Detecting long-range dependence in non-stationary time series

Holger Dette, Ruhr-Universität Bochum
Philip Preuß, Ruhr-Universität Bochum
Kemal Sen, Ruhr-Universität Bochum

February 15, 2016

Outline

Motivation

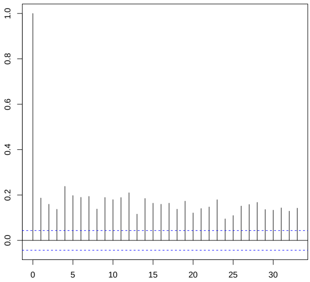


Figure: Sample autocovariance function of 2048 squared log-returns X_t^2 of the IBM stock (2005 - 2013)

X_t^2 might be considered as stationary long-range dependent.

Motivation

"Long-memory" features can also be as well explained by non-stationarity [Mikosch and Stărică (2004) or Chen et al. (2010)].

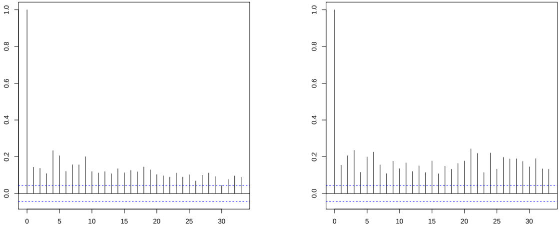


Figure: *Sample autocovariance function*

- Left panel: FARIMA(3,d,0)-fit to the squared IBM-returns
- Right panel: Fit of $X_{t,T}^2 = \hat{\sigma}^2(t/T)Z_t^2$
  - Z_t i.i.d.
  - σ piecewise constant

Motivation

Several authors point out the importance to discriminate between stationary long-range dependence and non-stationarity [see Stărică and Granger (2005), Perron and Qu (2010), Chen et al. (2010)].

- Künsch (1986) discriminates between LRD and SRD with changing trend
- Berkes et al. (2006), Baek and Pipiras (2012) and Yau and Davis (2012) test for
  - H_0 : one change point in mean in a short-range dependent process
  - H_1 : stationarity and long range dependence

Goal

Develop a test for the null hypothesis

$$H_0 : \text{no long-range dependence}$$

$$H_0 : \text{long-range dependence}$$

in a framework which is flexible enough to deal with different types of non-stationarity.

Locally stationary long-memory processes

Model:

- $(\{X_{t,T}\}_{t=1,\dots,T})_{T\in\mathbb{N}}$ locally stationary process [Dahlhaus (1997)]
- MA(∞) representation:

$$X_{t,T} \approx \mu(t/T) + \sum_{l=0}^{\infty} \psi_l(t/T) Z_{t-l}, \qquad t = 1, \dots, T$$

  - μ is twice continuously differentiable
  - $\{Z_t\}_{t\in\mathbb{Z}}$ i.i.d. $\mathcal{N}(0,1)$ (for simplicity)

Time-varying spectral density

Time varying spectral density

$$f(u,\lambda) = \frac{1}{2\pi}\left|\sum_{l=0}^{\infty}\psi_l(u)\exp(-i\lambda l)\right|^2$$

Assumption: AR(∞)-representation

$$f(u,\lambda) \quad = \quad \frac{1}{2\pi}|1-e^{i\lambda}|^{-2d_0(u)}\left|1+\sum_{l=1}^{\infty}a_{l,0}(u)\exp(-i\lambda l)\right|^{-2}$$

where $d_0 : [0,1] \to [0,1/2)$ is the (continuous) **time-vayring long-memory parameter**.

Hypotheses

$$\mathrm{H}_0 : d_0(u) = 0 \quad \forall u \in [0,1]$$ (non-stationarity **and no** long-range dependence)

vs. $$\mathrm{H}_1 : d_0(u) > 0 \quad \text{for some } u \in [0,1]$$ (non-stationarity **and** long-range dependence)

Hypotheses

$$\mathrm{H}_0 : d_0(u) = 0 \quad \forall u \in [0,1] \quad \text{(non-stationarity **and no** long-range dependence)}$$

vs.

$$\mathrm{H}_1 : d_0(u) > 0 \quad \text{for some } u \in [0,1] \quad \text{(non-stationarity **and** long-range dependence)}$$

Note: This is equivalent to

$$\mathrm{H}_0 : F = \int_0^1 d_0(u)du = 0 \quad \text{vs.} \quad \mathrm{H}_1 : F = \int_0^1 d_0(u)du > 0,$$

"Sieve" estimation

Approximate the spectral density with a sieve of semi-parametric models

- Choose a sequence $k = k(T) \in \mathbb{N}$, which diverges "slowly" to infinity as the sample size T grows (for example $k = \log(T)$).
- Decompose the sample into M blocks of length N and denote by u_j the midpoint of the j^{th} block.
- On each block we fit a time-varying FARIMA(k,d,0) model with spectral density

$$f_{\theta_k(u_j)}(\lambda) = |1 - e^{i\lambda}|^{-2d(u_j)} \frac{1}{2\pi} \left|1 + \sum_{l=1}^{k} a_l(u_j) e^{-i\lambda l}\right|^{-2}$$

and parameter $\theta_k(u_j) = (d(u_j), a_1(u_j), \ldots, a_k(u_j))$.

Sieve procedure

- Estimate $\theta_k(u_j)$ by a localized Whittle-estimator, that is

$$\hat{\theta}_{N,k}(u_j) \quad = \quad \arg \min_{\theta_k \in \Theta_{u_j,k}} \mathcal{L}^{\hat{\mu}}_{N,k}(\theta_k, u_j)$$

where

$$\mathcal{L}^{\hat{\mu}}_{N,k}(\theta_k, u_j) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \Big(\log(f_{\theta_k}(\lambda)) + \frac{I^{\hat{\mu}}_N(u_j, \lambda)}{f_{\theta_k}(\lambda)} \Big) d\lambda$$

is the local Whittle likelihood and

$$I^{\hat{\mu}}_N(u_j, \lambda) = \frac{1}{2\pi N} \Big| \sum_{p=0}^{N-1} \Big[X_{u_j T - N/2+1+p, T} - \hat{\mu}((u_j T - N/2 + 1 + p)/T) \Big] e^{-ip\lambda} \Big|^2 .$$

the mean-corrected local periodogram.

- Resulting estimator

$$\hat{\theta}_{N,k}(u_j) \quad = \quad (\hat{d}_N(u_j), \hat{a}_{N,1}(u_j), \ldots, \hat{a}_{N,k}(u_j)).$$

Estimator

- Estimate $d_0(u_j)$ by the first component $\hat{d}_N(u_j)$ of $\hat{\theta}_{N,k}(u_j)$.
- Esimate $F = \int_0^1 d_0(u)du$ by the mean

$$\hat{F}_T = \frac{1}{M} \sum_{j=1}^{M} \hat{d}_N(u_j).$$

Some technical assumptions

- For each $u \in [0,1]$ the parameter

$$\tilde{\theta}_{0,k}(u) = \arg \min_{\theta_k \in \Theta_{u,k}} \frac{1}{4\pi} \int_{-\pi}^{\pi} \Big(\log(f_{\theta_k}(\lambda)) + \frac{f(u,\lambda)}{f_{\theta_k}(\lambda)} \Big) d\lambda$$

exists and is uniquely determined.

Some technical assumptions

- For each $u \in [0,1]$ the parameter

$$\tilde{\theta}_{0,k}(u) = \arg \min_{\theta_k \in \Theta_{u,k}} \frac{1}{4\pi} \int_{-\pi}^{\pi} \Big(\log(f_{\theta_k}(\lambda)) + \frac{f(u,\lambda)}{f_{\theta_k}(\lambda)} \Big) d\lambda$$

  exists and is uniquely determined.
- Assumption regarding the approximation error by parametric models:

$$\sup_{u \in [0,1]} \int_{-\pi}^{\pi} \Big| f(u,\lambda) - f_{\theta_{0,k}(u)}(\lambda) \Big| d\lambda \quad = \quad O(N^{-1+\epsilon})$$

  where $\theta_{0,k}(u) = (d_0(u), a_{1,0}(u), ..., a_{k,0}(u))$ is the FARIMA$(k, d, 0)$-parameter
- (satisfied for geometrically decaying AR coefficients $a_{l,0}(u) \longrightarrow k = \log T$)

Local window estimator

The mean-function $\mu(u)$ is estimated by

$$\hat{\mu}_L(u) = \frac{1}{L}\sum_{p=0}^{L-1} X_{\lfloor uT \rfloor - L/2+1+p,T}.$$

Local window estimator

The mean-function $\mu(u)$ is estimated by

$$\hat{\mu}_L(u) = \frac{1}{L}\sum_{p=0}^{L-1} X_{\lfloor uT \rfloor - L/2+1+p,T}.$$

Note:

$$L^{1/2-D-\alpha} \max_{t=1,\ldots,T} \left|\mu(t/T) - \hat{\mu}(t/T)\right| = o_p(1)$$

for every $\alpha > 0$ where $D = \sup_{u\in[0,1]} d_0(u) < 1/2$.

Asymptotic properties of $\hat{F}_T$ under H_0

Theorem

If $F = 0$ and the conditions

$$N^{1+4\epsilon}/L^{1-\delta} \to 0,\ L^{5/2-\delta}/T^2 \to 0,$$
$$k^6\sqrt{T}/N^{1-\epsilon} \to 0,\ k^4\log^2(T)/N^{\epsilon/2} \to 0,\ k^2N^2/T^{\frac{3}{2}} \to 0$$

are satisfied as $M, N, T \to \infty$ for $0 < \epsilon, \delta < 1/6$, then

$$\boxed{\sqrt{T}\hat{F}_T/\sqrt{W_T} \xrightarrow{D} \mathcal{N}(0,1)}$$

where

$$\begin{aligned} W_T &= \left[\int_0^1 \Gamma_k^{-1}(\theta_{0,k}(u))du\right]_{1,1} \\ \Gamma_k(\theta_{0,k}(u)) &= \frac{1}{4\pi}\int_{-\pi}^{\pi} f^2_{\theta_{0,k}(u)}(\lambda)\nabla f^{-1}_{\theta_{0,k}(u)}(\lambda)\nabla f^{-1}_{\theta_{0,k}(u)}(\lambda)^T\, d\lambda. \end{aligned}$$

Asymptotic properties of $\hat{F}_T$ under H_1

Theorem

If $F > 0$ *and the conditions* $N^{\epsilon}k^5/L^{1/2-D-\delta} \to 0, \quad L^{5/2-D-\delta}/T^2 \to 0,$

$$k^6/N^{1-2\epsilon} \to 0, \quad k^4 \log^2(T)/N^{\epsilon/2} \to 0,$$
$$k^4/N^{1-2D-2\epsilon} \to 0, \quad k^2 N^{5/2}/T^2 \to 0$$

are satisfied as $M, N, T \to \infty$ *for* $0 < \delta < \epsilon < \min\{1/2 - D, 1/6\}$, *then*

$$\boxed{\hat{F}_T \xrightarrow{P} F > 0.}$$

Test for long-memory

- Estimate the asymptotic variance consistently by

$$\hat{W}_T = \Big[\frac{1}{M}\sum_{j=1}^{M}\Gamma_k^{-1}(\hat{\theta}_{N,k}(u_j))\Big]_{11}$$

- Consistent asymptotic level α-test: Reject the null hypothesis (of no long-range dependence), whenever

$$\sqrt{T}\hat{F}_T/\sqrt{\hat{W}_T} \geq u_{1-\alpha},$$

where $u_{1-\alpha}$ denotes the $(1-\alpha)$-quantile of the standard normal distribution.

Finite sample properties

Choice of regularization parameters:

- Choose $L = N^{1.05}$.
- Choose k with the AIC criterion, that is

$$\hat{k} = \arg\min_k \frac{1}{T} \sum_{j=1}^{T/2} \left(\log(h_{\hat{\theta}_{k,s}}(\lambda_j)) + \frac{I^{\hat{\mu}}(\lambda_j)}{h_{\hat{\theta}_{k,s}}(\lambda_j)} \right) + \frac{k+1}{T},$$

where

- $\lambda_j = 2\pi j/T \ (j = 1, \ldots, T)$,
- $h_{\hat{\theta}_{k,s}}(\lambda)$ estimated spectral density of a FARIMA$(k, d, 0)$-process and
- $I^{\hat{\mu}_L}(\lambda) = \frac{1}{2\pi N} \left| \sum_{t=1}^{T} \left[X_{t,T} - \hat{\mu}_L(t/T) \right] e^{-it\lambda} \right|^2$ mean-corrected periodogram.

Note: We use the same k in all blocks

Approximation of the nominal level

- time-vayring AR(1)-error process

$$
\begin{aligned}
X_{t,T} &= \mu_i(t/T) + Y_{t,T} \quad t = 1, \ldots, T \\
Y_{t,T} &= 0.6\frac{t}{T} Y_{t-1,T} + Z_{t,T}, \quad t = 1, \ldots, T,
\end{aligned}
$$

with

$$
\begin{aligned}
\text{(smooth mean)} \qquad \mu_1(t/T) &= 1.2\frac{t}{T}, \\
\text{(change in mean)} \qquad \mu_2(t/T) &= \begin{cases} 0.65 & \text{for } t = 1, \ldots, T/2 \\ 1.3 & \text{for } t = T/2 + 1, \ldots T. \end{cases}
\end{aligned}
$$

Approximation of the nominal level

- time-vayring AR(1)-error process

$$\begin{aligned} X_{t,T} &= \mu_i(t/T) + Y_{t,T} \quad t = 1, \dots, T \\ Y_{t,T} &= 0.6\frac{t}{T} Y_{t-1,T} + Z_{t,T}, \quad t = 1, \dots, T, \end{aligned}$$

with

$$\begin{aligned} \text{(smooth mean)} \qquad \mu_1(t/T) &= 1.2\frac{t}{T}, \\ \text{(change in mean)} \qquad \mu_2(t/T) &= \begin{cases} 0.65 & \text{for } t = 1, \dots, T/2 \\ 1.3 & \text{for } t = T/2+1, \dots T. \end{cases} \end{aligned}$$

- time-vayring MA(1)-process

$$X_{t,T} = Z_{t,T} + 0.55 \sin\left(\pi\frac{t}{T}\right) Z_{t-1,T}, \quad t = 1, \dots, T$$

where $\{Z_{t,T}\}_{t=1,\dots,T}$ is Gaussian white noise with variance 1.

Approximation of the nominal level

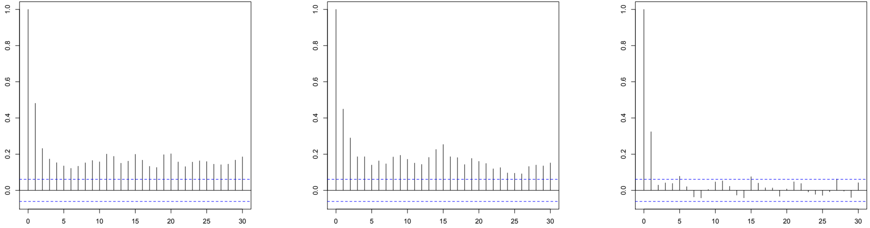


Figure: *Autocovariance functions (T=1024): Left panel: tvAR(1)-error process with smooth mean function. Middle panel: tvAR(1)-error process with a change in mean. Right panel: tvMA(1)-process.*

Approximation of the nominal level

			smooth mean		change in mean		tvMA(1)-process	
T	N	M	5%	10%	5%	10%	5%	10%
256	64	4	0.090	0.128	0.094	0.145	0.085	0.122
256	32	8	0.151	0.228	0.165	0.255	0.182	0.261
512	128	4	0.061	0.095	0.070	0.114	0.069	0.099
512	64	8	0.089	0.130	0.089	0.126	0.081	0.107
1024	256	4	0.046	0.072	0.077	0.119	0.069	0.106
1024	128	8	0.059	0.087	0.061	0.088	0.064	0.093

Table: *Simulated level of the new test.*

Power of the test

Alternative procedures designed for testing

- H_0 : Short-range dependence and change in mean
- H_1 : Stationarity and long-range dependence

Three tests:

- Berkes et al. (2006) estimate a change point and consider two CUSUM statistics in the samples before and after the change point.
- Baek and Pipiras (2012) remove the mean effect and reject for large values of the local Whittle estimate of the long-range dependence parameter.
- Yau and Davis (2012) use a parametric likelihood ratio test assuming two (not necessarily equal) ARMA(p, q) models before and after a change in the mean function.

Power of the test

All competing procedures are designed to detect stationary long-range dependent alternatives. Simulate a stationary FARIMA(1,d,1)-process

$$(1+0.25B)(1-B)^{0.1}X_T = (1-0.3B)Z_{t,T}, \quad t=1,\dots,T.$$

			new test		Baek/Pipiras		Berkes et. al		Yau/Davis	
T	N	M	5%	10%	5%	10%	5%	10%	5%	10%
256	64	4	0.094	0.136	0.087	0.149	0.045	0.093	0.178	0.210
256	32	8	0.138	0.216						
512	128	4	0.146	0.196	0.119	0.177	0.022	0.055	0.140	0.176
512	64	8	0.138	0.214						
1024	256	4	0.328	0.406	0.127	0.197	0.018	0.079	0.152	0.206
1024	128	8	0.152	0.218						

Table: *Rejection frequencies of the new test and three competing procedures.*

Power of the test

But the new test is consistent against more general non-stationary alternatives.

We simulated data from a time-varying FARIMA$(1, d, 0)$-process

$$(1 + 0.2\frac{t}{T}B)(1 - B)^{d(t/T)}X_{t,T} = Z_{t,T}, \quad t = 1, \dots, T$$

with long-memory function $d(t/T) = 0.1 + 0.3t/T$.

			new test		Baek/Pipiras		Berkes et. al		Yau/Davis	
T	N	M	5%	10%	5%	10%	5%	10%	5%	10%
256	64	4	0.288	0.354	0.248	0.330	0.037	0.080	0.250	0.306
256	32	8	0.290	0.436						
512	128	4	0.530	0.590	0.356	0.468	0.006	0.041	0.182	0.226
512	64	8	0.348	0.458						
1024	256	4	0.746	0.770	0.562	0.656	0.026	0.102	0.204	0.267
1024	128	8	0.412	0.512						

Table: *Rejection frequencies of the new test and three competing procedures.*

Power of the test

We simulated data from a time-varying FARIMA$(0, d, 1)$-process

$$(1 - B)^{d(t/T)} X_{t,T} = (1 - 0.35\frac{t}{T} B) Z_{t,T}, \quad t = 1, \ldots, T$$

with long-memory function $d(t/T) = 0.1 + 0.3t/T$.

			new test		Baek/Pipiras		Berkes et. al		Yau/Davis	
T	N	M	5%	10%	5%	10%	5%	10%	5%	10%
256	64	4	0.260	0.330	0.230	0.322	0.039	0.088	0.296	0.366
256	32	8	0.276	0.394						
512	128	4	0.528	0.590	0.342	0.456	0.010	0.036	0.268	0.322
512	64	8	0.314	0.414						
1024	256	4	0.774	0.796	0.546	0.656	0.024	0.086	0.228	0.292
1024	128	8	0.414	0.492						

Table: *Rejection frequencies of the new test and three competing procedures.*

Data example

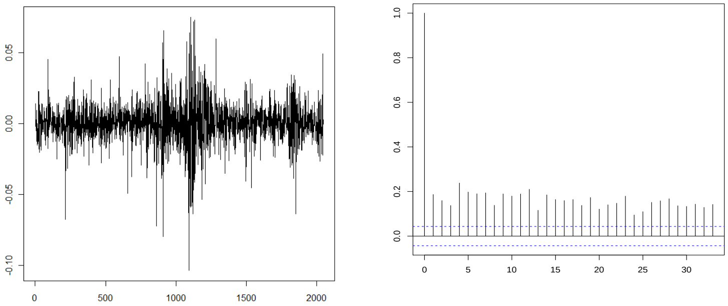
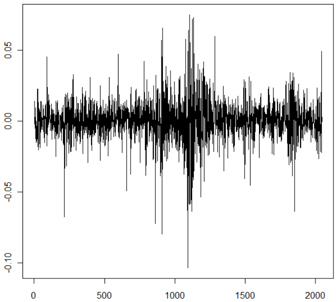


Figure: Log-returns of the IBM stock (2005 - 2013) and sample autocovariance function of the squared log-returns.

Data example



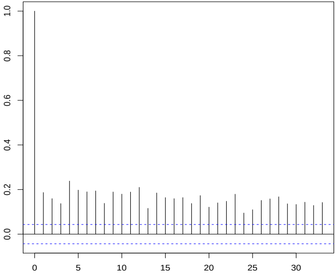


Figure: Log-returns of the IBM stock (2005 - 2013) and sample autocovariance function of the squared log-returns.

P-value of new test: 0.971

Constrained versus unconstrained inference

- **Note:** we consider a constrained testing problem (under the null hypothesis the function $d_0 : [0,1] \to [0,1/2)$ is boundary point of the parameter space):

$$\begin{aligned} & \mathrm{H}_0 : d_0(u) = 0 \quad \forall u \in [0,1] \\ \text{vs.} \qquad & \mathrm{H}_1 : d_0(u) > 0 \quad \text{for some } u \in [0,1] \end{aligned}$$

Constrained versus unconstrained inference

- **Note:** we consider a constrained testing problem (under the null hypothesis the function $d_0 : [0,1] \to [0,1/2)$ is boundary point of the parameter space):

$$\begin{aligned} & \mathrm{H}_0 : d_0(u) = 0 \quad \forall u \in [0,1] \\ \text{vs.} \qquad & \mathrm{H}_1 : d_0(u) > 0 \quad \text{for some } u \in [0,1] \end{aligned}$$

- In general tests for these type of hypotheses are not asymptotically normal distributed [see Chernoff (1954)]!

Constrained versus unconstrained inference

- **Note:** we consider a constrained testing problem (under the null hypothesis the function $d_0 : [0,1] \to [0,1/2)$ is boundary point of the parameter space):

$$\begin{aligned} & \mathrm{H}_0 : d_0(u) = 0 \quad \forall u \in [0,1] \\ \text{vs.} \quad & \mathrm{H}_1 : d_0(u) > 0 \quad \text{for some } u \in [0,1] \end{aligned}$$

- In general tests for these type of hypotheses are not asymptotically normal distributed [see Chernoff (1954)]!
- **However:**
  (1) We do not use the information $d_0(u) \geq 0$ in the construction of the test statistic
  (2) We form averages of $d_0(u_i) \geq 0$ $(i = 1, \ldots, M)$
- **The resulting test statistic is asymptotically normal distributed**

Constrained versus unconstrained inference - final example

- Let $X_1, \ldots, X_n$ i.i.d. , $\mathbb{E}[X_i^2] = 1$; $\mu = \mathbb{E}[X_i] \geq 0$;

$$\mathrm{H}_0 : \mu = 0 \qquad \text{vs.} \qquad \mathrm{H}_1 : \mu > 0$$

- **Unconstrained test:** rejects H_0 for large values of

$$\sqrt{n}\, \overline{X}_n$$

  using quantiles of the normal distribution.
- **Constrained test:** rejects H_0 for large values of

$$\max\{\sqrt{n}\, \overline{X}_n, 0\}$$

  using quantiles of the distribution $\max\{Z, 0\}$, where $Z \sim \mathcal{N}(0, 1)$.

Constrained versus unconstrained inference - final example

- Let $X_1, \ldots, X_n$ i.i.d. , $\mathbb{E}[X_i^2] = 1$; $\mu = \mathbb{E}[X_i] \geq 0$;

$$\mathrm{H}_0 : \mu = 0 \qquad \text{vs.} \qquad \mathrm{H}_1 : \mu > 0$$

- **Unconstrained test:** rejects H_0 for large values of

$$\sqrt{n}\,\overline{X}_n$$

  using quantiles of the normal distribution.

- **Constrained test:** rejects H_0 for large values of

$$\max\{\sqrt{n}\,\overline{X}_n, 0\}$$

  using quantiles of the distribution $\max\{Z, 0\}$, where $Z \sim \mathcal{N}(0,1)$.

- **In the present context:** unconstrained inference is more powerful than constrained inference (due to averaging)

References

- Baek and Pipiras (2012): Statistical tests for a single change in mean against long-range dependence, Journal of Time Series Analysis, 33, 131-151.
- Berkes, Horváth, Kokoszka, Shao (2006): On discriminating between long-range dependence and changes in mean, The Annals of Statistics, 34, 1140-1165.
- Chen, Härdle, Pigorsch (2010): Localized realized volatility modeling, Journal of the American Statistical Association, 105, 1376-1393.
- Chernoff, H. (1954). On the distribution of the likelihood ratio. Annals of Mathematical Statistics, 25, 573 - 578.
- Dahlhaus (1997): Fitting time series models to nonstationary processes, The Annals of Statistics, 25, 1-37.
- Künsch (1986): Discrimination between monotonic trends and long-range dependence, J. Appl. Prob., 23, 1025-1030.
- Mikosch and Stărică (2004): Non-stationarities in financial time series, the long-range dependence, and the IGARCH effects, The Review of Economics and Statistics, 86, 378-390.
- Perron and Qu (2010): Long-Memory and Level Shifts in the Volatility of Stock Market Return Indices, Journal of Business and Economic Statistics, 23, 275-290.
- Preuss, Sen, Dette (2014-2016): Detecting long-range dependence in non-stationary time series, arXiv..
- Stărică and Granger (2005): Nonstationarities in stock returns, The Review of Economics and Statistics, 87, 503-522.
- Yau and Davis (2012): Likelihood inference for discriminating between long-memory and change-point models, Journal of Time Series Analysis, 33, 649-664.