# WAVELETS, LONG MEMORY AND BOOTSTRAP - AN APPROACH TO DETECTING DISCONTINUITIES

*Jan Beran* and Yevgen Shumeyko***
*Department of Mathematics and Statistics*
*University of Konstanz*
*78457 Konstanz, Germany*
*E-mail: jan.beran@uni-konstanz.de*; yevgen.shumeyko@uni-konstanz.de***

## 1. Introduction

*Weakly stationary time series are defined by a slow decay of autocorrelations and a pole of the spectral density at the origin. Sample paths tend to exhibit local spurious trends and cycles of varying length and magnitude. Visual separation of stationary long-memory components and deterministic trend functions is difficult. Statistical methods such as kernel and local polynomial smoothing and wavelet thresholding have been developed however in the literature for this purpose (see e.g. Hall and Hart 1990, Csörgö and Mielniczuk 1995, Ray and Tsay 1997, Robinson 1997, Beran and Feng 2002a,b,c, Wang 1996, Johnstone and Silverman 1997, Yang 2001, Li and Xiao 2007, Kulik and Raimondo 2009, Beran and Shumeyko 2011a; also see Ghosh and Draghicescu 2002a,b for quantile smoothing). For kernel and local polynomial trend estimation, data driven iterative algorithms are available (Ray and Tsay 1997, Beran and Feng 2002a,b, Ghosh and Draghicescu 2002b). For wavelet thresholding, most results deal with the minimax approach. Often, minimax solutions are not suitable for a concrete data analysis. Beran and Shumeyko (2011a) therefore suggest a data driven wavelet method and derive asymptotically optimal tuning parameters (also see Li and Xiao 2007 for related results). This is discussed in the next section (section 2). The derivation of the asymptotic mean squared error leads to a natural decomposition of the trend estimator into a smooth and a discontinuous component. An application of this decomposition is testing for discontinuities in the trend function (Beran and Shumeyko 2011b). This is discussed in section 3 below. To obtain a method that is applicable to relatively short time series, a bootstrap statistic and an algorithm for computing its finite sample distribution are defined. The blockwise approach used here is similar to Lahiri (1993) (also compare with Percival et al. 2000). The validity and consistency of the test procedure is shown under the assumption of Gaussian residuals. A generalization to subordinated processes is possible but not pursued here in detail. For literature on structural breaks in the long-memory context see for example references in Sibbertsen (2004) and Banerjee and Urgab (2005).*

## 2. Data adaptive wavelet estimation

*Suppose we observe*

$$Y_i = g(t_i) + \xi_i \tag{1}$$

*($i = 1, 2, ..., n$) with $t_i = i/n$, $g \in L^2([0,1])$ and $\xi_i$ a Gaussian zero mean stationary process with autocovariances of the form*

$$\gamma(k) = E(\xi_i \xi_{i+k}) \underset{k\to\infty}{\sim} C_\gamma |k|^{-\alpha}, \tag{2}$$

*$\alpha \in (0,1)$, $C_\gamma > 0$, and spectral density*

$$f(\lambda) = (2\pi)^{-1} \sum \gamma(k) \exp(-ik\lambda) \underset{\lambda\to 0}{\sim} C_f |\lambda|^{\alpha-1}$$

*(see e.g. Beran 1994, Samorodnitsky 2007). Furthermore let* $\phi(t)$ *and* $\psi(t)$ *be continuous piecewise differentiable father and mother wavelets respectively, with compact support* $[0, N]$ *(for some* $N \in \mathbb{N}$*) such that*

$$(3) \qquad \int_0^N \phi(t)dt = \int_0^N \phi^2(t)dt = \int_0^N \psi^2(t)dt = 1,$$

$$(4) \qquad \psi(0) = \psi(N) = 0$$

*and, for any* $J \geq 0$*, the system* $\{\phi_{Jk}, \psi_{jk}, k \in \mathbb{Z}, j \geq 0\}$ *with*

$$\psi_{jk}(t) = N^{1/2} 2^{(J+j)/2} \psi(N2^{J+j}t - k), \quad \phi_{Jk}(t) = N^{1/2} 2^{J/2} \phi(N2^J t - k),$$

*is an orthonormal basis in* $L^2(\mathbb{R})$*. The number of vanishing moments of* $\psi$ *is denoted by* $r \in \mathbb{N}$*, i.e.*

$$(5) \qquad \int_0^N t^k \psi(t)\, dt = 0, \; k = 0, 1, \ldots, r-1$$

*and*

$$(6) \qquad \int_0^N t^r \psi(t)\, dt = \nu_r \neq 0.$$

*For every* $J \geq 0$*,* $g \in L^2[0,1]$ *has a unique orthogonal expansion*

$$(7) \qquad g(t) = \sum_{k=-N+1}^{N2^J-1} s_{Jk}\phi_{Jk}(t) + \sum_{j=0}^{\infty} \sum_{k=-N+1}^{N2^{J+j}-1} d_{jk}\psi_{jk}(t),$$

$$(8) \qquad s_{Jk} = \int_0^1 g(t)\phi_{Jk}(t)\, dt, \; d_{jk} = \int_0^1 g(t)\psi_{jk}(t)\, dt.$$

*A (hard) thresholding wavelet estimator*

$$(9) \qquad \hat{g}(t) = \sum_{k=-N+1}^{N2^J-1} \hat{s}_{Jk}\phi_{Jk}(t) + \sum_{j=0}^{q} \sum_{k=-N+1}^{N2^{J+j}-1} \hat{d}_{jk}\, I(|\hat{d}_{jk}| > \delta_j)\psi_{jk}(t),$$

*(*$J$ *=decomposition level,* $q$ *=smoothing parameter,* $\delta_j$ *=threshold) with*

$$(10) \qquad \hat{s}_{Jk} = \frac{1}{n}\sum_{i=1}^{n} Y_i\phi_{Jk}(t_i), \; \hat{d}_{jk} = \frac{1}{n}\sum_{i=1}^{n} Y_i\psi_{jk}(t_i)$$

*can be understood as a combination of a smoothing component*

$$(11) \qquad \hat{g}_{low}(t) = \sum_{k=-N+1}^{N2^J-1} \hat{s}_{Jk}\phi_{Jk}(t)$$

*and a higher resolution component*

$$(12) \qquad \hat{g}_{high}(t) = \sum_{j=0}^{q} \sum_{k=-N+1}^{N2^{J+j}-1} \hat{d}_{jk}\, I(|\hat{d}_{jk}| > \delta_j)\psi_{jk}(t).$$

*The first component provides a good estimate of smooth functions whereas the second component is useful for modeling discontinuities. The asymptotically optimal rate of the integrated mean squared error (MISE) is attained for*

$$(13) \qquad J = J_n = \frac{\alpha}{2r+\alpha}\log_2 n + C, \; (C \in \mathbb{R}),$$

$$(14) \qquad q = q_n = \log_2 n - J,$$

$$(15) \qquad 2^{J+j}\delta_j^2 \to 0, \; 2^{(J+j)(2r+1)}\delta_j^2 \to \infty, \; \delta_j^2 \geq \frac{4eC_\psi^2 N^{-1+\alpha}(\ln n)^2}{n^\alpha 2^{(J+j)(1-\alpha)}}$$

*(Li and Xiao 2007, Beran and Shumeyko 2011a) where $C_\psi$ is a certain constant and $g$ is assumed to be $r$ times differentiable almost everywhere with a piecewise continuous $r$th derivative (where it exists). Explicit expressions for $q$ and $J$ including constants that minimize the actual asymptotic value of the $MISE$ are also derived in Beran and Shumeyko (2011a). In principle, these results can be used for an interative plug-in algorithm that allows for optimal data driven wavelet estimation of $g$.*

### 3. Bootstrap based testing for jumps

*Consider the null hypothesis $H_0 : g \in C[0,1]$ against the alternative that $g$ is only piecewise continuous, with a finite number of isolated jumps. The idea of the test statistic proposed in Beran and Shumeyko (2011b) is to compare residuals obtained by subtracting $\hat{g}$ with residuals where only the smooth component $\hat{g}_{low}$ was removed. Thus, let*

$$X_i = Y_i - \hat{g}(t_i), \;\; X_{i,low} = Y_i - \hat{g}_{low}(t_i). \tag{16}$$

*For blocks of size $l$ define block sums*

$$\zeta_i = X_i + \cdots + X_{i+l-1}, \;\; \zeta_{i,low} = X_{i,low} + \cdots + X_{i+l-1,low} \tag{17}$$

*and draw $\zeta_1^*, \ldots, \zeta_k^*$ randomly with replacement to obtain*

$$T_{kl}^* = a_l^{-1}\left(k^{-1/2}\sum_{i=1}^{k}\zeta_i^*\right), \;\; T_{kl,low}^* = a_l^{-1}\left(k^{-1/2}\sum_{i=1}^{k}\zeta_{i,low}^*\right) \tag{18}$$

*where $a_l = C_\gamma^{1/2} l^{1-\alpha/2}$. It can be shown that under $H_0$ both statistics have the same asymptotic distribution whereas this is no longer the case under $H_1$. More specifically, under $H_1$, the conditional variance (given the observed data) of $T_{kl}^*$ is equal to*

$$Var_n\left(T_{kl}^*\right) = \tilde{\sigma}^2 + o_p(1), \tag{19}$$

*with $\tilde{\sigma}^2 = 2\sigma^2(1-\alpha)^{-1}(2-\alpha)^{-1}$ whereas*

$$Var_n\left(T_{kl,low}^*\right) = \tilde{\sigma}^2 + v_n + o_p(1) \tag{20}$$

*with $v_n \to \infty$. Asymptotic normality of both statistics under $H_0$ and $H_1$ can also be derived. This leads to testing $H_0 : var(T_{kl,low}^*) = var(T_{kl}^*)$ against $H_1 : var(T_{kl,low}^*) > var(T_{kl}^*)$. To obtain a method that is also applicable for moderate sample sizes, a bootstrap based procedure is developed by comparing*

$$W_{low} = \tilde{\sigma}^{-2}\sum_{i=1}^{m}\left(T_{kl,low}^{*(i)} - \bar{T}_{kl,low}^*\right)^2, \tag{21}$$

*(with $T_{kl,low}^{*(i)}$ representing bootstrapped values of $T_{kl,low}^*$) with quantiles of*

$$W = \tilde{\sigma}^{-2}\sum_{i=1}^{m}\left(T_{kl}^{*(i)} - \bar{T}_{kl}^*\right)^2. \tag{22}$$

### REFERENCES

Banerjee, A. and Urgab, G. (2005). Modelling structural breaks, long memory and stock market volatility: an overview. *Journal of Econometrics 129, 1–34.*

Beran, J. (1994). Statistics for long-memory processes. *Chapman & Hall, London.*

Beran, J. and Feng, Y. (2002a). SEMIFAR models - a semiparametric framework for modelling trends, long-range dependence and nonstationarity. *Computational Statistics & Data Analysis, Vol. 40, No. 2, 393-419.*

Beran, J. and Feng, Y. (2002b). Data driven bandwidth choice for SEMIFAR models. *Journal of Computational and Graphical Statistics, Vol. 11, No. 2, 690-713.*

Beran, J. and Feng, Y. (2002c). Local polynomial fitting with long memory, short memory and antipersistent errors. *Annals of the Institute of Statisitcal Mathematics, Vol. 54, No. 2, 291-311.*

Beran, J. and Shumeyko, Y. (2011a). On asymptotically optimal wavelet estimation of trend functions under long-range dependence. *Bernoulli, in press.*

Beran, J. and Shumeyko, Y. (2011b). Bootstrap testing for discontinuities under long-range dependence. *Preprint.*

Csörgö, S. and Mielniczuk, J. (1995). Nonparametric regression under long-range dependent normal errors. *The Annals of Statistics, 23, 1000–1014.*

Ghosh, S. and Draghicescu, D. (2002a). Predicting the distribution function for long-memory processes. *International Journal of Forecasting, 18, 283-290.*

Ghosh, S. and Draghicescu, D. (2002b). An algorithm for optimal bandwidth selection for smooth nonparametric quantiles and distribution functions. *In: Statistics in Industry and Technology: Statistical Data Analysis based on the L1-norm and related methods, Birkhäuser Verlag, Basel, Switzerland, pp. 161-168.*

Hall, P. and Hart, J. D. (1990). Nonparametric regression with long-range dependence. *Stochastic Processes and Their Applications, 36, 339–351.*

Johnstone, I. M. and Silverman, B.W. (1997). Wavelet threshold estimators for data with correlated noise. *Journal of the Royal Statistical Society, Series B, 59, 319-351.*

Kulik, R. and Raimondo, M. (2009). $L^p$-wavelet regression with correlated errors and inverse problems. *Statistica Sinica, 19, 1479-1489.*

Lahiri S.N. (1993) On the moving block bootstrap under long range dependence. *Statistics & Probability Letters, 18, 405-413.*

Li, L. and Xiao, Y. (2007). Mean integrated squared error of nonlinear wavelet-based estimators with long memory data. *Annals of the Institute of Statistical Mathematics, Vol. 59, No. 2, 299-324.*

Percival, D.B., Sardy, S. and Davison, A.C. (2000). Wavestrapping time series: adaptive wavelet-based bootstrapping. *In: Nonlinear and Nonstationary Signal Processing, Fitzgerald, W.J., Smith, R.L., Walden, A.T. (eds.), Cambridge University Press, Cambridge, p. 442-471.*

Ray, B. K. and Tsay, R. S. (1997). Bandwidth selection for kernel regression with long-range dependent errors. *Biometrika, 84, 791-802.*

Robinson, P. M. (1997). Large sample inference for nonparametric regression with dependent errors. *The Annals of Statistics, 25, 2054-2083.*

Samorodnitsky, G. (2007). Long-range dependence. *Now Publishers Inc., Hanover, USA.*

Sibbertsen, P. (2004). Long memory versus structural breaks: An overview. *Statistical Papers, Vol. 45, No. 4, 465-515.*

Wang, Y. (1996). Function estimation via wavelet shrinkage for long-memory data. *The Annals of Statistics, 24, 466-484.*

Yang, Y. (2001). Nonparametric regression with dependent errors. *Bernoulli, Vol. 7, No. 4, 633-655.*