These blocks are located at $t = 201, \ldots, 250$ and $t = 401, \ldots, 450$, respectively. The model is described by (11) with $\alpha_0 = 0.15, \alpha_1 = 0.25$, $f(u) = u$, $\beta_0 = 0$ and $\beta_1 = 1$. Figure 5 shows the observations for $t = 1, \ldots, 512$, along with the forecasts $\widehat{Y}_{200+j}$ $\widehat{Y}_{400+j}$ for $j = 1, \ldots, 50$ and their respective 95% prediction bands. These multi-step-ahead predictors (heavy line) are based on the fitted model with a truncated state space representation of $m = 80$. The true values of the series are depicted by the light line for the available data and by the dotted line for the data

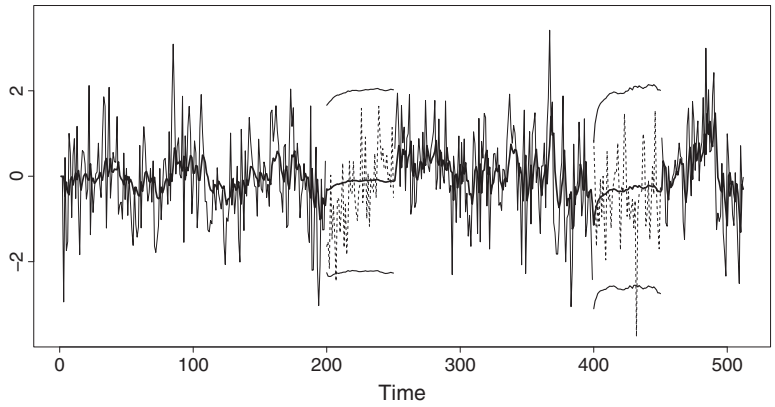

Figure 5. Prediction of an LSFN model: multi-step forecasts of two blocks of 50 missing values and 95% prediction bands

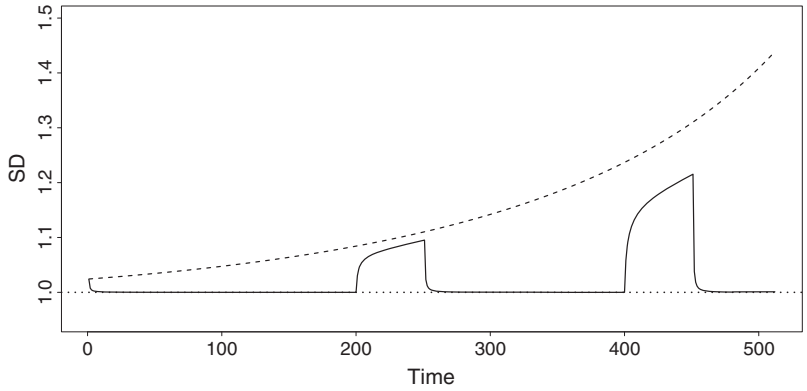

Figure 6. Prediction of LSFN model. Broken line, SD of $Y_{t,T}$; dotted line, SD of the noise $\sigma(u)$; heavy line, empirical prediction error SD, $\Delta_{t,T}^{\frac{1}{2}}$

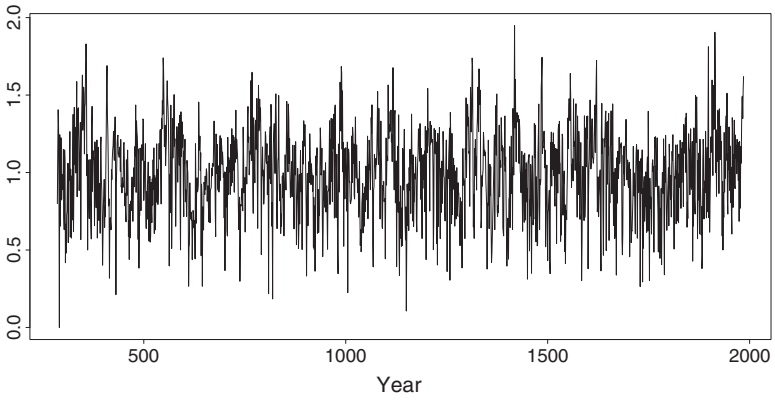

Figure 7. Tree ring data

Copyright © 2011 John Wiley & Sons, Ltd

gaps. In Figure 6, the broken line represents the evolution of $\sigma(Y_{t,T})$. The heavy line corresponds to the empirical error prediction SD, $\Delta_{t,T}^{1/2}$, in the presence of data gaps for $t = 201 - 250$ and $t = 401 - 450$, while the dotted line represents the noise SD, $\sigma(u) = 1$. Notice that the prediction error SD increases right after the beginning of each data gap and it decays to $\sigma(u) = 1$ as new observations become available.

DATA APPLICATION

This section discusses the application of the Kalman estimation and prediction techniques developed above to the analysis of a time series consisting of pine tree ring measurements at Nevada, from AD 286 to AD 1985. The data, reported by D. A. Graybill and available at the National Climatic Data Center, are displayed in Figure 7. This type of time series is often used as proxies of temperature and precipitation conditions in paleoclimatology (cf. Tan *et al*, 2003).

The sample autocorrelation function (ACF) of the data, displayed in Figure 8(a), shows significant autocorrelations at large lags. The corresponding *variance plot* is shown in Figure 8(b). In a variance plot, the broken line represents the expected behavior of the variance of the sample mean of a block of k observations for a short-memory process. On the other hand, the heavy line represents the expected behavior of the variance for a long-memory process (cf. Section 4.5.3 of Palma, 2007). From both panels, this series seems to exhibit long-range dependence.

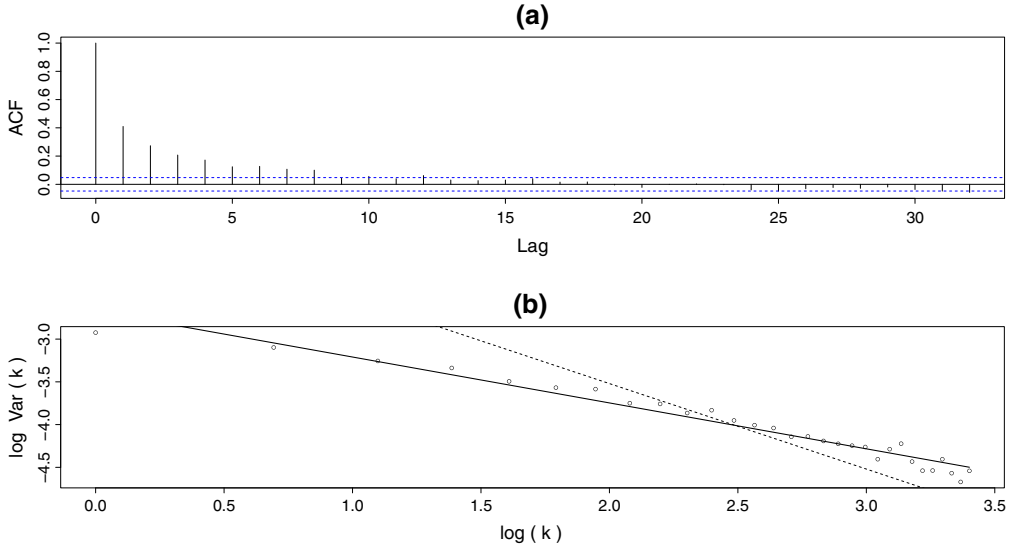

Figure 8. Tree ring data: (a) sample ACF; (b) variance plot

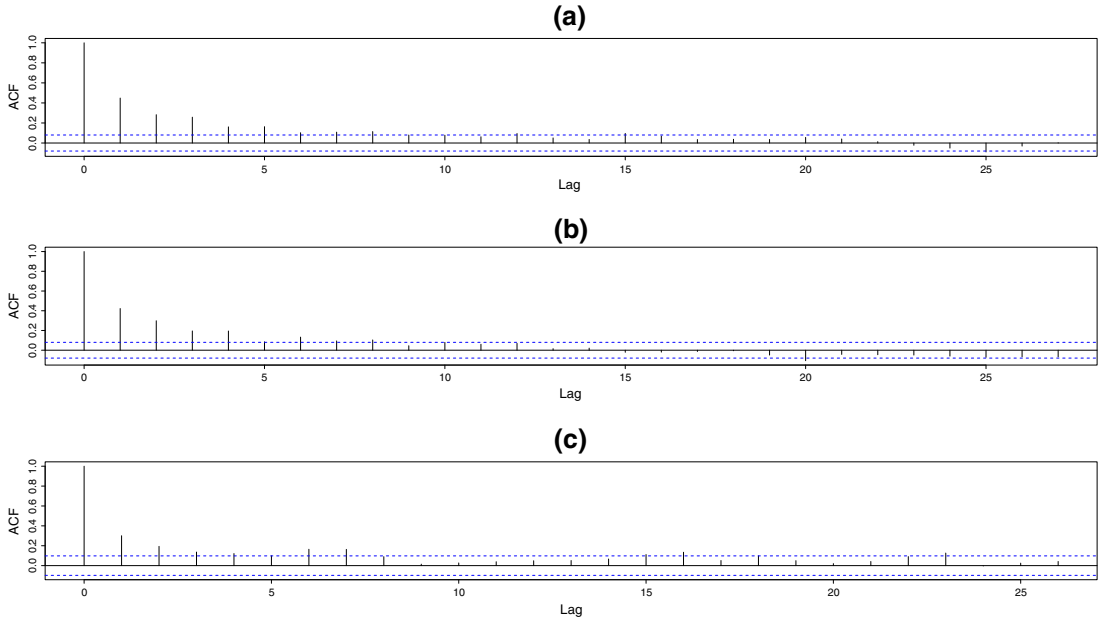

Figure 9. Tree ring data. Sample ACF: (a) observations 1–400;(b) observations 651–1250; (c) observations 1301–1700

Copyright © 2011 John Wiley & Sons, Ltd

Nevertheless, a closer look at the empirical ACF of the data reveals that the degree of persistence seems to vary over time; see Figure 9, where the full sample is divided into three segments: observations 1–400, 651–1050 and 1301–1700, respectively. Figure 9(a–c) shows the empirical ACF for these three periods, providing evidence of possible changes in the degree of dependence.

To gain some insight about the shape of the time-varying long-memory function $d(\cdot)$ and the noise SD $\sigma(\cdot)$, the series is divided into several windows. Within each window, the long-memory parameter and scale are estimated with a stationary FN(d) process. In this case, each window has 500 observations with shift 120. For instance, the first window is $Y_{1,T}, \ldots, Y_{501,T}$, the second is $Y_{121,T}, \ldots, Y_{620,T}$ and so on.

Figure 10 shows these heuristic estimates represented by dots. Since these rough estimates suggest a linear behavior of $d(u)$ and $\sigma(u)$, a model specified by (11) with $f(u) = u$ is considered. The state space estimates are also plotted in Figure 10. In this figure, the dotted line represents the parameter estimates of d and σ for the stationary case, while

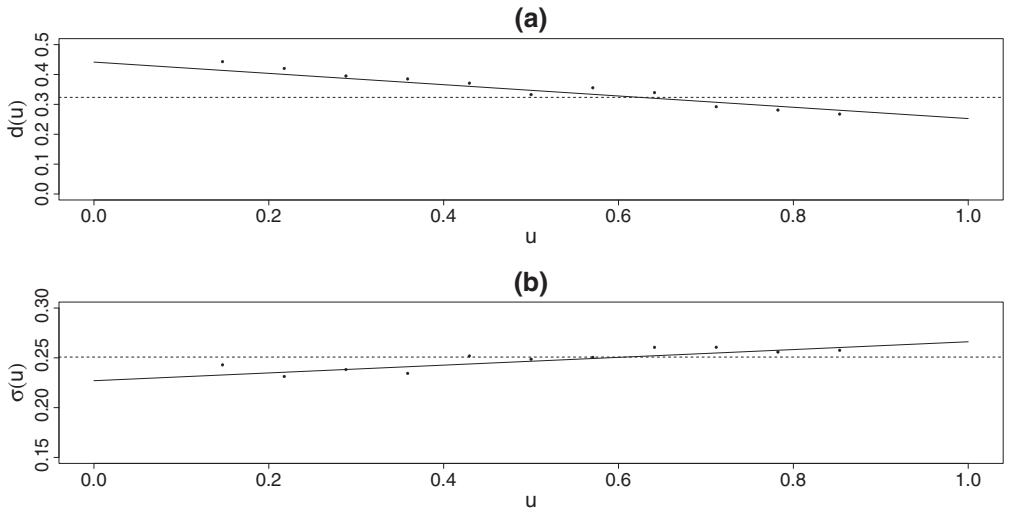

Figure 10. State space estimation of $d(u)$ and $\sigma(u)$ (heavy line), heuristic estimates (dots) and estimates of d and σ under a stationary fractional noise model (broken line)

Table V. Parameters estimated with a truncated $m = 50$

Parameter	Estimate	SD	t-value
α_0	0.4900	0.0378	12.9630
α_1	−0.2742	0.0655	−4.1862
β_0	0.2123	0.0077	27.5714
β_1	0.0566	0.0143	3.9580

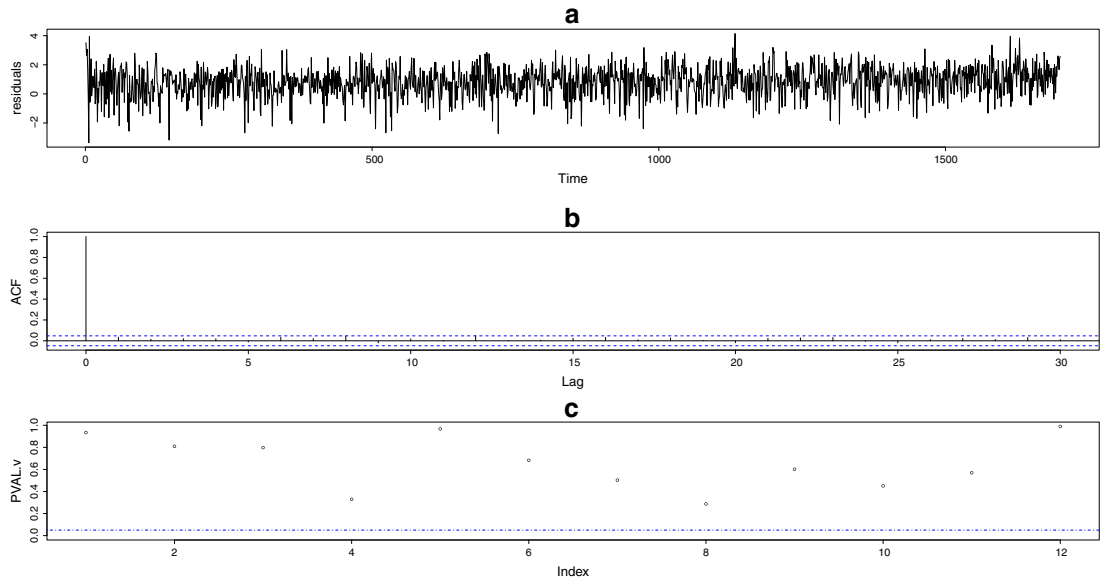

Figure 11. Tree ring data: residual analysis: (a) standardized residuals from the fitted model; (b) sample ACF; (c) Ljung–Box tests

Copyright © 2011 John Wiley & Sons, Ltd

the heavy line indicates the time-varying Kalman estimates of $d(u)$ and $\sigma(u)$, using truncation $m = 80$, and the lineal specification (11).

Table V reports the Kalman estimates. Note that, according to the fourth column of this table, the parameters are statistically significant at the 5% level. Higher-order polynomials were also fitted to $d(u)$ and $\sigma(u)$ but their coefficients were not statistically significant at that level.

A nice feature of the Kalman recursions (7) is that they directly provide the residuals of the model, $e_t = Y_{t,T} - \widehat{Y}_{t,T}$, along with their SD, $\Delta_{t,T}^{\frac{1}{2}}$. Figure 11 shows three panels exploring the structure of the standardized residuals $r_t = e_t \Delta_{t,T}^{-\frac{1}{2}}$. Panel (a) displays the standardized residuals from the fitted LSFN model. Panel (b) shows the sample ACF, and panel (c) exhibits the Ljung–Box whiteness tests. From panel (b), it seems that there are no significant autocorrelations in the residuals. This conclusion is formally supported by the Ljung–Box tests, considering up to $K = 12$ (see panel (c)). This graph indicates that the white noise hypothesis is not rejected at the 5% level of significance.

ACKNOWLEDGEMENT

This research was partially supported by Fondecyt Grant 1085239.

REFERENCES

Brockwell PJ, Davis RA. 1991. *Time Series: Theory and Methods*. Springer: New York.

Cavanaugh JE, Wang Y, Davis JW. 2003. Locally self-similar processes and their wavelet analysis. In *Stochastic Processes: Modelling and Simulation*, vol. 21, Shanbhag DN, Rao CR (eds), Handbook of Statistics. North-Holland: Amsterdam; 93–135.

Dahlhaus R. 1997. Fitting time series models to nonstationary processes. *Annals of Statistics* **25**: 1–37.

Dahlhaus R. 2000. A likelihood approximation for locally stationary processes. *Annals of Statistics* **28**: 1762–1794.

Dahlhaus R, Polonik W. 2006. Nonparametric quasi-maximum likelihood estimation for Gaussian locally stationary processes. *Annals of Statistics* **34**: 2790–2824.

Dahlhaus R, Polonik W. 2009. Empirical spectral processes for locally stationary time series. *Bernoulli* **15**: 1–39.

Hannan EJ, Deistler M. 1988. *The Statistical Theory of Linear Systems*. Wiley: New York.

Jensen MJ, Witcher B. 2000. Time-varying long memory in volatility: detection and estimation with wavelets. , EURANDOM, Eindhoven.

Last M, Shumway R. 2008. Detecting abrupt changes in a piecewise locally stationary time series. *Journal of Multivariate Analysis* **99**: 191–214.

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

Palma W. 2010. On the sample mean of locally stationary long-memory processes. *Journal of Statistical Planning and Inference* **140**: 3764–3774.

Palma W, Olea R. 2010. An efficient estimator for locally stationary Gaussian long-memory processes. *Annals of Statistics* **38**: 2958–2997.

Priestley MB. 1965. Evolutionary spectra and non-stationary processes. *Journal of the Royal Statistical Society, Series B* **27**: 204–237.

Shumway RH, Stoffer DS. 2010. *Time Series Analysis and Its Applications: With R Examples* (3rd edn). Springer: New York.

Tan M, Liu TS, Hou J, Qin X, Zhang H, Li T. 2003. Cyclic rapid warming on centennial-scale revealed by a 2650-year stalagmite record of warm season temperature. *Geophysical Research Letters* **30**: 191–194.

Wang Y, Cavanaugh JE, Song C. 2001. Self-similarity index estimation via wavelets for locally self-similar processes. *Journal of Statistical Planning and Inference* **99**: 91–110.

Authors' biographies:

Wilfredo Palma is a Professor of Statistics at Pontificia Universidad Catolica de Chile. His research interests include time series analysis and forecasting.

Ricardo Olea is an Assistant Professor at the Department of Statistics at Pontificia Universidad Catolica de Chile. He works on time series methods for stationary and nonstationary data.

Guillermo Ferreira received his Ph.D. degree in Statistics from the Pontificia Universidad Catolica de Chile. Since 2009 he is an Assistant Professor of Statistics at Universidad de Concepcion. His research interests are time series analysis and financial econometrics.

Authors' addresses:

Wilfredo Palma and **Ricardo Olea** Department of Statistics Pontificia Universidad Catt'olica de Chile, Santiago 22, Chile.

Guillermo Ferreira Department of Statistics, Universidad de Concepción, Concepción, Chile.

Copyright © 2011 John Wiley & Sons, Ltd