EcoDep Seminar
Forecasting highly persistent time series with bounded spectrum processes

Federico Maddanu [1]
CY University

February 28, 2022

[1]Address for Correspondence: AGM Laboratory, Cergy University, Av. Adolphe Chauvin 95302 CERGY-PONTOISE, France. Email: federico.maddanu@gmail.com.

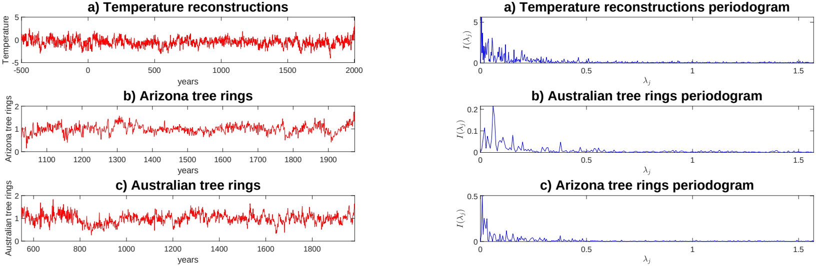
Introduction

1. Long memory models have been widely considered in the geophysical literature to investigate on the dependence structure of many climate time series.
2. According to Beran (1994), a stochastic process exhibits long memory if the autocovariance function $\gamma(k)$ is not absolutely summable or, equivalently, if the spectral density $f(\lambda)$ is unbounded at some frequency $\lambda_0 \in [0, \pi]$, such that $\gamma(k) \sim k^{2d-1} C_\gamma(k, d)$ as $k \to \infty$ and $f(\lambda) \sim \lambda^{-2d} C_f(k, d)$ as $\lambda \to \lambda_0$ where d is defined as the memory parameter and $C_\gamma(k, d)$ and $C_f(\lambda, d)$ are two slowly varying function.

Introduction

1. We focus on the analysis of three climate time series where the persistence in the data is mostly concentrated at the long-run frequency. The first series concerns yearly central European summer temperature reconstructions based on Austrian Alps Tree ring data. The remaining series are two yearly tree ring records.
2. Probably, we remember from our childhood that the age of a tree could be found by counting its rings. Moreover, the ring widths can be used to estimate past temperature or precipitation over the lifetime of the tree, such that we can learn about climate history for hundreds and thousands of years.

Figure: (a) Paleo-temperature reconstruction, (b) Australian pine tree rings and (c) Arizona tree ring series: time series plot and periodogram.



- A possible candidate model to analyse these data is the the fractional noise (FN) process (cf. Hosking (1981) and Andvel (1986))

$$(1-L)^d y_t = \varepsilon_t \quad , \quad \varepsilon_t \sim WN(0, \sigma^2) \tag{1}$$

which displays long memory if $d \in (0, 0.5)$ with spectral density

$$f(\lambda) = \frac{\sigma^2}{2\pi}\left(2\sin(\frac{\lambda}{2})\right)^{-2d} \tag{2}$$

unbounded at the long-run frequency.

- The FN process can be generalized by the larger class of Fractional equal-root Autoregressive Moving Average (FerARMA) processes, which encompasses both long memory and short memory specifications (eg. the standard ARMA processes).

The FerARMA class of processes

- The FerARMA class is defined by the following fractional process:

$$\phi(L)^{d_1} y_t = \psi(L)^{d_2} \varepsilon_t, \tag{3}$$

  where $L^k y_t = y_{t-k}$, $\phi(z) = \sum_{k=1}^{p} \phi_k z^k$, $\psi(z) = \sum_{k=1}^{q} \psi_k z^k$ and $\varepsilon_t \sim WN(0, \sigma^2)$, such that the process is stationary and invertible if the roots of the $\phi(z)$ and $\psi(z)$ polynomials lie outside the unit circle, while the parameters d_1, d_2 belong to the interval $(0, 1]$, .
- The standard ARMA(p, q) process is obtained if $d_1 = d_2 = 1$.
- A long memory process can be obtained eg. if $d_1 \in (0, 1)$ with a unit root in the AR polynomial, $\phi(1) = 0$.
- **Our purpose is to introduce the FerARMA class as an alternative way to fit and forecast highly persistent time series.**

The FerARMA class of processes

- In our context, we will focus on a reduced form of (3) with $p, q \in \{0, 1\}$, that is the FerARMA$(1, d_1, d_2, 1)$ process

$$(1 - \phi L)^{d_1} y_t = (1 - \psi L)^{d_2} \varepsilon_t \tag{4}$$

  where $\varepsilon_t \sim N(0, \sigma^2)$. The process is second order stationary and exhibits short memory for any value of d_1 and d_2 in $(0, 1]$ and ϕ and ψ in $[0, 1)$.
- Notice that if $\phi = d_2 = 1$ we obtain a long memory FIMA process (c.f. Hosking, 1981) with MA(1) errors, which is stationary if $d_1 < 0.5$ with an unbounded spectral density at the origin.

Comparison with respect to long memory models

- The FerARMA model is stationary also for values of $d > 0.5$, allowing for the support of the Durbin-Levinson recursion in the computation of the best linear predictor (see Brockwell and Davis, 1986).

- The FerARMA process displays a bounded and continuous spectrum, allowing for a CLT of the Whittle estimator (c.f. Whittle, 1953) under less restrictive assumptions wrt the asymptotic theory on long memory models given in Fox and Taqqu (1986) and Velasco and Robinson (2000) for the stationary and no-stationary case, respectively.

The Fractional equal root AR(1) or Spolia process

The FerAR(1,d) or Spolia process is an interesting special case of (4). It was initially introduced by Spolia et al. (1980) and, more recently, it has been formalized in Peiris (2003). It is described by:

$$y_t = (1 - \phi L)^{-d} \varepsilon_t \tag{5}$$

where $\varepsilon_t \sim N(0, \sigma^2)$. The process is stationary if $\phi \in (0, 1)$. The spectral density is given by:

$$f(\omega) = \frac{\sigma^2}{2\pi} |1 - \phi e^{-\imath\omega}|^{-2d} = \frac{\sigma^2}{2\pi} (1 + \phi^2 - 2\phi \cos(\omega))^{-d} \quad \omega \in [-\pi, \pi] \tag{6}$$

The ACF is given by:

$$\gamma(k) = \sigma^2 \frac{\Gamma(k+d)}{\Gamma(d)\Gamma(k+1)} \phi^k F(d, k+d; k+1; \phi^2) \qquad k \in \mathbb{Z} \qquad (7)$$

where $\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx$ is the Euler gamma function, while

$$F(a, b; c; z) = \sum_{j=0}^{\infty} \frac{\Gamma(a+j)\Gamma(b+j)\Gamma(c)}{\Gamma(a)\Gamma(b)\Gamma(c+j)\Gamma(j+1)} z^j$$

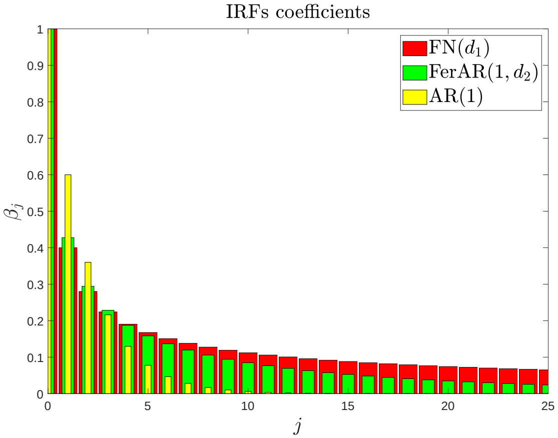
is the hypergeometric function defined for $|z| < 1$.

The Fractional equal root AR(1) or Spolia process

- Since both the AR(1) and FN specification are special case of the Spolia process (for $d = 1$ and $\phi = 1$, respectively) it may be interesting to compare them in terms of persistence.
- The persistence of a process is characterized by the speed at which its impulse response function (IRF) $\sum_{j=0}^{\infty} \beta_j$ vanishes, where $\beta_j \sim \frac{1}{\Gamma(d)} j^{d-1} \phi^j$ as $j \to \infty$.
- The next figure highlights the hybrid characteristic of the Spolia process in a suitable parameter's set showing how the Spolia IRF coefficients decays slower than in the case of the AR(1) process but faster with respect to the FN specification.

The Fractional equal root AR(1) or Spolia process

Figure: IRFs coefficients according to the FN(d_1) process with parameter $d_1 = 0.40$, the AR(1) process with parameter $\phi_1 = 0.60$ and the FerAR($1, d_2$) process with parameters $d_2 = 0.45$ and $\phi_2 = 0.95$.



The Fractional equal root AR(1) process with MA(1) errors

Another interesting special case of (4) is the Spolia process with MA(1) errors (that is the FerARMA$(1, d, 1, 1)$), formalised in Shitan and Peiris (2011). It is defined by:

$$(1-\phi L)^d y_t = (1-\psi L)\varepsilon_t \tag{8}$$

where $\varepsilon_t \sim N(0, \sigma^2)$. The spectral density is:

$$f(\omega) = \frac{\sigma^2}{2\pi} \frac{1+\psi^2 - 2\psi\cos(\omega)}{(1+\phi^2 - 2\phi\cos(\omega))^d} \quad \text{with} \quad \omega \in [-\pi, \pi] \tag{9}$$

The variance is given as:

$$\gamma(0) = \sigma^2 \left[(1+\psi^2)F(d, d; 1; \psi^2) - 2\phi\psi d F(d, 1+d; 2, \phi^2) \right] \tag{10}$$

The Fractional equal root AR(1) process with MA(1) errors

The ACF for $|k| \geq 1$ follows according to:

$$\gamma(k) = \frac{\sigma^2 \phi^{|k|-1} \Gamma(|k| + d - 1)}{\Gamma(d)\Gamma(|k| + 2)} \times \tag{11}$$

$$\times \Big\{ \phi(1 + \psi^2)(|k| + 1)(|k| + d - 1)F(d, |k| + d; |k| + 1; \psi^2) +$$

$$-\psi\phi^2(|k| + d)(|k| + d - 1)F(d, |k| + 1 + d; |k| + 2, \phi^2) +$$

$$-\psi|k|(|k| + 1)F(d, |k| - 1 + d; |k|, \phi^2) \Big\}$$

Parameters Estimation: The Exact likelihood

We consider two full likelihood approaches to estimate the class of FerARMA models, through the maximization of the exact and Whittle likelihood (cf. Whittle, 1953).

The Exact likelihood

Let $y = \left[y_1, \quad \cdots \quad, y_T\right]'$ be T realizations of a random variable $Y \sim N(0, \Sigma_T)$ where Σ_T is the covariance matrix. The exact log-likelihood function can be computed as:

$$LL(\theta) = -\frac{T}{2}\log(2\pi) - \frac{1}{2}y'\Sigma_T^{-1}(\theta)y - \frac{1}{2}\log|\Sigma_T(\theta)| \qquad (12)$$

where $\theta = [\theta_1, \cdots, \theta_p]' \in \Theta \subset \mathbb{R}^p$ is the parameter vector. Because the inversion of a large dimensional covariance matrix could be laborious, we compute Σ_T^{-1} via the **Durbin-Levinson recursion**.

Parameters Estimation: The Durbin-Levinson recursion

According to Brockwell and Davis (1986), the best one-step ahead linear predictor $\hat{y}_{t+1|t}$ of y_{t+1} in terms of $y_{t+1-T}, \cdots, y_t$ is:

$$\hat{y}_{t+1|t} = \sum_{j=1}^{n} \varphi_{T,j} y_{t+1-j}$$

where the vector $\varphi_T = [\varphi_{T,1}, \cdots, \varphi_{T,T}]'$ is the solution of the Yule-Walker equation:

$$\begin{bmatrix} \varphi_{T,1} \\ \vdots \\ \varphi_{T,T} \end{bmatrix} = \Sigma_T^{-1} \begin{bmatrix} \gamma(1) \\ \vdots \\ \gamma(T) \end{bmatrix}$$

with Σ_T as the $T \times T$ covariance matrix of the process y_t. The notation $y_{t+1|t}$ means $E(y_{t+1}|\mathcal{I}_t)$ where $\mathcal{I}_t$ is the information set at time t.

Parameters Estimation: The Durbin-Levinson recursion

For $m = 2, \cdots, T$ and initialising $\varphi_{1,1} = \frac{\hat{\gamma}(1)}{\hat{\gamma}(0)}$ and $\nu_1 = \hat{\gamma}(0)\left(1 - \frac{\hat{\gamma}(1)^2}{\hat{\gamma}(0)^2}\right)$, the coefficients collected in φ_T can be computed recursively via the Durbin and Levinson recursion:

$$\varphi_{m,m} = \left(\hat{\gamma}(m) - \begin{bmatrix}\varphi_{m-1,1} & \cdots & \varphi_{m-1,m-1}\end{bmatrix} \begin{bmatrix}\hat{\gamma}(m-1) \\ \vdots \\ \hat{\gamma}(1)\end{bmatrix}\right) \nu_{m-1}^{-1}$$

$$\begin{bmatrix}\varphi_{m,1} \\ \vdots \\ \varphi_{m,m-1}\end{bmatrix} = \begin{bmatrix}\varphi_{m-1,1} \\ \vdots \\ \varphi_{m-1,m-1}\end{bmatrix} - \varphi_{m,m} \begin{bmatrix}\varphi_{m-1,m-1} \\ \vdots \\ \varphi_{m-1,1}\end{bmatrix}$$

$$\nu_m = \nu_{m-1}(1 - \varphi_{m,m}^2)$$

where $\varphi_{m,m}$ are the partial autocorrelations, $\hat{\gamma}(k)$ is the estimated ACF and ν_T is the one-step-ahead mean square prediction error.

Parameters Estimation: The Durbin-Levinson recursion

Through the Durbin-Levinson recursion, we can obtain the inverse covariance matrix via the following decomposition:

$$\Sigma_T^{-1} = C_T' D_T C_T$$

where $D_T = \text{diag}(\nu_0^{-1}, \nu_1^{-1}, \cdots, \nu_{T-1}^{-1})$ with $\nu_0 = \hat{\gamma}(0)$ and:

$$C_T = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\varphi_{1,1} & 1 & 0 & \cdots & 0 \\ -\varphi_{2,2} & -\varphi_{2,1} & 1 & \cdots & 0 \\ \vdots & \vdots & \ldots & \ddots & \vdots \\ -\varphi_{T-1,T-1} & -\varphi_{T-1,T-2} & -\varphi_{T-1,T-3} & \cdots & 1 \end{bmatrix}$$

The Whittle likelihood

The standard Whittle likelihood is an approximation of the exact log-likelihood where the covariance terms in the time domain are substituted by spectral terms in the frequency domain. It is given by the following function:

$$L_w(\theta) = -\frac{T}{4\pi}\int_{-\pi}^{\pi} \log f(\omega;\theta)d\omega - \frac{T}{4\pi}\int_{-\pi}^{\pi} \frac{I(\omega)}{f(\omega;\theta)}d\omega \tag{13}$$

where $f(\omega;\theta)$ is the spectral density and $I(\omega)$ is the periodogram defined as:

$$I(\omega) = \frac{1}{2\pi T}\left|\sum_{t=0}^{T-1} y_{t+1}e^{-i\omega t}\right|^2$$

The main advantage of the Whittle approach is that computations may be simplified considerably with respect to the exact likelihood estimation.

Parameters Estimation: Asymptotic theory

It follows from the standard asymptotic theory by Hannan (1973) that for the Whittle estimator $\hat{\boldsymbol{\theta}} = \text{argmax}_{\boldsymbol{\theta} \in \Theta}\{L_W(\boldsymbol{\theta})\}$ holds

Theorem

Let y_t be the FerARMA$(1, d_1, d_2, 1)$ process and let $\boldsymbol{\theta} \in \Theta \subset S$ be the true value of the parameters, where Θ is assumed to be compact, $\boldsymbol{\theta} = \begin{bmatrix} d_1 & d_2 & \phi & \psi & \sigma \end{bmatrix}'$ and $S = (0,1]^2 \times [0,1)^2 \times (0,\infty)$. It holds that

$$\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}$$

and

$$\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N(0, \boldsymbol{\Omega}(\boldsymbol{\theta})^{-1})$$

, where the matrix $\boldsymbol{\Omega}(\boldsymbol{\theta})$ has (i,j) elements

$$\frac{1}{4\pi}\int_{-\pi}^{\pi} \frac{\partial \log f(\lambda;\boldsymbol{\theta})}{\partial \theta_i} \frac{\partial \log f(\lambda;\boldsymbol{\theta})}{\partial \theta_j} d\lambda \ .$$

In the following, we implement a Monte Carlo experiment simulating 10^3 times a sample size of $T \in \{250, 500, 1000, 2500\}$ realizations from the Spolia process and, then, estimating the model both via the Whittle and exact likelihood. We consider different true values of $d, \phi \in \{0.2, 0.6, 0.9\}$, while the variance parameter is set to 1.
The process is simulated via the Durbin Levinson method (cf. Hosking, 1981) and initial guesses have been found via the Method of Moments considering the Spolia sample autocorrelation function, such that:

$$\phi_0 = 2\frac{\hat{\rho}_2}{\hat{\rho}_1} - \hat{\rho}_1 \quad , \quad d_0 = \frac{\hat{\rho}_1}{\phi_0}$$

where $\hat{\rho}_k$ is the sample autocorrelation function.

Simulation study: Exact vs Whittle estimation

Table: Bias from the estimation of 10^3 realizations of the Spoila process with sample size $T = 2500$. Absolute lower values are in bold text.

T=2500	WHITTLE			EXACT		
	Bias($\hat{d}$)	Bias($\hat{\phi}$)	Bias($\hat{\sigma}$)	Bias($\hat{d}$)	Bias($\hat{\phi}$)	Bias($\hat{\sigma}$)
$d = 0.2, \phi = 0.2$	**0.3090**	**0.1369**	-0.0013	0.3130	0.1372	-**0.0009**
$d = 0.2, \phi = 0.6$	0.0885	-0.0330	-0.0008	**0.0873**	-**0.0318**	-**0.0004**
$d = 0.2, \phi = 0.9$	**0.0089**	-**0.0202**	-0.0010	0.0093	-0.0212	-**0.0006**
$d = 0.6, \phi = 0.2$	0.0155	**0.0974**	-0.0008	**0.0134**	0.0986	-**0.0004**
$d = 0.6, \phi = 0.6$	0.0182	-0.0045	-0.0007	**0.0180**	-**0.0042**	-**0.0004**
$d = 0.6, \phi = 0.9$	**0.0013**	-**0.0018**	-0.0005	0.0021	-0.0022	-**0.0003**
$d = 0.9, \phi = 0.2$	-**0.1316**	**0.0738**	-0.0004	-0.1326	0.0741	-**0.0000**
$d = 0.9, \phi = 0.6$	**0.0028**	0.0003	-0.0007	**0.0028**	0.0006	-**0.0005**
$d = 0.9, \phi = 0.9$	**0.0017**	-**0.0016**	-**0.0003**	0.0024	-**0.0016**	-0.0009

Simulation study: FerARMA fits long memory

Table: Bias from the estimation of 10^3 realizations of the Spoila process with sample size $T = 2500$. Here, we approximate both long memory and unit root processes considering true values of the d and ϕ parameters close to one. Lower values are in bold text.

T=2500	WHITTLE			EXACT		
	Bias($\hat{d}$)	Bias($\hat{\phi}$)	Bias($\hat{\sigma}$)	Bias($\hat{d}$)	Bias($\hat{\phi}$)	Bias($\hat{\sigma}$)
$d = 0.20, \phi = 0.99$	**0.0012**	**-0.0056**	-0.0012	0.0021	-0.0066	**-0.0003**
$d = 0.60, \phi = 0.99$	**0.0034**	**-0.0022**	**0.0001**	0.0462	-0.0449	0.0138
$d = 0.99, \phi = 0.20$	**-0.1837**	**0.0755**	-0.0006	-0.1840	0.0757	**-0.0002**
$d = 0.99, \phi = 0.60$	-0.0291	**0.0148**	-0.0012	**-0.0290**	0.0150	**-0.0011**
$d = 0.99, \phi = 0.99$	**-0.0020**	**-0.0010**	0.0176	0.0099	-0.0016	**-0.0003**

Simulation study: FerARMA fits long memory

- Many studies in the literature (eg. Granger and Hyung, 2004 and Perron and Qu, 2010) that highlight the difficulty in distinguish between the long memory property generated by a short memory process with occasional shifts in the mean and the one generated a standard fractional long memory process.
- Let us investigate on this issue showing as the Spolia process is able to capture the long memory features generated by a WN with occasional breaks in the mean:

$$y_t = \varepsilon_t + \sum_{i=1}^{t} \eta_i b_i \tag{14}$$

where $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$, $\eta_t \sim N(0, \sigma_\eta^2)$, $E(\varepsilon_t \eta_t) = 0$ and b_i follows an i.i.d. binomial distribution, such that $b_i = 1$ with probability p $b_i = 0$ with probability $1 - p$.

Simulation study: FerARMA fits long memory

Table: Whittle estimates according to the Spolia and FN models in terms of Monte Carlo expectations with 10^3 repetitions, $T = 10^3$ and considering the WN with shifts in the mean as DGP. The value in brackets is the AIC.

Whittle estimates of the ϕ parameter

Model	$\sigma_\eta^2 \setminus p(pT)$	$p(pT) = 0.0025(5)$	$p(pT) = 0.005(10)$	$p(pT) = 0.01(20)$	$p(pT) = 0.05(100)$
FerAR$(1, d)$	$\sigma_\eta^2 = 0.01$	0.2655(-1658)	0.4088(-1643)	0.6324(-1630)	0.9843(-1554)
	$\sigma_\eta^2 = 0.05$	0.6136(-1623)	0.8386(-1591)	0.9646(-1557)	0.9988(-1418)
	$\sigma_\eta^2 = 0.1$	0.7762(-1598)	0.9500(-1558)	0.9905(-1501)	0.9990(-1327)
	$\sigma_\eta^2 = 0.2$	0.8943(-1560)	0.9797(-1508)	0.9979(-1435)	0.9991(-1220)

Whittle estimates of the d parameter

	$\sigma_\eta^2 \setminus p(pT)$	$p(pT) = 0.0025(5)$	$p(pT) = 0.005(10)$	$p(pT) = 0.01(20)$	$p(pT) = 0.05(100)$
	$\sigma_\eta^2 = 0.01$	0.1023(-1658)	0.0803(-1643)	0.0895(-1630)	0.1319(-1554)
	$\sigma_\eta^2 = 0.05$	0.0938(-1623)	0.1026(-1591)	0.1297(-1557)	0.2388(-1418)
	$\sigma_\eta^2 = 0.1$	0.1084(-1598)	0.1299(-1558)	0.1725(-1501)	0.2947(-1327)
	$\sigma_\eta^2 = 0.2$	0.1289(-1560)	0.1697(-1508)	0.2224(-1435)	0.3544(-1220)

Whittle estimates of the d_{LM} parameter

	$\sigma_\eta^2 \setminus p(pT)$	$p(pT) = 0.0025(5)$	$p(pT) = 0.005(10)$	$p(pT) = 0.01(20)$	$p(pT) = 0.05(100)$
FN(d_{LM})	$\sigma_\eta^2 = 0.01$	0.0211(-1661)	0.0339(-1646)	0.0544(-1634)	0.1301(-1556)
	$\sigma_\eta^2 = 0.05$	0.0565(-1626)	0.0903(-1594)	0.1272(-1559)	0.2389(-1422)
	$\sigma_\eta^2 = 0.1$	0.0864(-1601)	0.1256(-1561)	0.1719(-1505)	0.2955(-1330)
	$\sigma_\eta^2 = 0.2$	0.1190(-1563)	0.1684(-1512)	0.2220(-1439)	0.3573(-1220)

Simulation study: conclusion

- The Whittle likelihood leads to accurate parameters estimates if the sample size is large.
- Both the ϕ and d parameters affect the memory of the process and they could come into conflict with each other when either one or both of them are quite small. In this case, the Spolia process tends to be more similar to a white noise sequence.
- It should be clear that the Spolia process estimates are more accurate in the case of high persistent data.
- In approximating long memory under non stationarity ($d = 0.60$ and $\phi = 0.99$) the Whittle approach leads to very much more accurate parameters estimates with respect to the exact likelihood, because of the drawbacks of the Durbin-Levinson recursion in estimating the model in proximity of the non stationary region.
- The Spolia process approaches a FN as the size and numbers of breaks increase along with the persistence in the simulated data.

Testing long memory against FerARMA

- Let us address the issue whether the FerARMA model is the better model in a given situation or if a long memory specification should be preferred.
- Define the spectral density of the FerARMA specification, with parameter $\boldsymbol{\theta} = [d, \boldsymbol{\theta}_1, \phi]'$, as

$$f(\lambda; d, \boldsymbol{\theta}_1, \phi) = (1 - 2\cos(\lambda)\phi + \phi^2)^{-d} f_1(\lambda; \boldsymbol{\theta}_1), \quad \text{with} \quad \lambda \in [-\pi, \pi], \tag{15}$$

  where $f_1(\lambda; \boldsymbol{\theta}_1)$ is the bounded spectrum of a generic short memory stationary process.
- We test the null $H_0 : \phi = 1$ of a long memory specification against the alternative $H_1 : \phi < 1$ of a short memory FerARMA specification, via the Whittle likelihood ratio test:

$$LR = 2T(L_W(\hat{d}, \hat{\boldsymbol{\theta}}_1, \hat{\phi}) - L_W(\hat{d}, \hat{\boldsymbol{\theta}}_1, 1)), \tag{16}$$

Testing long memory against FerARMA

- It follows from Taniguchi and Amano (2009) that

Theorem

Let y_t be a FerARMA process with spectral density $f(\lambda; d, \boldsymbol{\theta}_1, \phi)$ defined in (15), then under the null hypothesis $H_0 : \phi = 1$, it holds that

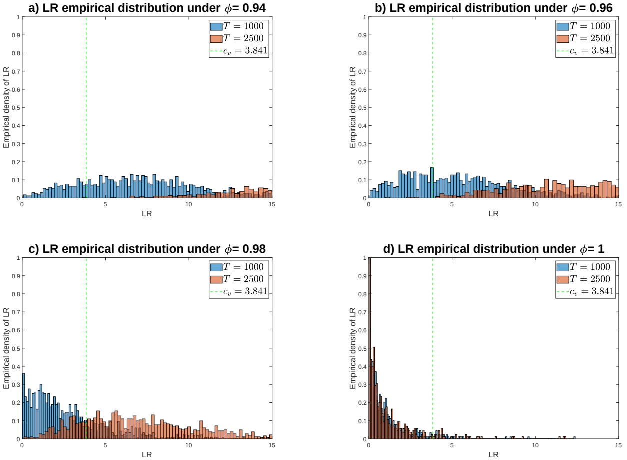
$$LR \xrightarrow{d} \chi_1^2,$$

where LR is defined in (16) and χ_1^2 denotes the chi-squared distribution with 1 degree of freedom.

- Now, consider 10^3 Monte Carlo repetitions where at each trial a sample size of $T \in \{1000, 2500\}$ realizations is generated from the Spolia process, then the LR statistic is computed, under the true values $\phi \in \{0.94, 0.96, 0.98, 1.00\}$, $d = 0.55$ and $\sigma = 1$.

Testing long memory against FerARMA

Figure: Empirical densities of the *LR* statistic. The dashed green line corresponds to $c_v = \Upsilon_1(0.95) = 3.8415$ where $\Upsilon_1(z)$ is the inverse cumulative χ^2_1 distribution s.t. if $LR > c_v$, then $H_0 : \phi = 1$ is rejected at the confidence level of 95%.



Empirical Illustration

- Let us move on an empirical illustration considering the problem of forecasting the yearly records of the tree ring and temperature reconstruction time series.
- Our purpose is to compare long memory models with respect to the FerARMA class

$$\text{FerAR}: \ (1-\phi L)^d y_t = \varepsilon_t \quad , \quad \text{FerARMA}: \ (1-L)^d y_t = (1-\psi L)\varepsilon_t$$

  in terms of forecasting performance on the climate series.
- As long memory models, we consider

$$\text{FN}: (1-L)^d y_t = \varepsilon_t, \qquad \text{FARI}: (1-L)^d (1-\varphi L) y_t = \varepsilon_t$$

$$\text{FIMA}: (1-L)^d y_t = (1-\psi L)\varepsilon_t, \quad \text{FerIMA}: (1-L)^d y_t = (1-\psi L)^d \varepsilon_t$$

  where $\varepsilon_t \sim WN(0, \sigma^2)$ and $\begin{bmatrix}\varphi & \psi\end{bmatrix}' \in (-1,1)^2$.

Empirical Illustration

Table: Whittle estimates of the parameters.

Temperature reconstruction	$\hat{d}$	$\hat{\sigma}^2$	$\hat{\phi}$	$\hat{\varphi}$	$\hat{\psi}$	$L_W(\hat{\boldsymbol{\theta}})$
FerAR	0.5058	0.3805	0.9743			2258.7
FerARMA	0.6781	0.3791	0.9409		0.1845	2263.4
FN	0.4857	0.3822				2255.1
FARI	0.5132	0.3821		-0.0418		2255.8
FIMA	0.5184	0.3820			0.0483	2255.9
FerIMA	0.5203	0.3820			0.0969	2255.9

Australian tree rings	$\hat{d}$	$\hat{\sigma}^2$	$\hat{\phi}$	$\hat{\varphi}$	$\hat{\psi}$	$L_W(\hat{\boldsymbol{\theta}})$
FerAR	0.5034	0.0203	0.9663			2242.9
FerARMA	0.5924	0.0203	0.9515		0.1061	2243.9
FN	0.4829	0.0205				2241.2
FARI	0.5011	0.0205		-0.0280		2241.4
FIMA	0.5051	0.0205			0.0335	2241.4
FerIMA	0.5074	0.0205			0.0723	2241.4

Arizona tree rings	$\hat{d}$	$\hat{\sigma}^2$	$\hat{\phi}$	$\hat{\varphi}$	$\hat{\psi}$	$L_W(\hat{\boldsymbol{\theta}})$
FerAR	0.5019	0.0222	0.9882			3336.5
FerARMA	0.5134	0.0222	0.9868		0.0156	3336.6
FN	0.4935	0.0222				3335.9
FARI	0.4908	0.0222		0.0042		3335.9
FIMA	0.4902	0.0222			-0.0052	3335.9
FerIMA	0.4897	0.0222			-0.0121	3335.9

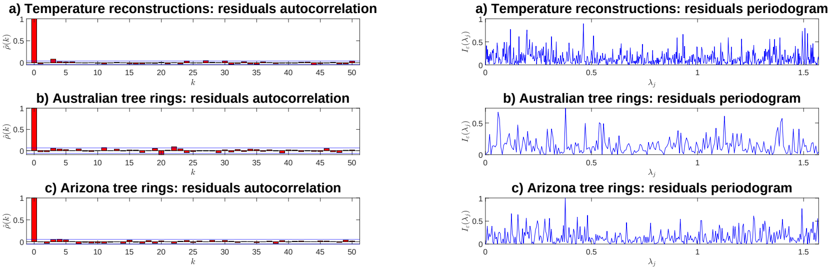
Table: Lagrange Multiplier test by Robinson (1994) according to the FN, FARI$(1, d)$, FIMA$(d, 1)$ and FerIMA$(d, 1)$ models. The test cannot reject the null of non stationarity under long memory ($H_0 : d = 0.5$) in all the cases.

	FN		FARI$(1, d, 1, 1)$		FIMA$(1, d, 1, 1)$		FerIMA$(1, d, 1, 1)$	
	p-value	*LM*	p-value	*LM*	p-value	*LM*	p-value	*LM*
Temperature reconstr.	0.7867	0.0733	0.3535	0.8608	0.3912	0.7351	0.3398	0.9110
Australian tree rings	0.4839	0.4900	0.8776	0.0237	0.7747	0.0819	0.9303	0.0077
Arizona tree rings	0.8999	0.0158	0.9263	0.0086	0.9133	0.0118	0.9335	0.0070

Table: Likelihood ratio test according to the FerAR$(1, d)$ and FerARMA$(d, 1, 1, 1)$ models. The test rejects the null hypothesis of long memory ($H_0 : \phi = 1$) for most of the series and models.

	FerAR		FerARMA	
	p-values	*LR*	p-values	*LR*
Temperature reconstr.	0.0071	7.2402	0.0001	15.0147
Australian tree rings	0.0652	3.3995	0.0259	4.9643
Arizona tree rings	0.2742	1.1954	0.2610	1.2636

Figure: Autocorrellation and periodogram of the residuals obtained by fitting the Spolia process on the temperature reconstructions, Australian and Arizona tree ring series.



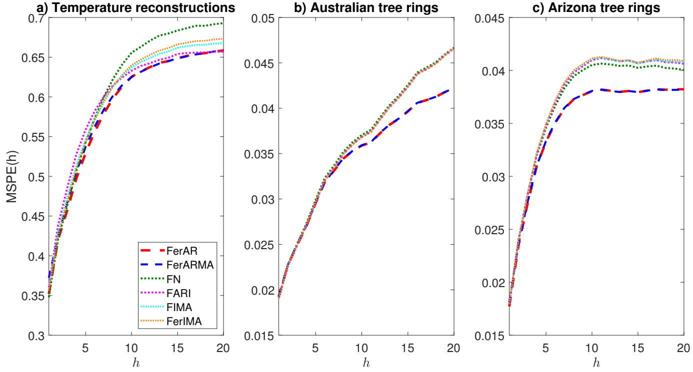
Empirical Illustration

- Let us illustrate that the FerARMA class is able to improve predictive accuracy in a recursive forecasting experiment with respect to the long memory specifications.
- Starting from $T_0 = 600$ we estimate the models using the first $T_0 + t$ observations, for $t = 0, 1, \cdots, T - T_0 - 1$, and then we predict $\hat{y}_{T_0+t+h|T_0+t}$, for $h = 1, 2, \cdots, 20$. So that the Mean Square Prediction Error (MSPE), given by

$$MSFE(h) = (T - T_0 - h + 1)^{-1} \sum_{t=T_0}^{T-h} (y_{t+h} - \hat{y}_{t+h|t})^2,$$

is then used to compare the forecasting performance of the various models.

Figure: MSPEs as a function of the prediction horizon *h* for the (a) Paleo-temperature reconstructions, (b) Australian pine tree rings and (c) Arizona tree ring series.



Summary

1. The FerARMA process is here proposed, in its short memory specification, as an alternative way to forecast highly persistent time series.
2. The main advantage concerns the stationarity condition, which holds also for values of $d > 0.5$, enabling for the use of the DL recursion in the computation of the best linear predictor. This is not possible in the long memory case under non stationarity.
3. The possibility of regulating the persistence in the data through two parameters (d and ϕ), likely provides more flexibility to the model in capturing the dependence structure of the data.
4. These peculiarities allow to the FerARMA processes to perform generally better with respect to its long memory counterparts in forecasting several kind of climate time series.

Discussion

1. In our opinion, the FerARMA class should be highly considered in empirical studies on datasets showing elevate persistence, but for which a long memory specification may not be the most appropriate.
2. For instance, in climate science the Spolia process may be considered as a possible choice in modelling the **climate noise**, which arises from the sampling variability of high-frequency climate variables (see Leith, 1973).
3. A common choice in the climate literature is to model the climate noise via an AR(1) (*red noise*) process.
4. However, there is also debate whether the FN (*pink noise*) specification may be a better choice, since the climate noise could be related to the interaction of several variables, which aggregation can, according to Granger (1980), generates long memory.
5. In such context, the Spolia process may be introduced as a possible alternative.

Merci Beaucoup pour Votre Attention!