Chilean Journal of Statistics
Vol. 4, No. 2, September 2013, 133–149

TIME SERIES
REVIEW PAPER

Statistical analysis of locally stationary processes

Guillermo Ferreira [1], Ricardo Olea [2], and Wilfredo Palma[2,*]

[1]Department of Statistics , Universidad de Concepción, Concepción, Chile,
[2]Department of Statistics, Pontificia Universidad Católica de Chile , Santiago , Chile,

(Received: 01 March Month 2013 · Accepted in final form: 30 August 2013)

Abstract

This paper provides an overview of locally stationary processes, a helpful methodology for handling nonstationary time series. These techniques allow for the smooth evolution of the model parameters. This work reviews estimation and predictions techniques, illustrating the application of these methods to real-life data examples. These examples show that the locally stationary methods provide a useful theoretical and practical framework for the statistical analysis of nonstationary time series data.

Keywords: Kalman filter · State space system · Nonstationarity · Long-range dependence · Local stationarity · Time-varying models.

Mathematics Subject Classification: Primary 62M10 · Secondary 60G15.

1. Introduction

During the last decades, locally stationary (LS) processes have been playing an important role in time series analysis. They have provided a sound statistical methodology for modeling data exhibiting nonstationary features without resorting to data transformations, trend removals and other related techniques. The theory of LS processes is based on the principle that a nonstationary process can be locally approximated by a stationary one if the time variation of the model parameters is sufficiently smooth. The idea of developing techniques for handling directly nonstationary processes dates back to the sixties. For example, Priestley (1965), Priestley and Tong (1973), Tong (1973) and others developed the concept of evolutionary spectra. In the ninities, Dahlhaus (1996a, 1997) provided a formal definition of a family of LS processes which has generated a pleyade of related works, see for example Dahlhaus (2000), Jensen (2000), Dahlhaus (2006, 2009), Chandler and Polonik (2006), Palma and Olea (2010) and Palma et al. (2013), among others. Other classes of LS processes have been discussed for example by Wang et al. (1973), Cavanaugh et al. (2003) and Last and Shumway (2008).

In this paper we provide a brief overview of this important field of research in time

*Corresponding author. Email: wilfredo@mat.puc.cl

ISSN: 0718-7912 (print)/ISSN: 0718-7920 (online)

http://www.soche.cl/chjs

series analysis, discussing estimation and prediction techniques as well as applications to real-life data. In the context of parameter estimation, we discuss the calculation of exact maximum likelihood estimators (MLE) by means of state space systems and the computation of approximate MLE by means of an adaptation of the Whittle approach to the nonstationary case. Additionally, we address the calculation of one-step and multi-step ahead predictors, based on finite or infinity past.

The remaining of this article is structured as follows. A class of LS processes is defined in Section 2 following Dahlhaus (1997). Section 3 discusses a state space framework for modeling that class of LS processes models. Section 4 is devoted to a revision of estimation methods including state space and Whittle techniques. Section 5 addresses methods for obtaining both one-step and multi-step ahead forecats for LS processes . Conclusions are provided in Section 6.

2. Locally Stationary Processes

Recall that a Gaussian stationary $\{Y_t\}$ can be written in terms of a Cramer spectral representation,

$$Y_t = \int_{-\pi}^{\pi} A(\lambda)\, e^{i\lambda t}\, dB(\lambda), \tag{1}$$

where $A(\lambda)$ is a transfer function and $B(\lambda)$ is a Brownian motion on $[-\pi, \pi]$. In a series of seminal papers, see for example Dahlhaus (1997), this expression is extended allowing the transfer function to evolve in time as follows,

$$Y_{t,T} = \int_{-\pi}^{\pi} A^0_{t,T}(\lambda)\, e^{i\lambda t}\, dB(\lambda), \tag{2}$$

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 $A : (0, 1] \times \mathbb{R} \to \mathbb{C}$ with $A(u, -\lambda) = \overline{A(u, \lambda)}$ such that

$$\sup_{t,\lambda} |A^0_{t,T}(\lambda) - A\left(\tfrac{t}{T}, \lambda\right)| \leq \tfrac{K}{T}, \tag{3}$$

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.

Example 2.1 Consider the following time-varying version of the first order moving average process, denoted for simplicity as LSMA(1),

$$Y_{t,T} = \sigma\left(\tfrac{t}{T}\right)\left[1 + \theta\left(\tfrac{t}{T}\right)\varepsilon_{t-1}\right], \tag{4}$$

$t = 1, \ldots, T$, where $\{\varepsilon_t\}$ is a zero-mean and unit variance white noise sequence. The covariance structure of this model is,

$$\kappa_T(s,t) = \begin{cases} \sigma^2\left(\tfrac{t}{T}\right)\left[1 + \theta^2\left(\tfrac{t}{T}\right)\right], & s = t, \\ \sigma\left(\tfrac{t}{T}\right)\sigma\left(\tfrac{t-1}{T}\right)\theta\left(\tfrac{t}{T}\right), & s = t-1, \\ \sigma\left(\tfrac{t}{T}\right)\sigma\left(\tfrac{t+1}{T}\right)\theta\left(\tfrac{t+1}{T}\right), & s = t+1, \\ 0 & \text{otherwise.} \end{cases}$$

In this case, the transfer function of the process is given by

$$A^0_{t,T}(\lambda) = A\left(\tfrac{t}{T}, \lambda\right) = \sigma\left(\tfrac{t}{T}\right)\left[1 + \theta\left(\tfrac{t}{T}\right) e^{i\lambda}\right]. \tag{5}$$

Furthermore, the time-varying spectral density is

$$f\left(\tfrac{t}{T}, \lambda\right) = |A\left(\tfrac{t}{T}, \lambda\right)|^2 = \sigma^2\left(\tfrac{t}{T}\right)\left[1 + \theta^2\left(\tfrac{t}{T}\right) + 2\theta\left(\tfrac{t}{T}\right)\cos\lambda\right]. \tag{6}$$

Example 2.2 An extension of the previous model is the time-varying MA(∞) moving average expansion

$$Y_{t,T} = \sigma\left(\tfrac{t}{T}\right)\sum_{j=0}^{\infty}\psi_j\left(\tfrac{t}{T}\right)\varepsilon_{t-j}, \tag{7}$$

$t = 1, \ldots, T$, where $\{\varepsilon_t\}$ is a zero-mean and unit variance Gaussian white noise and $\{\psi_j(u)\}$ are coefficients satisfying $\psi_0(u) = 1$ and $\sum_{j=0}^{\infty}\psi_j(u)^2 < \infty$ for all $u \in [0,1]$. This model will be denoted LSMA(∞) hereafter. The time-varying spectral density of (7) 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 $\lambda \in [-\pi, \pi]$. For simplicity, if $|\psi_j(u)| \leq K\exp(-aj)$ for $j \geq 1$ and $u \in [0,1]$ with K and a positive constants, model (7) will be called a *short-memory process*. On the other hand, if $|\psi_j(u)| \leq Kj^{d-1}$ for $u \in [0,1]$ and some $d \in (0, 1/2)$, model (7) will be called a *long-memory process*. Another characterization is based on the spectral density. It is said that a 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 2.3 Consider the LS autoregressive process LSAR(1) defined as

$$Y_{t,T} = \phi(\tfrac{t}{T})Y_{t-1,T} + \varepsilon_t, \tag{8}$$

for $T = 1, \ldots, T$. Suppose that $\phi(u) = \phi(0)$ for $u < 0$, and there exists a positive constant $K < 1$ such that $|\phi(u)| \leq K$ for $u < 1$. Thus, an expanded Wold expansion of this process is given by,

$$Y_{t,T} = \sum_{j=0}^{\infty}\psi_j(t,T)\,\varepsilon_{t-j}, \tag{9}$$

where $\psi_0(t,T) = 1$ for all t,T, and for $j \geq 1$,

$$\psi_j(t,T) = \prod_{k=0}^{j-1}\phi(\tfrac{t-k}{T}). \tag{10}$$

From this, we conclude that the transfer function can be written as

$$A^0_{t,T}(\lambda) = 1 + \sum_{j=1}^{\infty}\prod_{k=0}^{j-1}\phi(\tfrac{t-k}{T})e^{i\lambda j}. \tag{11}$$

Note that by Theorem 2.3 of Dahlhaus (1996b), this transfer function satisfies the condition (3). The spectral density of the limiting process is $f_\theta(u,\lambda) = \sigma(u)^2|1 - \phi(u)e^{i\lambda j}|^{-2}$. This

process satisfies definition (2) and its spectral density is bounded at the origin for all u. Thus, this is a short-memory process.

Example 2.4 Observe that a stationary fractional noise process (FN) with long-memory parameter d is given by

$$Y_t = \sigma \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \tag{12}$$

where $\psi_j = \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)}$, where $\Gamma(\cdot)$ is the Gamma function. A nonstationary extension of this 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 $d(\cdot)$ 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

$$\kappa_T(s,t) = \sigma\left(\tfrac{s}{T}\right)\sigma\left(\tfrac{t}{T}\right)\frac{\Gamma\left[1-d\left(\frac{s}{T}\right)-d\left(\frac{t}{T}\right)\right]\Gamma\left[s-t+d\left(\frac{s}{T}\right)\right]}{\Gamma\left[1-d\left(\frac{s}{T}\right)\right]\Gamma\left[d\left(\frac{s}{T}\right)\right]\Gamma\left[s-t+1-d\left(\frac{t}{T}\right)\right]}, \tag{13}$$

for $s,t = 1,\ldots,T$, $s \geq t$. From this expression, and for large $s-t$ we have that

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

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 and then this is a long-memory process.

3. State-Space Representations

Given that the state space (SS) systems provide a very useful framework for the efficient calculation of estimates and forecasts, in this section we review the application of this representations to the case of locally stationary processes. Consider the following state space system,

$$\begin{aligned} X_{t+1,T} &= F_{t,T}X_{t,T} + V_{t,T}, \\ Y_{t,T} &= G_{t,T}X_{t,T} + W_{t,T}, \end{aligned} \tag{14}$$

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 (7) can be represented by a state space system (14) by generalizing the infinite-dimensional equations given in (Hannan and Deistler, 1988, p.22) to the nonstationary case.

According to Palma et al. (2013), the process (7) can be represented by the following

infinite-dimensional state space system

$$\begin{aligned} X_{t+1,T} &= \begin{bmatrix} 0 \\ I_\infty \end{bmatrix} X_{t,T} + \begin{bmatrix} 1\ 0\ 0 \cdots \end{bmatrix}' \varepsilon_{t+1}, \\ Y_{t,T} &= \sigma(\tfrac{t}{T}) \begin{bmatrix} 1\ \psi_1(\tfrac{t}{T})\ \psi_2(\tfrac{t}{T})\ \psi_3(\tfrac{t}{T})\ \cdots \end{bmatrix} X_{t,T}, \end{aligned} \tag{15}$$

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. In some cases, this state space representation may not be minimal. For instance, for LSAR(2) processes, the state space is 2-dimensional:

$$X_{t+1,T} = \begin{bmatrix} a_1(\frac{t}{T})\ a_2(\frac{t}{T}) \\ 0 \qquad 1 \end{bmatrix} X_{t,T} + \varepsilon_{t+1}, \quad Y_{t,T} = [1\ 0]\, X_{t,T}.$$

It is ussually more practical to approximate the model by,

$$Y_{t,T} = \sigma\left(\tfrac{t}{T}\right) \sum_{j=0}^{m} \psi_j\left(\tfrac{t}{T}\right) \varepsilon_{t-j}, \tag{16}$$

for $t = 1, \ldots, T$ and some positive integer m. A finite-dimensional state state system for (16) is given by

$$\begin{aligned} X_{t+1,T} &= \begin{bmatrix} 0 & 0 \\ I_m & 0 \end{bmatrix} X_{t,T} + \begin{bmatrix} 1\ 0 \cdots\ 0 \end{bmatrix}' \varepsilon_{t+1} \\ Y_{t,T} &= \sigma(\tfrac{t}{T}) \begin{bmatrix} 1\ \psi_1(\tfrac{t}{T})\ \psi_2(\tfrac{t}{T})\ \psi_3(\tfrac{t}{T})\ \cdots \psi_m(\tfrac{t}{T}) \end{bmatrix} X_{t,T}, \end{aligned} \tag{17}$$

for $t = 1, \ldots, T$, where I_r denotes the $r \times r$ identity matrix hereafter. Let $r_m = \mathrm{Var}[\sum_{j=m+1}^{\infty} \psi_j(u)\varepsilon_{t-j}]$ be the variance of the truncation error for approximating $\{Y_{t,T}\}$ by the finite moving average expansion (16). Then, according to Palma et al. (2013), the asymptotic magnitude of the truncation error when approximating (7) by (16) is, $r_m \sim \mathcal{O}(e^{-am})$ for a short-memory process and $r_m \sim \mathcal{O}(m^{2d-1})$ for a long-memory process, for large m, where $a > 0$ and $d = sup_u d(u) < 1/2$.

4. Estimation

4.1 Whittle Method

Let $\theta \in \Theta$ be a parameter vector specifying model (2) where the parameter space Θ is a subset of a finite-dimensional Euclidean space. Given a sample $\{Y_{1,T}, \ldots, Y_{T,T}\}$ of the process (2) we can estimate θ by minimizing the Whittle log-likelihood function

$$\mathcal{L}_T(\theta) = \frac{1}{4\pi} \frac{1}{M} \int_{-\pi}^{\pi} \sum_{j=1}^{M} \left\{ \log f_\theta(u_j, \lambda) + \frac{I_N(u_j, \lambda)}{f_\theta(u_j, \lambda)} \right\} d\lambda, \tag{18}$$

where $f_\theta(u, \lambda) = |A_\theta(u, \lambda)|^2$ is the time-varying spectral density of the limiting process specified by the parameter θ, $I_N(u, \lambda) = \frac{|D_N(u,\lambda)|^2}{2\pi H_{2,N}(0)}$ is a tapered periodogram with

$$D_N(u, \lambda) = \sum_{s=0}^{N-1} h\left(\frac{s}{N}\right) Y_{[uT]-N/2+s+1,T}\, e^{-i\lambda s}, \quad H_{k,N} = \sum_{s=0}^{N-1} h\left(\frac{s}{N}\right)^k e^{-i\lambda s},$$

$T = S(M-1)+N$, $u_j = t_j/T$, $t_j = S(j-1)+N/2$, $j = 1, \ldots, M$ and $h(\cdot)$ is a data taper.

The Whittle estimator of the parameter vector θ is given by

$$\widehat{\theta}_T = \arg\min \mathcal{L}_T(\theta), \tag{19}$$

where the minimization is over a parameter space Θ.

The analysis of the asymptotic properties of the Whittle locally stationary estimates (19) is discussed, for example, in Dahlhaus (1997) and Palma and Olea (2010).

4.2 State Space Method

Consider the state space representation (15) of $Y_{t,T}$. The Kalman filter equations can be used for estimating model parameters, state vectors, future observations and missing values. Let $\Delta_{t,T} = \text{Var}\left(Y_{t,T} - \widehat{Y}_{t,T}\right)$ be the prediction error variance and let $\Omega_{t,T} = \text{Var}\left(X_{t,T} - \widehat{X}_{t,T}\right) = (\omega_{i,j}(t,T))$ be the state prediction error variance-covariance matrix. The Kalman recursive equations are defined in Palma et al. (2013) as follows for the initial conditions $Y_{0,T} = (0, 0, \ldots)$, $\widehat{X}_1 = \text{E}(X_1) = (0, 0, \ldots)$ and $\Omega_{1,T} = \text{E}(X_1, X_1') = \text{diag}\{1, 1, \ldots\}$:

$$\begin{aligned}
\Delta_{t,T} &= \sigma^2\left(\tfrac{t}{T}\right)\sum_{i,j=1}^{\infty}\psi_{i-1}\left(\tfrac{t}{T}\right)\,\omega_{i,j}(t,T)\,\psi_{j-1}\left(\tfrac{t}{T}\right),\\
\Theta_{t\,T}(i) &= \sigma\left(\tfrac{t}{T}\right)\sum_{j=1}^{\infty}\omega_{i-1,j}(t,T)\,\psi_{j-1}\left(\tfrac{t}{T}\right),\\
\omega_{t+1,T}(i,j) &= \omega_{t,T}(i+1,j+1) + q_{i,j} - \delta(t)\,\Theta_{t,T}(i)\,\Theta_{t,T}(j)/\Delta_{t,T},\\
\widehat{Y}_{t,T} &= \sigma\left(\tfrac{t}{T}\right)\sum_{j=1}^{\infty}\psi_{j-1}\left(\tfrac{t}{T}\right)\,\widehat{X}_{t,T}(j),\\
\widehat{X}_{t+1,T}(i) &= \widehat{X}_{t,T}(i-1) + \Theta_{t,T}(i)(Y_{t,T} - \widehat{Y}_{t,T})/\Delta_{t,T},
\end{aligned} \tag{20}$$

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

Let θ be the model parameter vector, then the log-likelihood function (up to a constant) can be obtained from (20), $\mathcal{L}(\theta) = \sum_{t=1}^{T}\log\Delta_{t,T} + \sum_{t=1}^{T}\frac{\left(Y_{t,T}-\widehat{Y}_{t,T}\right)^2}{\Delta_{t,T}}$. Hence the exact MLE provided by the Kalman equations (20) is given by $\widehat{\theta} = \arg\max_{\theta\in\Theta}\mathcal{L}(\theta)$, where Θ is a parameter space. Note that the Kalman equations (20) can be applied directly to the general state space representation (14) or to the truncated representation (17), yielding in this case an approximate MLE.

4.3 Simulation Studies

4.3.1 Whittle Method

In order to gain some insight into the finite sample performance of the Whittle estimator discussed in Section 4.1 we report next a number of Monte Carlo experiments for the LSFN

model

$$Y_{t,T} = \sigma(t/T)\,(1-B)^{-d(t/T)}\,\varepsilon_t, \tag{21}$$

for $t = 1, \ldots, T$ with $d(u) = \alpha_0 + \alpha_1\, u,\ \sigma(u) = \beta_0 + \beta_1\, u$ and Gaussian white noise $\{\varepsilon_t\}$ with unit variance. Denote the parameter vector by $\boldsymbol{\alpha} = (\alpha_0, \alpha_1)$ for $d(\cdot)$ and $\boldsymbol{\beta} = (\beta_0, \beta_1)$ for the noise scale $\sigma(\cdot)$.

The samples of the LSFN process are generated by means of the innovation algorithm, see for example (Brockwell and Davis, 1991, p.172). In this implementation, the covariances of the process $\{Y_{t,T}\}$ is given by (13). The Whittle estimates in these Monte Carlo simulations have been computed by using the cosine bell data taper $h(x) = \frac{1}{2}[1-\cos(2\,\pi\,x)]$.

Table 1 reports the results from Monte Carlo simulations for several parameter values, based on 1000 replications. These tables show the average of the estimates as well as their theoretical and empirical standard deviations (SD) given by

$$\Gamma_{\boldsymbol{\alpha}} = \left[\frac{\pi^2}{6\,(i+j+1)}\right]_{i,j=0,1}, \qquad \Gamma_{\boldsymbol{\beta}} = 2\left[\int_0^1 \frac{u^{i+j}\,du}{\sigma^2(u)}\right]_{i,j=0,1}. \tag{22}$$

Observe from this table that the estimated parameters are close to their true values. Besides, the empirical SD are close to their theoretical counterparts. These simulations suggest that the finite sample performance of the proposed estimators seems to be very good in terms of bias and standard deviations.

Estimated

4.3.2 Kalman Filter Method

Consider the LSMA(∞) model described by (7) with

$$\psi_j(\,u) = \phi(\,u)^j, \quad \phi(\,u) = \alpha_0 + \alpha_1\, u, \quad \sigma(\,u) = \beta_0 + \beta_1\, u, \tag{23}$$

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 $\sigma(\cdot)$. In this case the theoretical SD are based on the formulas given by Theorem 3.1 of Dahlhaus (2000) Dahlhaus (2000) for the MLE of short-memory LS models:

$$\Gamma(\boldsymbol{\alpha})_{i,j} = \left[\int_0^1 \frac{u^{i+j-2}}{1-[\phi(u)]^2}\,du\right]_{i,j=1,2}. \tag{24}$$

Table 2 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 have been randomly selected for each simulation. Finally, we use sample size T=1024 and 1000 replications. 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.

5. Prediction Methods

In this section, we describe a procedure for obtaining one-step and multi-step predictors along with prediction bands by means of state space models. We will focus on the prediction

of future values of time series generated by LS processes with short or long memory and Gaussian innovations.

The Kalman filter described in Section 4.2 provides the state vector, $\widehat{\mathrm{X}}_{t+1,T}$, and its mean square error (MSE) based on the information available at time t. These estimates are given by

$$\widehat{\mathrm{X}}_{t+1,T} = F_{t,T}\mathrm{X}_{t,T} + \Theta_{t,T}\Delta_{t,T}^{-1}\Upsilon_{t,T}, \tag{25a}$$

$$\Omega_{t+1,T} = F_{t,T}\Omega_{t,T}F'_{t,T} + Q_{t,T} - \Theta_{t,T}\Delta_{t,T}^{-1}\Theta'_{t,T}, \tag{25b}$$

where $\Upsilon_{t,T} = Y_{t,T} - \widehat{Y}_{t,T}$ is the innovation, $\Delta_{t,T}$ is its variance given by $\Delta_{t,T} = G_{t,T}\Omega_{t,T}G'_{t,T} + R_{t,T}$ and $\Theta_{t,T} = F_{t,T}\Omega_{t,T}G'_{t,T}$.

5.1 Missing Observations

The analysis of missing observations in time series has been addressed by several authors including Harvey (1989) and Durbin and Koopman (2009), among others. The state space method and its associated Kalman filter algorithm provides a simple methodology for handling missing values.

In order to describe this procedure, assume that the set of observations $Y_{t,T}$ for $t =$

Table 1. Whittle estimation for model (21): Sample size $T = 1024$, block size $N = 128$ and shift $S = 64$.

Parameters		Estimates		Theoretical SD		Estimated SD	
α_0	α_1	$\widehat{\alpha}_0$	$\widehat{\alpha}_1$	$\sigma(\widehat{\alpha}_0)$	$\sigma(\widehat{\alpha}_1)$	$\widehat{\sigma}(\widehat{\alpha}_0)$	$\widehat{\sigma}(\widehat{\alpha}_1)$
0.10	0.20	0.080	0.213	0.049	0.084	0.060	0.112
0.15	0.25	0.133	0.277	0.049	0.084	0.066	0.109
0.20	0.20	0.187	0.221	0.049	0.084	0.067	0.113
0.20	0.25	0.197	0.258	0.049	0.084	0.057	0.092
0.25	0.20	0.252	0.204	0.049	0.084	0.057	0.091
0.10	0.20	0.078	0.215	0.049	0.084	0.059	0.114
0.15	0.25	0.132	0.275	0.049	0.084	0.067	0.115
0.20	0.20	0.193	0.214	0.049	0.084	0.067	0.114
0.20	0.25	0.195	0.262	0.049	0.084	0.058	0.088
0.25	0.20	0.254	0.198	0.049	0.084	0.055	0.089
β_0	β_1	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\sigma(\widehat{\beta}_0)$	$\sigma(\widehat{\beta}_1)$	$\widehat{\sigma}(\widehat{\beta}_0)$	$\widehat{\sigma}(\widehat{\beta}_1)$
0.50	0.50	0.498	0.509	0.027	0.056	0.030	0.064
0.50	0.50	0.499	0.518	0.027	0.056	0.032	0.066
0.50	0.50	0.499	0.518	0.027	0.056	0.031	0.064
0.50	0.50	0.500	0.521	0.027	0.056	0.030	0.062
0.50	0.50	0.506	0.513	0.027	0.056	0.030	0.060
1.00	-0.50	0.998	-0.494	0.038	0.056	0.042	0.067
1.00	-0.50	1.001	-0.492	0.038	0.056	0.044	0.067
1.00	-0.50	1.004	-0.495	0.038	0.056	0.043	0.066
1.00	-0.50	1.003	-0.491	0.038	0.056	0.041	0.062
1.00	-0.50	1.008	-0.498	0.038	0.056	0.040	0.062

Table 2. Estimation of model (23) with $(\alpha_0,\, \alpha_1, \beta_0,\, \beta_1) = (-0.2, 0.7, 0.6, 0.4)$

	m=40			m=80		
% NA	0%	10%	20%	0%	10 %	20%
$\widehat{\alpha}_0$	-0.2019	-0.2032	-0.1982	-0.2032	-0.2018	-0.1988
$\widehat{\alpha}_1$	0.7011	0.7033	0.6961	0.7036	0.7027	0.6950
$\widehat{\beta}_0$	0.6052	0.5821	0.5554	0.6058	0.5945	0.5831
$\widehat{\beta}_1$	0.3891	0.3608	0.3370	0.3945	0.3912	0.3856
$\sigma(\widehat{\alpha}_0)$	0.0613	0.0646	0.0686	0.0613	0.0646	0.0686
$\sigma(\widehat{\alpha}_1)$	0.1027	0.1082	0.1148	0.1026	0.1082	0.1149
$\sigma(\widehat{\beta}_0)$	0.0305	0.0308	0.0311	0.0306	0.0317	0.0330
$\sigma(\widehat{\beta}_1)$	0.0604	0.0607	0.0612	0.0606	0.0628	0.0654
$\widehat{\sigma}(\widehat{\alpha}_0)$	0.0613	0.0647	0.0679	0.0608	0.0641	0.0687
$\widehat{\sigma}(\widehat{\alpha}_1)$	0.1024	0.1073	0.1152	0.1028	0.1091	0.1150
$\widehat{\sigma}(\widehat{\beta}_0)$	0.0313	0.0314	0.0319	0.0310	0.0316	0.0323
$\widehat{\sigma}(\widehat{\beta}_1)$	0.0604	0.0604	0.0619	0.0600	0.0626	0.0658

$n+1, \ldots, T-n$ is missing, the vector $\Upsilon_{t,T}$ and the matrix $\Theta_{t,T}$ of the Kalman filter are set to zero for these values, that is, $\Upsilon_{t,T} = 0$ and $\Theta_{t,T} = 0$, and the Kalman updates become

$$\widehat{\mathrm{X}}_{t+1,T} = F_{t,T}\mathrm{X}_{t,T}, \tag{26a}$$

$$\Omega_{t+1,T} = F_{t,T}\Omega_{t,T}F'_{t,T} + Q_{t,T}, \tag{26b}$$

for $t = n+1, \ldots, T-n$. Forecasts of $Y_{n+k,T}$ together with their forecast error, can be obtained by treating $Y_{t,T}$ for $t > n$ as missing observations and continuing the Kalman filter beyond $t = n$ with $\Upsilon_{t,T} = 0$ and $\Theta_{t,T} = 0$ for $t > n$.

5.2 One step forecasting

Let us suppose we have observations $\{Y_{1,T}, \ldots Y_{n,T}\}$ which follow the state space model (14) and we wish to forecast $Y_{n+1,T}$. In this case the forecast is fairly straightforward calculation. Namely, from (14), we have $Y_{n+1,T} = G_{n+1,T}X_{n+1,T} + W_{n+1,T}$ so the one-step in-sample predictor is given by

$$\widehat{Y}_{n+1,T} = G_{n+1,T}\widehat{X}_{n+1,T},$$

where $\widehat{X}_{n+1,T}$ is the estimate (26a) of $X_{n+1,T}$ produced by the Kalman filter and $G_{n+1,T} = \left[1, \psi_1\left(\frac{n+1}{T}\right), \psi_2\left(\frac{n+1}{T}\right), \ldots\right]$. The corresponding MSE is given by

$$\Delta_{n+1,T} = G_{n+1,T}\Omega_{n+1,T}G'_{n+1,T},$$

where $\Omega_{n+1,T}$ is given by (26b). On the other hand, out-of-sample predictor $\widehat{Y}_{T+1}$ can be obtained from the Kalman filter equations by redefining the sample size $\tilde{T} = T+1$ and merely treating Y_{T+1} as missing value. Thus, the best linear mean square predictor, is given by

$$\widehat{Y}_{T+1,\tilde{T}} = G_{T+1,\tilde{T}}\widehat{X}_{T+1,\tilde{T}},$$

where $\widehat{X}_{T+1,\tilde{T}}$ is produced by the Kalman filter relation (26a) and $G_{T+1,\tilde{T}} = [1, \psi_1(1), \psi_2(1), \ldots]$. Finally, the MSE is given by $\Delta_{T+1,\tilde{T}} = G_{T+1,\tilde{T}}\Omega_{T+1,\tilde{T}}G'_{T+1,\tilde{T}}$, with $\Omega_{T+1,\tilde{T}} = F_{T,\tilde{T}}\Omega_{T,\tilde{T}}F'_{T,\tilde{T}} + Q_{T,\tilde{T}}$.

5.3 Multi-step forecasting

Let $\widehat{Y}_{n+k,T} = \mathrm{E}(Y_{n+k,T}|Y_{n,T}, Y_{n-1,T}, \ldots, Y_{1,T})$ be the k-step-in-sample predictor based on the finite past for $1 \leq n+k \leq T$. Following the procedure described in the previous section, these forecasts are obtained from the Kalman recursive equations given by (26), for more details see Palma et al. (2013).

Assuming that future prediction errors are Gaussian, the k-step ahead prediction bands for Y_{n+k} are given by

$$\left[\ \widehat{Y}_{n+k,T} - z_{1-\alpha/2}\sqrt{\Delta_{n+k,T}},\ \widehat{Y}_{n+k,T} + z_{1-\alpha/2}\sqrt{\Delta_{n+k,T}}\ \right], \tag{27}$$

where $z_{1-\alpha/2}$ is the $\left(1-\frac{\alpha}{2}\right)$-percentile of the standard normal distribution, and $\widehat{Y}_{n+k,T}$ and $\Delta_{n+k,T}$ are the k-steps ahead prediction of Y_{n+k} and its MSE given by

$$\widehat{Y}_{n+k,T} = G_{n+k,T}F^k\widehat{\mathrm{X}}_{n,T}, \tag{28a}$$

$$\Delta_{n+k,T} = G_{n+k,T}\left(F^k\right)\Omega_{n,T}\left(F^k\right)' G'_{n+k,T} + G_{n+k,T}\sum_{j=0}^{k-1}\left[\left(F^j\right)Q\left(F^j\right)'\right]G'_{n+k,T} \tag{28b}$$

for $k = 1, \ldots, T-n$, where $G_{n+k,T} = \left[1, \psi_1\left(\frac{n+k}{T}\right), \psi_2\left(\frac{n+k}{T}\right), \ldots\right]$.

Finally, the k-step linear predictor for out-sample $\widehat{Y}_{T+k,T}$ with $k > 0$, is obtained by re-defining the sample size $\tilde{T} = T+k$ and considering the observations $\tilde{T}+1, \ldots, \tilde{T}+k$ as missing data. In practice, the unknown parameters involved in (28) and (27) are substituted by consistent estimates such as MLE, which can be obtained by means of the Kalman recursive equations.

5.4 Simulation Studies

This section reports the results from several Monte Carlo experiments carried out to analyze the finite sample behavior of one-step and multi-step predictors of short and long-memory LS processes. In order to study the behavior of the prediction bands defined in (27), we consider their coverage and length, and the proportion of observations lying out to both left and right. We compare these measures with the nominal coverage. In order to do so, we consider simulations of a particular series generated by the specific LSMA process discussed in Section 2, generating $R = 1000$ future values $Y_{n+k,T}$ from that series. Then, for the LSMA process, we obtained a prediction band denoted by (L, U) and estimate the coverage by

$$\widehat{\alpha} = \#(L \leq Y^r_{n+k} \leq U)/R,$$

where $Y^r_{n+k}(r = 1, 2, \ldots, R)$ are the values generated previously. We have carried out 1000 Monte Carlo experiments and report average coverage, average length and coverage proportion of observations on the left and on the right for each prediction interval.

Thus, the best in-sample linear predictor (28a) was calculated for $k = 1, 5$ and 15. We also consider different sample sizes $T = 112, 512$ and 1024, with $n = T-k$. In order to check the suitability of $\widehat{Y}_{n+k,T}$, let us consider two models: the first is a short-memory LSAR(1) process and the second is a long-memory LSFN process. Both models have time-varying parameters. For the LSAR(1) case, we consider time-varying $\phi(u)$ parameter and scale factor $\sigma(u)$ given by specification (23) with $(\alpha_0, \alpha_1, \beta_0, \beta_1) = (-0.4, 0.8, 0.5, 0.5)$. On the other hand, for the LSFN(d) case with time-varying long-memory parameter $d(u)$ and scale factor given by specification (21) with $(\alpha_0, \alpha_1, \beta_0, \beta_1) = (0.2, 0.25, 0.5, 0.5)$. The models are estimated using Kalman filters with truncation $m = 30$. Observations $t = 1, \ldots, n$ and values $t = n+1, \ldots, n+k$ are left for out-of-sample forecasting and then obtain, 95% prediction bands computed using (27). Finally, we compute the coverage of each of these bands as well as the length and the percentage of observations left out on the right and on the left of the limits of the prediction bands. Table 3 reports the Monte Carlo averages of

Table 3. Monte Carlo average coverages, length and percentage of observations left out on the right and on the left of the prediction bands for Y_{n+k} and the nominal coverage is 95%.

Model	Sample Size	k	Average Coverage	Coverage below/above	Average Length
LSAR(1)	128	1	0.9294	0.0339 / 0.0368	3.7237
		5	0.9163	0.0419 / 0.0418	3.8944
		15	0.8285	0.0858 / 0.0857	3.9162
	512	1	0.9477	0.0261 / 0.0262	3.8715
		5	0.9421	0.0300 / 0.0279	4.1536
		15	0.9267	0.0373 / 0.0360	4.1560
	1024	1	0.9464	0.0282 / 0.0254	3.8913
		5	0.9447	0.0297 / 0.0256	4.2385
		15	0.9410	0.0316 / 0.0274	4.1518
LSFN	128	1	0.9357	0.0267 / 0.0376	3.8571
		5	0.9164	0.0411 / 0.0425	4.3063
		15	0.7753	0.1132 / 0.1115	4.3905
	512	1	0.9379	0.0312 / 0.0309	3.9362
		5	0.9265	0.0364 / 0.0371	4.6673
		15	0.9032	0.0474 / 0.0494	4.7137
	1024	1	0.9480	0.0256 / 0.0264	3.9441
		5	0.9315	0.0356 / 0.0330	4.7418
		15	0.9135	0.0423 / 0.0442	4.9860

these quantities. The average coverage of the bands is smaller than the nominal coverage when the sample size is small ($T = 128$) and the prediction horizon is 5 and 15. When comparing the prediction bands of these models, we can observe that in both cases their average coverage get closer to their nominal values as the sample sizes increases to 512 and 1024 observations.

5.5 Data Application

In this section, we apply the locally stationary methodology for estimating, predicting and handling missing values to real-life time series consisting of tree-ring measurements.

5.6 Tree-Ring Data

Ring count is a typical procedure in studies of forest mass to determine growth and yield of both natural forests and forest plantations. This methodology is known as forest analysis, and it can be only implemented in species growing in temperate regions, where it is easy to identify the ring growth. In tropical climates, where there is little differentiation among seasons, growth rates are constant, making it difficult to clearly differentiate spring and winter wood. Consequently, this data set can be used as climate proxies and to indicate the chances of temperature and precipitation conditions in paleoclimatology, see Tan et al. (2003).

Figure 1 plots annual tree-ring width of the Pinus Longaeva form the Mammoth Creek, Utah, from 0 AD to 1989 AD. This data set, available at the National Climatic Data Center, was reported in Graybill (1990). Following to Ferreira et al. (2013), tree ring

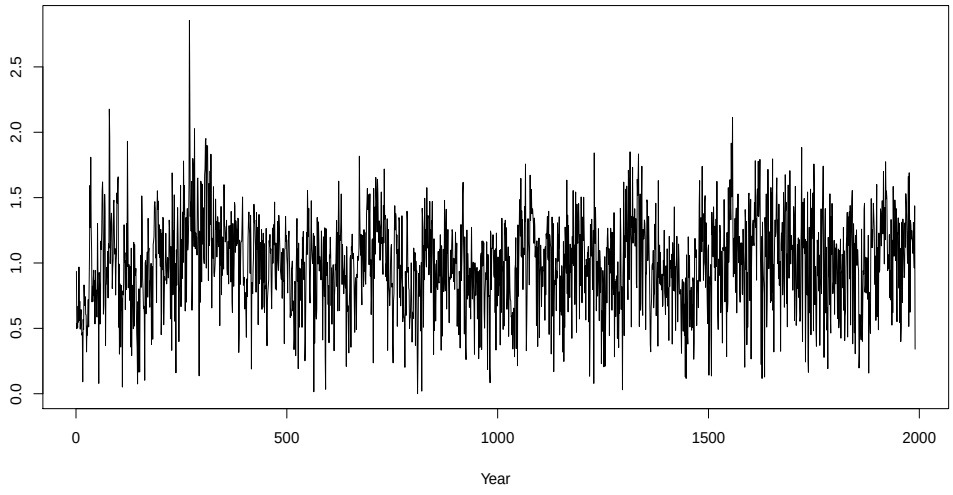

Figure 1. *Tree Ring Data.*

data can be represented by a model with a constant mean over time and a LS process, i.e. $Y_{t,T} = \mu + \varepsilon_{t,T}$, without loss of generality, we can define a zero-mean process by $X_{t,T} = Y_{t,T} - \widehat{\mu}$, with sample mean given by $\widehat{\mu} = 1.0017$.

Figure 2 shows the sample ACF of $X_{t,T}$ in panel (a), and the corresponding variances of the sample mean in panel (b). The dashed line corresponds to its expected behavior for a short-memory case with blocks of k observations, whereas the continuous line represents the expected behavior for a long-memory case. From both panels, this series seems to exhibit long-range dependence.

In addition, a closer look at the sample ACF of the data reveals that the degree of persistence seems to vary over time. Indeed, Figure 3 shows the sample autocorrelation of three segments of the sample: observations 1 to 500, observations 750 to 1250 and observations 1490 to 1990. This figure provides information for arguing possible changes in the degree of dependence. This represents a clear evidence of a nonstationary process. Therefore, it seems that the data has a time-varying long-memory structure. In order to

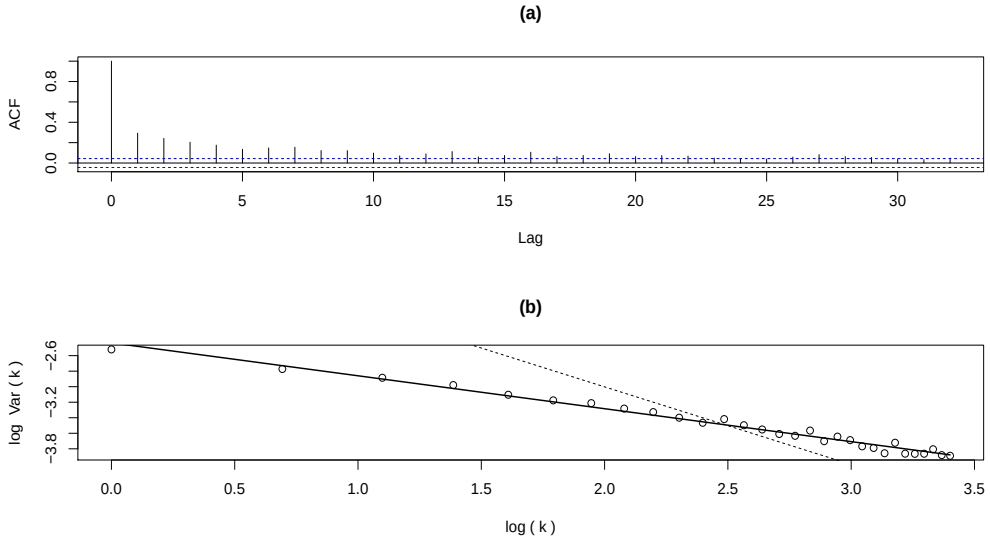

Figure 2. *Tree Ring Data. (a) Sample ACF, (b) Variance Plot*

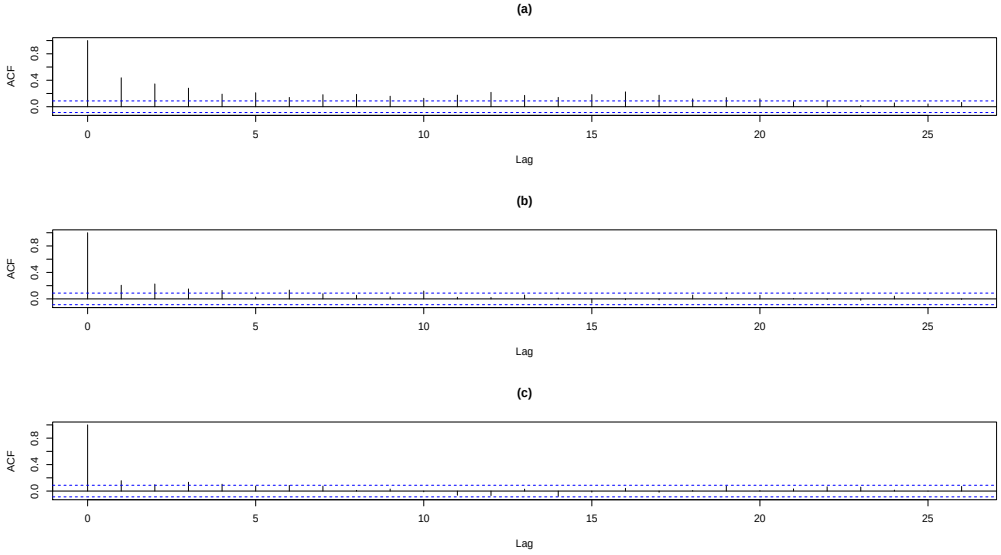

Figure 3. *Tree Ring Data. Sample ACF: (a) Observations 1 to 500,(b) Observations 750 to 1250, (c) Observations 1490 to 1990.*

handle these features, a LSFN process was used. Figure 4 shows an heuristic estimator of the long-memory parameter and the variance of the noise scale along with stationary fractional noise and locally stationary fractional noise model estimates of these quantities. From this figure we suggest a linear and quadratic function for $d(\,u)$ and $\sigma(\,u)$ respectively; i.e,

$$d(u) = \alpha_0 + \alpha_1\, u, \quad \sigma(u) = \beta_0 + \beta_1\, u + \beta_2\, u^2. \tag{29}$$

Table 4 reports the parameter estimates using the Kalman filter for model (21) with truncation level $m = 80$. The standard deviations and the t-tests have been obtained using (22) for $d(\,u)$ and $\sigma(\,u)$, respectively. As we can observe in this table, the parameters (α_0, α_1) and $(\beta_0, \beta_1, \beta_2)$ are statistically significant at the 5% level.

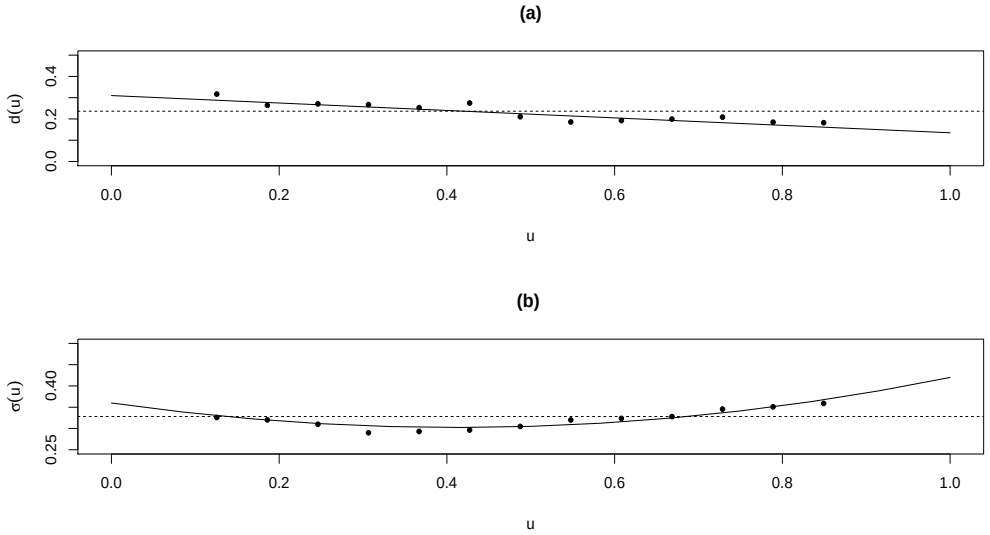

Figure 4. *(a) Estimates of the long-memory parameter. (b) Estimates of the noise variance. In all panels the heavy line represents the locally stationary FN model, the horizontal broken line indicates the stationary FN model and the dots represent the heuristic approach.*

Table 4. Tree Ring Data: Parameters estimated with a truncation m=80

Parameter	Estimates	Standard Deviation	$t-$value
α_0	0.3294943	0.03495660	9.425811
α_1	-0.2005137	0.06054661	-3.311725
β_0	0.3391996	0.01566141	21.65831
β_1	-0.1638700	0.07259525	-2.257310
β_2	0.2076965	0.07137396	2.909976

The residuals of the model, $e_t = X_{t,T} - \widehat{X}_{t,T}$, along with their SD, $\Delta_{t,T}^{1/2}$, are plotted in Figure 5. In particular, the figure exhibits three panels exploring the structure of the standardized residuals $r_t = e_t \Delta_{t,T}^{-1/2}$: panel (a) displays the standardized residuals for the LSFN estimates, panel (b) shows the sample ACF, and panel (c) presents the Ljung-Box statistic. From panel (b), it seems that there are no significant autocorrelations in the residuals. This conclusion is formally supported by the Ljung-Box tests. Consequently, the white noise hypothesis cannot be rejected at the 5% level.

In order to validate our estimates, we consider a data gap consisting of two blocks of 100 observations each removed from the sample. These blocks are located at $t = 501, \ldots, 600$ and $t = 1891, \ldots, 1990$, respectively. Figure 6 displays both forecasting $\widehat{Y}_{500+k}$, $\widehat{Y}_{1890+k}$ for $k = 1, 2, \ldots, 100$ and their respective 95% predictions bands. These multi-step-ahead prediction bands are based on formula (27). Notice that, as expected, most of the future observations fall inside the prediction bands.

The evolution of the prediction error standard deviations is depicted in Figure 7. The dashed line represents the SD of the process $\sigma(X_{t,T})$, the dotted line corresponds to the noise scale SD, $\sigma(u) = 0.3391996 - 0.1638700\, u + 0.2076965\, u^2$, and the continuous line denotes the estimated SD of the prediction error $\Delta_{t,T}^{1/2}$. From this figure, notice that the dashed line is an upper limit for the SD of the prediction error. In addition, it can be observed that $\Delta_{t,T}^{1/2}$ increases right after the beginning of the data gap and it decays to $\sigma(t/T)$ as new observations become available. Therefore we can conclude that $\sigma(t/T) \leq \Delta_{t,T}^{1/2} \leq \sigma(X_{t,T})$.

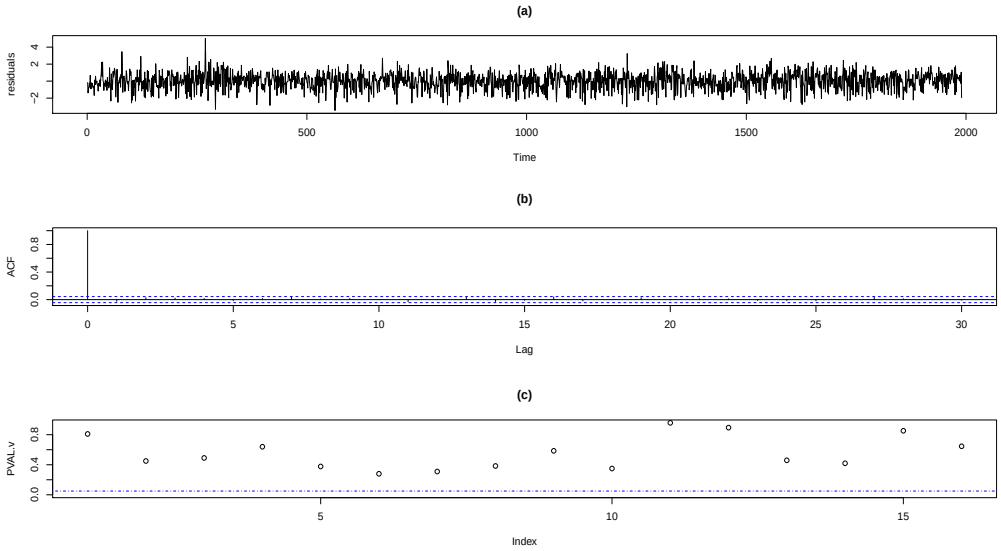

Figure 5. *Tree Ring Data: Residuals analysis (a)Residuals from the fitted model, (b) Sample ACF, (c) Ljung-Box tests .*

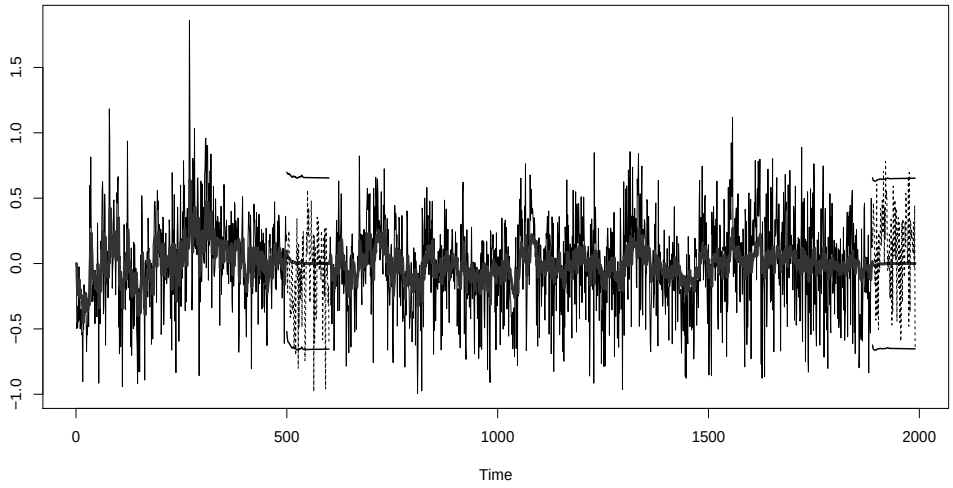

Figure 6. *Tree Ring Data: Multi-step forecasts of two blocks of missing values and 95% prediction bands.*

6. Conclusions

In this paper we have reviewed the locally stationary processes, including a number of estimation and prediction techniques. These models have proved to provide useful tools for analyzing nonstationary time series which often arise in several fields. As shown in this work, the LS methodologies allows for the modeling of real-life time series such as tree rings, sheding some light into the dependence structure of these series.

Acknowledgement

The first author would like to express his thanks for the support from DIUC 213.014.021-1.0, established by the the Universidad de Concepción. The second author gratefully ac-

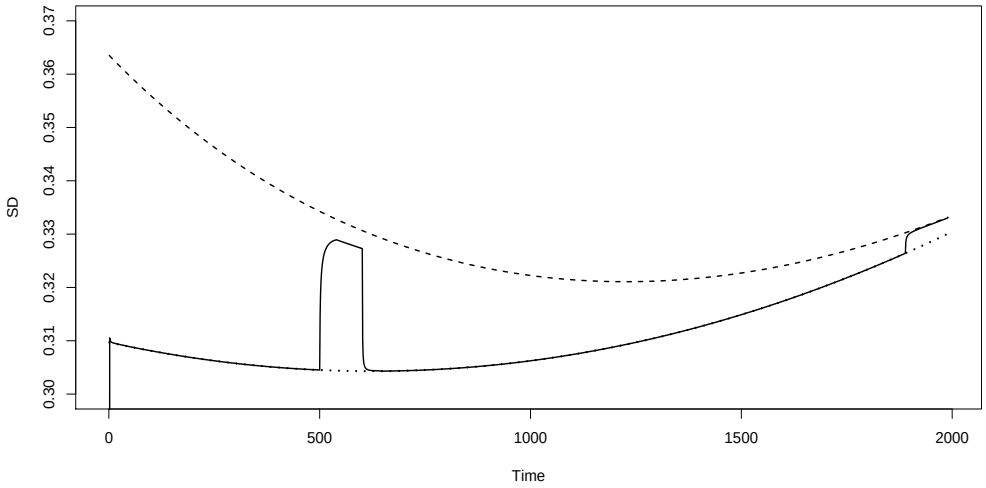

Figure 7. *Prediction of LSFN model: Broken line: Standard deviation of* $X_{t,T}$*, Dotted line: SD of the noise* $\sigma(t/T)$ *and Heavy line: Empirical prediction error SD,* $\Delta_{t,T}^{1/2}$.

knowledges the financial support from Project Fondecyt 11121128. Finally the third author gratefully acknowledges partial financial support from Fondecyt Grant 1120758 established by the Chilean Government.

References

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

Cavanaugh, J.E.. Wang, Y., Davis, J.W., 2003. Locally self-similar processes and their wavelet analysis. In Stochastic processes: modelling and simulation, Vol. 21, Handbook of Statist. North-Holland, Amsterdam, pp. 93–135.

Chandler, G., Polonik, W., 2006. Discrimination of locally stationary time series based on the excess mass functional. Journal of the American Statistical Association, 101, 240–253.

Dahlhaus, R., 1996a. Asymptotic statistical inference for nonstationary processes with evolutionary spectra. In Athens Conference on Applied Probability and Time Series Analysis, Vol. II (1995), Vol. 115, Lecture Notes in Statist. Springer, New York, pp. 145–159.

Dahlhaus, R., 1996b. On the Kullback-Leibler information divergence of locally stationary processes. Stochastic Processes and their Applications, 62, 139–168.

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

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

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

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

Durbin, J., Koopman, S.J., 2001. Time Series Analysis by State Space Methods, Vol. 24, Oxford Statistical Science Series. Oxford University Press, Oxford. 142

Ferreira, G., Rodriguez, A., Lagos, B., 2013. Kalman filter estimation for a regression

model with locally stationary long memory errors. Computational Statistics and Data analysis 62, 52–69.

Graybill, D.A., 1990. Pinus longaeva tree ring data. Mammoth Creek, Utah, National Climatic Data Center.

Hannan, E.J., Deistler, M., 1988. The Statistical Theory of Linear Systems. Wiley, New York.

Harvey, A.C., 1989. Forecasting Structural Time Series and the Kalman Filter. Cambridge University Press, Cambridge.

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

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

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. The Annals of Statistics, 38, 2958–2997.

W. Palma, R. Olea, and G. Ferreira. 2013. Estimation and forecasting of locally stationary processes. Journal of Forecasting 32, 86–96.

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

Priestley, M.B., Tong., H., 1973. On the analysis of bivariate non-stationary processes. Journal of the Royal Statistical Society. Series B. Methodological 35, 153–166, 179–188.

Tan, M., Liu, T.S., 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. 30, 191–194.

Tong, H., 1973. Some comments on spectral representations of non-stationary stochastic processes. Journal of Applied Probability 10, 881–885.

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