

Long memory and structural break analysis of environmental time series^(*)

Memoria lunga e break strutturali nelle serie storiche ambientali

Carmela Cappelli and Angela D'Elia

Dipartimento di Scienze Statistiche, Università di Napoli Federico II

e-mail: carmela.cappelli@unina.it

Riassunto: Nell'ultimo decennio un crescente interesse è stato prestato alla difficile discriminazione tra processi a memoria lunga e processi con *break* strutturali occasionali nel livello medio. Numerosi studi mostrano che la presenza di *break* strutturali può causare un decadimento molto lento nella funzione di autocorrelazione, caratteristica tipica dei processi a lunga memoria; viceversa, quando il processo generatore dei dati (DGP) è di tipo integrato frazionario si rischia di identificare dei *break* spurii. In questo lavoro viene proposta una strategia di analisi che può fornire indicazioni circa la presenza reale o spuria di *break* strutturali o di lunga memoria. La proposta viene illustrata mediante un'applicazione allo studio della serie delle portate annuali del fiume Nilo.

Keywords: atheoretical regression trees, long memory, structural breaks.

1. Introduction

In the last decade lot of interest, both in the econometric and statistical literature, has been paid to the issue of confusing long memory and occasional structural breaks in mean (see, among the others, Diebold and Inoue, 2001; Granger and Hyung, 2004; Smith, 2005).

Indeed, there is evidence that a stationary short memory process that encounters occasional structural breaks in mean can show a slow rate of decay in the autocorrelation function and other properties of $I(d)$ processes, where d can be a fraction. On the other hand, when the data generating process is an integrated or fractionally integrated process (with no breaks), several breaks can be detected spuriously: in fact, there's a relation between the number of breaks and the value of d .

The purpose of this paper is to propose a simple strategy that an applied researcher can carry out as a first check before proceeding with further analysis of the data. In particular, while most of the authors focussed on economic/financial series, we are going to consider the problem of long memory and/or occasional breaks detection in environmental data. At this aim the proposed strategy is applied to analyze the annual flow series (1871-1970) of the Nile river, since the presence of long memory in hydrological time series is a well known phenomenon and, on the other hand, the presence of structural breaks in these series represents a relevant environmental issue.

^(*) The authors acknowledge financial support from the Dipartimento di Scienze Statistiche - Università di Napoli Federico II. Although this is a joint research, C. Cappelli wrote sections 1 and 2.2, and A. D'Elia wrote sections 2.1 and 3.

2. Long memory and/or structural breaks? A detection strategy

In this section we briefly recall the fractional integrated model (that is the most popular long memory model); then, we introduce a method, recently proposed by Cappelli and Reale (2005), for detecting structural breaks in order to illustrate a strategy that helps distinguishing between long memory and occasional breaks.

2.1. The ARFIMA(0, d , 0) model

A process $Y_t \sim ARFIMA(0, d, 0)$ is defined as: $(1 - B)^d Y_t = a_t$, where a_t is a Gaussian White Noise: $a_t \sim WN(0, \sigma^2)$, B is the backward operator, such that $B^k Y_t = Y_{t-k}$, $k = 0, \pm 1, \dots$, and

$$(1 - B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-1)^k B^k. \quad (1)$$

Moreover $d \in (-0.5, 0.5)$: thus, Y_t is a stationary and invertible stochastic process. If $0 < d < 0.5$, then Y_t is a long memory process, whose autocorrelation function decays to zero at an hyperbolic rate:

$$\rho(k) \sim b k^{2d-1}, \quad k \rightarrow \infty, \quad (2)$$

where b is a constant not depending on k ; in such case $\sum_{k=1}^{\infty} \rho_k = +\infty$, and the spectral density at the origin is $f(0) = \frac{1}{2\pi} (1 + 2 \sum_{k=1}^{\infty} \rho_k) = +\infty$. On the contrary, if $-0.5 < d < 0$, we get an anti-persistent process, and $f(0) = 0$ (Beran, 1994).

As discussed in Granger and Hyung (2004), time series with finite occasional breaks may show several of the above long memory properties, both in time and frequency domain.

2.2. Atheoretical Regression Trees for the detection of structural breaks

Let y_t ($t = 1, \dots, T$) be an observed series whose structural breaks in mean are to be detected. The problem resorts to estimate the set of break-dates $(t_1, \dots, t_g, \dots, t_{G-1})$ that define a partition of the series $P(G) = \{(1, \dots, t_1), \dots, (t_{g-1} + 1, \dots, t_g), \dots, (t_{G-1} + 1, \dots, t_G = T)\}$, into G homogeneous intervals such that $\mu_g \neq \mu_{g+1}$. A common criterion is based on the least squares principle: the estimated break-dates are associated with the partition $P^*(G)$ such that

$$P^*(G) = \arg \min_{P(G)} WSS_{y|P(G)}, \quad (3)$$

where $WSS_{y|P(G)} = \sum_{g=1}^G \sum_{t=t_{g-1}+1}^{t_g} (y_t - \mu_g)^2$.

As shown in Cappelli and Reale (2005), the task of estimating the break-dates can be accomplished within the framework of Least Squares Regression Trees (LSRT) that are based on the same objective function as the (3). In particular, for the purpose of locating the break-dates, an arbitrary ascending (or descending) sequence of completely ordered numbers k is employed: for sake of simplicity, we take $k = 1, 2, \dots, T$. Regressing y_t on k , i.e. using k as unique covariate in LSRT, leads to creating and checking at any node all possible binary contiguous partitions of the y_t in the given node. The procedure,

called ART (Atheoretical Regression Trees), provides a partition of y_t into homogeneous subseries, represented by a tree diagram whose split values indicate candidate break-dates. The final set of breaks is selected by means of pruning and model selection criteria.

As said before, time series with occasional structural breaks can generate a strong persistence in the autocorrelation function, i.e. the observed long memory behavior can be due to neglected structural breaks and, on the contrary, long memory processes may cause breaks to be detected spuriously.

In order to have an insight on "what is what", we propose to fit long memory and structural break models separately. In case both provide plausible explanation of the DGP of the data at hand, the long memory and structural break analysis are repeated on the series break-free and on the filtered series, respectively. Then, the presence of breaks in the filtered series and/or of long memory behavior in the break-free series could indicate whether the series shows real breaks and long memory. A further insight can be obtained by estimating the difference parameter on the subseries identified by splitting the original series according to the estimated break-dates: in case of erroneous long memory identification, the subseries are expected to show short memory.

3. Application

In order to show how the above described strategy can help in dealing with the discrimination between long memory and occasional breaks, we have analyzed the series of the annual flows of the Nile river at Ashwan (1871-1970), on which we have identified an ARFIMA(0, d , 0) model.

The maximum likelihood estimate of the fractional parameter d was obtained by means of a modified version of the Whittle's profile log-likelihood, as discussed in D'Elia and Piccolo (2003), in order to consider the possibility of a normalizing transformation; in fact, it turned out that the original series does not need any preliminary transformation ($\hat{\lambda} = 0.906$). The presence of long memory is highly significant (Table 1), as we would also expect by the inspection of the estimated autocorrelation function (Figure 1).

At the same time, we have detected the breaks using the ART method, setting a minimum subseries length of 15 observations. The resulting tree shows 4 breaks, but only one break is selected by cross validation and BIC: indeed, this break occurred in 1898 when the first dam was built.

Table 1 reports the estimates of the difference parameter d and their standard error (in brackets) for the original series and the series break-free whereas, in Figure 1 both series and their corresponding estimated autocorrelation functions are depicted.

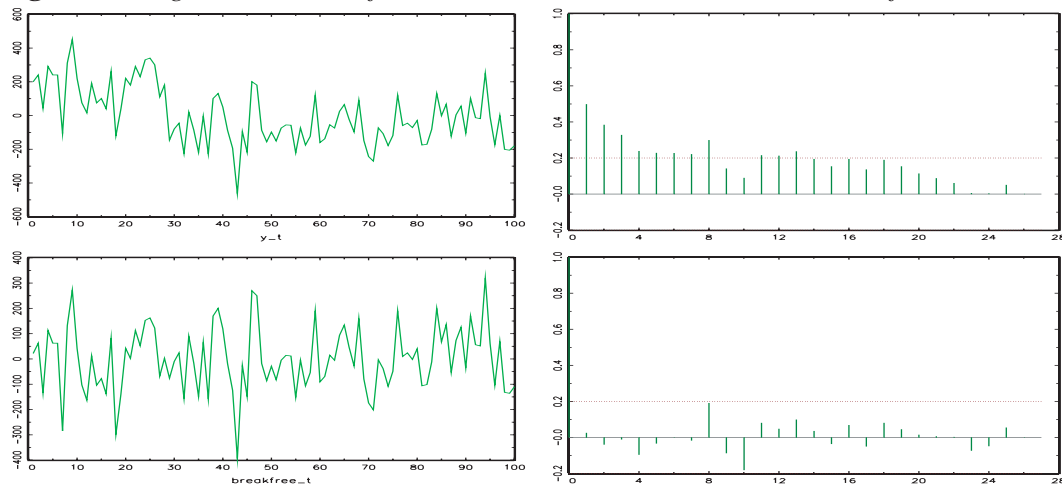
Table 1: *Estimated d on the original series and on the break-free series*

$\hat{d}$	No. of breaks	date of occurrence
0.381 (0.078)	0	
0.061 (0.078)	1	1898

We see that for the break-free series the fractional integrating parameter is not significant, and the estimated autocorrelation function does not exhibit any long memory

pattern. On the other hand, if we consider the filtered series (that is, without the long memory), again we find -using the ART- a break (in the neighborhood of year 1898). Then, we can conclude that, in the original series, the observed evidence of long memory is entirely due to the structural break of 1898 which, on the contrary, really characterizes the flows series.

Figure 1: *Original and break-free series, with estimated autocorrelation functions.*



Furthermore, the estimate of the fractional parameter d on the subseries⁽¹⁾ obtained by splitting at the break-date, confirms that there is not significant long memory in the DGP, being respectively: $\hat{d}_I = 0.132$ ($s.e. = 0.147$) and $\hat{d}_{II} = 0.095$ ($s.e. = 0.092$).

Of course, the same strategy can be applied for the analysis of other hydrological time series. In particular, we are currently analyzing a time series where a high number of breaks can be detected but, it seems that they are partly due to the non stationarity of the underlying long memory process.

References

- Beran J. (1994) *Statistics for Long-Memory Processes*, Chapman & Hall, New York.
- Cappelli, C., Reale, M. (2005), Dating multiple structural breaks occurring at unknown dates via Atheoretical Regression Trees, in: *S.Co. 2005: Modelli Complessi e Metodi Computazionali Intensivi per la Stima e la Previsione*, Provasi C. (Ed), Cleup, 479-484.
- D'Elia, A., Piccolo, D. (2003), Maximum likelihood estimation of ARFIMA models with a Box-Cox transformation, *Statistical Methods & Applications*, 12, 259-275.
- Diebold, F.X., Inoue, A. (2001), Long memory and regime switching, *Journal of Econometrics*, 105, 131-159.
- Granger, C.W.J., Hyung, N. (2004), Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns, *Journal of Empirical Finance*, 11, 399-421.
- Smith, A. (2005), Level shifts and the illusion of long memory in economic time series, *Journal of Business & Economic Statistics*, 23, 321-335.

⁽¹⁾ However, the estimate on the first one deserves caution due to the shortness of the subseries.