Nonparametric modeling of the second order structure of processes with time–varying memory

Georg M. Görg[1], Dana Draghicescu[2]
Vienna University of Technology, Department of Econometrics and System Theory
georg.goerg@gmail.com[1]
Hunter College of the City University of New York, Department of Mathematics and Statistics
dana.draghicescu@cuny.hunter.edu[2]

Abstract

In this paper we consider nonstationary time series, characterized by time-varying memory. We model the temporal changes of the memory parameter (or Hurst parameter) in the time domain, by using a moving window approach. Our estimation procedure incorporates a data-driven scheme for optimal bandwidth selection. The proposed methodology is illustrated on hydrological and financial data sets.

KEY WORDS: time–varying memory, Hurst parameter, fractional differencing, ARFIMA, nonstationary time series.

1. Introduction

The study of long-memory time series has been the object of intense research in the past decades. A second-order stationary process X_t is said to be long-range dependent (or to have long memory) if $Cov(X_t, X_{t+k}) \sim C(k)|k|^{2H-2}$, where C is a slowly varying function at infinity, and $H \in \mathbb{R}\backslash\{0\}$ is the so-called Hurst or memory parameter. A comprehensive study on long memory is given in [1], including detailed descriptions of several estimators of H, in both the time and frequency domains. Rigorous theoretical considerations on the estimation of H in the frequency domain are given in [7, 8].

More recently the study of nonstationary time series has received much attention. This interest is both theoretical and applied, a main motivation being the observed temporal nonstationarity in many real life processes. One way to describe nonstationarity is by allowing the memory parameter to change with time. This seems to be a realistic assumption, as illustrated in [2], where a test for a change of the long-memory parameter is proposed. For the estimation of time-dependent memory parameter H [3] use the R/S method, and [10] propose a statistical tool for the local analysis of self-similarity (LASS). [4] argue that detrending moving average (DMA) provides more accurate results than the standard wavelet transform of the higher order power spectrum technique.

In financial framework, one can interpret H as a measure of memory length, or influence of the past on investors in the stock market, since H contributes to the autocorrelation funcion of the series. By imposing a constant Hurst parameter on such processes, investors would always take the same amount of past information into account when making their investment decisions. This view of the world is neither reasonable, nor supported by empirical data. In this paper we propose a nonparametric method for the estimation of time-varying memory parameter in the time domain.

Our main contribution is that of providing an automatic scheme for optimal bandwidth selection, based on minimization of the variance of $\widehat{H}(t)$ for each t. The paper is organized as follows. Section 2 gives the theoretical framework, followed by a detailed description of a data-driven scheme for optimal bandwidth selection in Section 3. Two applications on hydrological and financial data sets are presented in Section 4. Concluding remarks are made in Section 5.

2. Theoretical framework

Let $x_1, \ldots, x_T$ be observations from an underlying process x_t with common mean $\mu = \mathbb{E}(x_t)$, variance $\sigma_x^2 = \mathbb{V}x_t = \mathbb{E}\left(x_t - \mu\right)^2$, and correlation function $\rho(i,j) = corr(x_i, x_j)$. Consider the sample mean estimator $\overline{x}_T := \frac{1}{T}\sum_{t=1}^{T} x_t$. It holds

$$\begin{aligned} \mathbb{V}(\overline{x}_T) &= \frac{\sigma_x^2}{T}\left(1 + \delta_T(\rho)\right), \qquad (1) \\ \text{with } \delta_T(\rho) &= \frac{1}{T}\sum_{i \neq j} corr(x_i, x_j). \end{aligned}$$

If x_t is i.i.d. then $\delta_T(\rho) = 0$ and we have the classic result of i.i.d. statistics, that the variance of the sample mean decays to zero with a rate of T^{-1}. Of course, we do not expect that observations in time are independent, thus in general $\delta_T(\rho) \neq 0$. For a covariance stationary process, the correlations $\rho(i,j)$ solely depend on the time lag $|i-j|$, thus $\delta_T(\rho)$ can be simplified to $\delta_T(\rho) = 2\sum_{k=1}^{T-1}(1 - \frac{k}{T})\rho(k)$, with $\rho(k) = corr(x_t, x_{t-k})$.

If $\lim_{T\to\infty} \delta_T(\rho) = c \in \mathbb{R}$ then the variance of the sample mean still goes to zero with rate T^{-1}. However, in practice one often finds a slower convergence rate, i.e.

$\mathbb{V}(\overline{x}_T) = o(T^{-\alpha})$ with $-1 < \alpha < 0$, which is in fact one definition for persistent long range dependence with long memory parameter $d = \frac{1+\alpha}{2}$ $(H = d + \frac{1}{2})$.[1]

We model the variance decay explicitly as $\mathbb{V}(\overline{x}_t) = \sigma_x^2 t^\alpha e^\varepsilon$, with standard assumptions on the error process ε (namely zero mean and common finite variance σ_ε^2). By taking logs we get

$$\log \mathbb{V}(\overline{x}_t) = \log \sigma_x^2 + \alpha \log t + \varepsilon. \quad (2)$$

So an estimate for α is given by the OLS slope estimator of

$$\begin{aligned} \log S_t^2 &= \log c + \alpha \log t + \varepsilon, \quad (3) \\ t &= t_{min}, t_{min}+1, \ldots, t_{max} < T, \end{aligned}$$

where S_t^2 is the sample variance of all calculated rolling sample means, with $t = t_{min}, \ldots, t_{max}$ observations per sampling (see Section 3 for details).

Typically, the points $\left(\log t, \log S_t^2\right)$ are scattered around a line with slope α. For short memory processes $\widehat{\alpha} \approx -1$, whereas for long memory in the data $\widehat{\alpha} \neq -1$.[2]

Table 1: Characterization of time series memory: Relation between α, H, and d

type of memory	α	$H = 1 + \frac{\alpha}{2}$	$d = H - \frac{1}{2}$
short memory	$= -1$	$= \frac{1}{2}$	$= 0$
anti-persistent	< -1	$< \frac{1}{2}$	< 0
persistent	> -1	$> \frac{1}{2}$	> 0

As mentioned in the Introduction, our assumption is that H, and consequently α, may change with time. A moving window approach is used in order to capture these possible changes. Thus, at any time point t, we only use the observations in a centered window of length $2b+1$. As with all nonparametric techniques, the choice of optimal window width (or, equivalently, of b) is crucial. In the next section we propose an automatic procedure for the choice of b_{opt}, based on minimizing $\mathbb{V}(\widehat{\alpha})$.

3. A data-driven scheme for optimal bandwidth selection

In order to get data for the estimation of $\mathbb{V}(\overline{x}_t)$, we use the method described in [1], Chapter 4, consisting of the following steps:

1. Let k_{min} and k_{max} be chosen constants, and let $2b_{max}+1$ be the maximum number of observations we base the estimations on. We estimate $\widehat{\alpha}(t)$ (and thus $\widehat{H}(t)$ as well) with data from the symmetric interval around X_t: $[X_{t-b_{max}}, X_{t+b_{max}}]$.
2. For different integers k in $(2 \leq) k_{min} \leq k \leq k_{max} (\leq \frac{2b+1}{2})$, and a sufficient number of subseries m_k of length k, we compute the sample means $\overline{x}_1(k), \ldots, \overline{x}_{m_k}(k)$ and the overall mean $\overline{x}(k) = \frac{1}{m_k}\sum_{j=1}^{m_k} \overline{x}_j(k)$.
3. For each k, take the sample variance of the sample mean $\overline{x}_j(k), j = 1, \ldots, m_k$ as $s^2(k) = \frac{1}{m_k - 1}\sum_{j=1}^{m_k}(\overline{x}_j(k) - \overline{x}(k))^2$.

Note that m_k is the number of intervals (boxes) needed to estimate the mean at a subsample size of k. We set $m_k := [\frac{2d+1}{k}]$ for each k, where $[x]$ denotes the integer part of x. Thus we get m_k non-overlapping boxes with length k, that will use (almost) all the observations in the overall sample.

The parameter k_{min} determines the minimum subsample size of x_k. If k_{min} is too small, the estimator will be very erratic, leading to a high variance, as well as biased estimates of the slope due to possible short term dependence in the data. The parameter k_{max} determines the maximum subsample size of the x_k. If k_{max} is too large, m_k will become small and similar problems as with k_{min} arise.

As a rule of thumb, we set $k_{min} = (2d+1)^{\frac{1}{3}}$ and $k_{max} = \frac{2d+1}{k.min} = (2d+1)^{\frac{2}{3}}$, based on the following heuristic reasons. Both k_{min} and k_{max} should get bigger, the bigger the number of observations gets. Moreover, if we have k_{min} observations for every estimation of the mean, we should also have at least k_{min} observations to estimate the variance of the sample mean with the maximum number of observations $k_{max} = \frac{2d+1}{k_{min}}$, so that we get k_{min} boxes, with each box containing k_{max} observations. Finally, the exponent $\frac{1}{3}$ is chosen both as a good approximation of what is reasonable given the length of the interval $2d+1$, and mathematical convenience. The optimal window length is then selected as

$$b_{opt} = \underset{d,\, k_{min}(d),\, k_{max}(d),\, m_k(d)}{\operatorname{argmin}} \mathbb{V}(\widehat{\alpha}), \quad (4)$$

and is obtained by numerical optimization on a grid.

Pseudo-code for numerical optimization

Step 1 Start with x_j for $j = b_{max} + 1$.

Step 2 Compute $\mathbb{V}(\hat{\alpha})$ for different values of d.

Step 3 Save the d for which $\mathbb{V}(\hat{\alpha})$ is minimal.

Step 4 If $j < T - (b_{max} + 1)$, j=j+1 and go to Step 2; otherwise stop.

Since we take symmetric intervals around X_t, the maximal bandwidth we can use is determined by how close we get to the edges of the time series. For instance, suppose we observe $\{X_t\}_{t=1}^{T=3000}$, and start with the point x_{700}. In this case b may not exceed 699, because we do not have any data prior to x_1 (the same argument holds at the

[1] The relationship between the convergence rate α, the Hurst parameter H, and the memory parameter d is given in Table 1.

[2] In practice on almost exclusively find persistent long memory processes, i.e. processes with Hurst parameter $H > \frac{1}{2}$.

end of the series). We therefore face a tradeoff between estimating $\widehat{\alpha}_t$ for as many time points t as possible, but on the other hand we want to have large bandwidths to detect long memory. Another problem is that the objective function takes its minimum value almost every time at $b = b_{max}$, if b_{max} is too low from the start.

4. Applications

In this section we illustrate the methodology previously described on hydrological and financial data. Again, our goal is to accurately estimate the Hurst parameter $\widehat{H}(t)$ for every point t, by choosing an optimal b that minimizes $\mathbb{V}(\hat{\alpha})$. We denote by $\widehat{\alpha}_{opt}$ and $\widehat{\alpha}_{max}$ the estimators of the slope α in (3) computed with b_{opt} and b_{max}, respectively. The same notation holds for $\widehat{H}_{opt}$ and $\widehat{H}_{max}$ (relationship given in Table 1). Additionally, we compute the ratio $\frac{\mathbb{V}(\hat{\alpha}_{opt})}{\mathbb{V}(\hat{\alpha}_{max})}$, for both time series, in order to assess the efficiency of our method.

4.1 Riverflow data

We look at data for downstream riverflow velocity (in m/s) sampled with 50 Hz over a period of one minute [9]. Having $T = 3000$ observations, we set $b_{max} = 699$. Differencing the data is not necessary, because several unit-root tests (KPSS, ADF, PP) indicate stationarity of the process.

Note the hyperbolical decay of the autocorrelation function in Figure 1, with significant lags up to 3 seconds (150 lags), indicating long memory behaviour of the process. The estimated mean of $\widehat{H}_{opt}(t)$ equals 0.73. We can also see that H is indeed time-varying (values range from $0.60 - 0.80$). Nevertheless $\widehat{H}_{opt}(t) \geq \frac{1}{2}$ for all t, meaning that this series is a time-varying, but always persistent long memory process. In Table 2 we can see that with our procedure, we have overall lowered the variance by 24%.

4.2 Financial data

Here we study EUR/CHF exchange rates with a set of 2040 daily returns from January 4, 1999 to November 6, 2006 ([5]). Here we set $b_{max} = 599$.

Considering the efficient market hypothesis, returns should be white noise (short memory). Hence we should get a value of a $\widehat{H}_t \approx 0.5$.

Figure 2 shows that $\widehat{H}_{opt}(t)$ is not very different to $\widehat{H}_{max}(t)$. Looking at the mean of $\widehat{H}_{opt}(t)$, that equals 0.53, we can not clearly infer an overall long term persistence. Again, $\widehat{H}_{opt}(t)$ is time-varying (values range from $0.43 - 0.62$), although in this case the variation could occur from random effects in the data. Nevertheless we observe the interesting effect that the process changes back and forth between anti-persistent long memory, short memory (possibly white noise) and persistent long memory. Although $\widehat{H}_{opt}(t) \approx \widehat{H}_{max}(t)$ we get a mean optimization gain of 18% (see Table 2).

Table 2: Summary statistics of $\frac{\mathbb{V}(\hat{\alpha}_{opt})}{\mathbb{V}(\hat{\alpha}_{max})}$

	Riverflow	EUR/CHF returns
Max.	1.00	1.00
3rd Qu.	0.90	0.93
Mean	**0.76**	**0.82**
Median	0.81	0.82
1st Qu.	0.64	0.72
Min.	0.23	0.47

5. Discussion

Estimating the variance of a process correctly is important in order to get accurate confidence intervals for forecasting. For instance, the widely used (short memory) ARMA models may underestimate the long-term dependence and the variance of the process, thus yielding too narrow confidence bands for their forecasts. On the other hand, ARIMA models are too general to give reasible forecasts and confidence bands. The method previously described might provide better insight.

This is an ongoing project (see [6]), detailed theoretical considerations and assessment of the corresponding forecasts will be addressed elsewhere. Our empirical study proved that the time-varying estimator that we propose is fast and accurate. It could be used as an exploratory, visual tool for the analysis of time-dependent data in many applied fields. An open problem that is worth investigating concerns the assumptions that need to be made on the function $H(t)$, in order to get *good* asymptotic properties. As seen in Figures 1 and 2, there are situations when the changes of H in time are quite abrupt, and where an additional smoothing might be necessary.

Acknowledgement. This research was conducted during Fall 2006, when Georg M. Görg visited the Department of Mathematics, City College, CUNY as an exchange student.

References

[1] Jan Beran, *Statistics for long-memory processes*, Chapman & Hall, 1994.

[2] Jan Beran and N. Terrin, *Testing for change of the long-memory parameter*, Biometrika **83** (1996), no. 3, 627 – 638.

[3] D. O. Cajueiro and B. M. Tabak, *Testing for fractional dynamics in the brazilian term structure of interest rates*, Tech. report, Banco Central do Brazil, 2004.

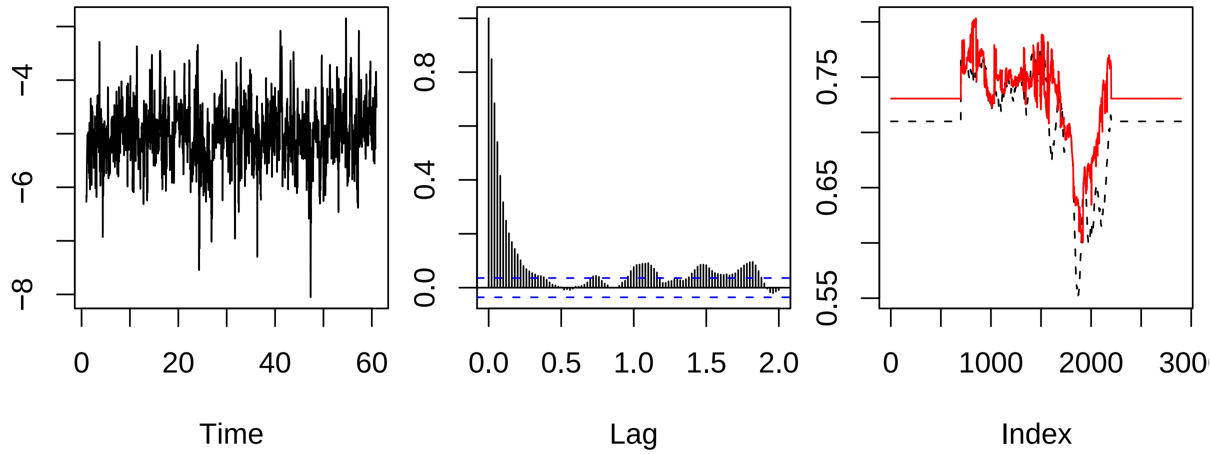

Figure 1: Downstream riverflow velocity in m/s (50 Hz, 1 minute) (left), corresponding autocorrelation function (center), and $\widehat{H}_{max}(t)$ (black) versus $\widehat{H}_{opt}(t)$ (red) with $b_{max} = 699$ (right).

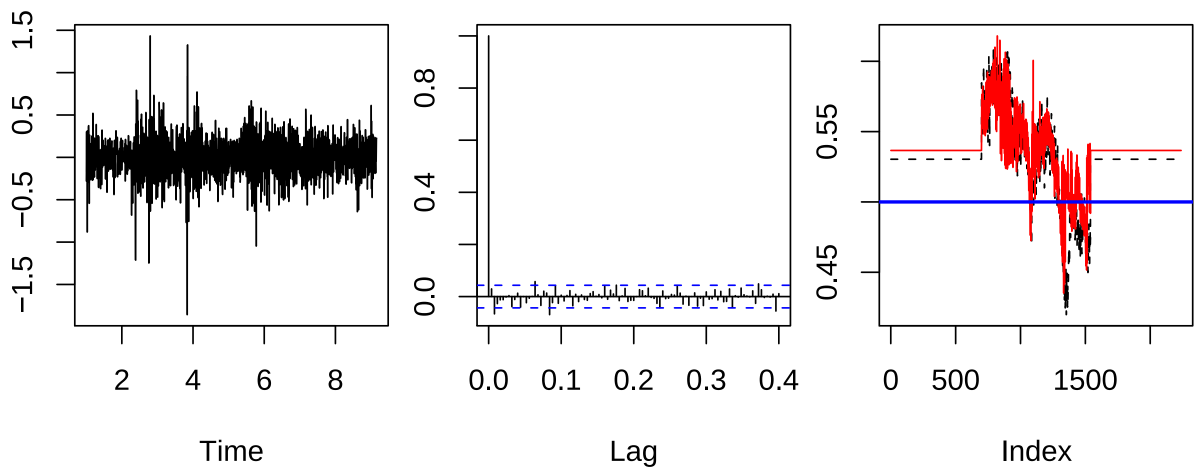

Figure 2: Daily log returns of EUR/CHF exchange rate, January 4, 1999 – November 6, 2006 (left), corresponding autocorrelation function (center), and $\widehat{H}_{max}(t)$ (black) versus $\widehat{H}_{opt}(t)$ (red) with $b_{max} = 599$ (right). Blue horizontal line corresponds to the threshold value $H = 0.5$.

[4] A. Carbone, G. Castelli, and H. E. Stanley, *Time-dependent Hurst exponent in financial time series*, Physica A **344** (2004), 267 – 271.

[5] European Central Bank, *Statistical Data Warehouse: EUR/CHF exchange rate data*, Available at: http://sdw.ecb.int, 2006.

[6] Georg M. Görg, *Long Memory versus Structural Breaks: A Time–Varying Memory Approach*, Master's thesis, Vienna University of Technology, Austria, 2007, Adviser: Manfred Deistler.

[7] Peter M. Robinson, *Gaussian semiparametric estimation of long range dependence*, The Annals of Statistics **23** (1995), no. 5, 1630 – 1661.

[8] ______, *Log-periodogram regression of time series with long range dependence*, The Annals of Statistics **23** (1995), no. 3, 1048 – 1072.

[9] Gabriel Singer, Tom Battin, Iris Hödl, K. Besemer, and E. Sollböck, *The lunzerinnen. Introducing a new dimension of experimental research in stream ecology. Acoustic Doppler Velocimeter Data from experimental streamside flumes.*, Unpublished, 2006.

[10] S. Stoev, M. S. Taqqu, C. Park, G. Michailidis, and J. S. Marron, *LASS: a tool for the local analysis of self-similarity*, preprint, 2004.