INFORMATICA, 2011, Vol. 22, No. 1, 97–114

© 2011 *Vilnius University*

On Comparison of the Estimators of the Hurst Index of the Solutions of Stochastic Differential Equations Driven by the Fractional Brownian Motion

Kęstutis KUBILIUS[1*], Dmitrij MELICHOV[2]
[1]*Vilnius University Institute of Mathematics and Informatics*
Akademijos 4, LT-08663 Vilnius, Lithuania
[2]*Vilnius Gediminas Technical University*
Saulėtekio 11, LT-10223 Vilnius, Lithuania
e-mail: kestutis.kubilius@mii.vu.lt

Received: August 2010; accepted: November 2010

Abstract. This paper presents a study of the Hurst index estimation in the case of fractional Ornstein–Uhlenbeck and geometric Brownian motion models. The performance of the estimators is studied both with respect to the value of the Hurst index and the length of sample paths.

Keywords: fractional Brownian motion, Hurst index, Ornstein–Uhlenbeck process, geometric Brownian motion.

1. Introduction

In this paper, we consider the stochastic differential equation

$$X_t = c + \int_0^t f(X_s)\,\mathrm{d}s + \int_0^t g(X_s)\,\mathrm{d}B_s^H, \quad t \in [0;1],\ c \in \mathbb{R}, \tag{1}$$

where B_t^H is the fractional Brownian motion (fBm) with the Hurst index $1/2 < H < 1$. fBm is a continuous-time Gaussian process $\{B_t^H, t \geqslant 0\}$ with $B_0^H = 0$ and the covariance function

$$\mathbf{E}\big[B_t^H B_s^H\big] = \frac{1}{2}\big(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\big).$$

The Hurst index H determines the correlation between the increments of fBm. If $H < 1/2$, these increments are negatively correlated, if $H > 1/2$, the increments are positively correlated, and if $H = 1/2$, the process is a regular Brownian motion. In

*Supported by the Research Council of Lithuaniam contract No. MIP-66/2010.

this paper the case $H > 1/2$ is considered since it allows to model various phenomena possessing the long-range dependance property. The second integral in (1) is that of Riemann–Stieltjes defined pathwise. For $0 < \alpha \leqslant 1$, $\mathcal{C}^{1+\alpha}(\mathbb{R})$ denotes the set of all $\mathcal{C}^1$-functions $g: \mathbb{R} \to \mathbb{R}$ such that

$$\sup_x |g'(x)| + \sup_{x \neq y} \frac{|g'(x) - g'(y)|}{|x - y|^\alpha} < \infty.$$

Let f be a Lipschitz function, $g \in \mathcal{C}^{1+\alpha}(\mathbb{R})$, $0 < \alpha \leqslant 1$. It is known (Lyons, 1994; Dudley, 1999; Kubilius, 2000; Nualart and Răşcanu, 2002) that there exists a unique strong solution of (1).

The fractional Brownian motion and processes based on it have found many applications in fields as diverse as economics and finance, physics, chemistry, medicine and environmental studies. Computer science is not an exception – fBm has been employed in telecommunication traffic studies (for an overview, see Li, 2010 and the references therein), fractal image compression (Davis, 1996), image object detection (see, i.e., Wenlu and Weixin, 1997) and other studies. Obviously, if some phenomenon can be modeled by fBm or a process based on it, the estimation of the fBm's parameter, the Hurst index H, is an important problem.

In 1961, Gladyshev (1963) derived a limit theorem for a statistic based on the first order quadratic variations of fBm. This yielded an estimator of H which was strongly consistent. In 1997, another estimator was introduced by Istas and Lang (1997) which again employed the first order quadratic variations and it was asymptotically normal for $H \in (1/2; 3/4)$. In 2005, Bégyn (2005) considered the second order quadratic variations along general subdivisions for processes with Gaussian increments. A more complete survey about asymptotic behavior of quadratic variations for Gaussian processes can be found in the thesis Bégyn (2006). In 2006, Berzin and León (2008) proposed a CLT for estimators of H and the diffusion function g for several specific cases of (1). In 2008–2010, Kubilius and Melichov (2008, 2009, 2010) studied the behavior of the first and second order quadratic variations of the pathwise solution of (1) and have shown that the quadratic variation based estimators remain strongly consistent in that case as well.

The goal of this paper is to compare the behavior of the estimators based on quadratic variations with some of the other known estimators, namely the naive and ordinary least squares Gladyshev and η-summing oscillation estimators, the variogram estimator and the IR estimator. These estimators are described in Section 1.2. Most of them were examined for Gaussian processes. The models chosen for comparison of these estimators were the fractional Ornstein–Uhlenbeck (O-U) and the fractional geometric Brownian motion (gBm). The initial inference about the behavior of these estimators was drawn for the O-U process which is Gaussian, while the gBm process was used to check how the estimators behave in a non-Gaussian case.

In order to achieve that, a sufficient amount of fBm sample paths is required. These sample paths were generated using the circulant matrix embedding method, as described in Coeurjolly (2000) and the references therein. Let n denote the length of the sample path. The circulant matrix embedding method uses a fast Fourier transform which bypasses matrix computations and therefore is sufficiently fast even for large values of n. 100 sample paths of the length $n = 2^{14} + 1$ were generated for each value of $H \in \{0.55, 0.6, \ldots, 0.95\}$ on the unit interval $t \in [0; 1]$.

The next step would be to use the generated fBm data to construct sample paths of the considered processes. However, it's not always possible to find and use the explicit solution of (1), therefore this solution needs to be replaced by a time discrete approximation. For a process X_t, its Milstein approximation at points t_k^n, $k = 1, \ldots, n$ is defined as

$$X_k^n = X_{k-1}^n + f\big(X_{k-1}^n\big)\Delta t_k + g\big(X_{k-1}^n\big)\Delta B_k^H + \frac{1}{2} g\big(X_{k-1}^n\big) g'\big(X_{k-1}^n\big)\big(\Delta B_k^H\big)^2,$$

where g' denotes the derivative of g and $X_0^n = c$. The fractional Ornstein–Uhlenbeck (O-U) and the fractional geometric Brownian motion (gBm) processes are defined as

$$\mathrm{d}X_t = -\mu X_t \,\mathrm{d}t + \sigma \,\mathrm{d}B_t^H, \quad X_0 = c, \tag{O-U}$$

$$\mathrm{d}X_t = \mu X_t \,\mathrm{d}t + \sigma X_t \,\mathrm{d}B_t^H, \quad X_0 = c. \tag{gBm}$$

The solutions of these equations are, respectively,

$$X_t = e^{-\mu t}\left(c + \sigma \int_0^t e^{\mu s}\,\mathrm{d}B_s^H\right) \quad \text{and} \quad X_t = c \exp\big(\mu t + \sigma B_t^H\big).$$

In fact, for the O-U process the Milstein approximation is reduced to the Euler one due to $g'(X_{k-1}^n) = (\sigma)' = 0$. The constants were chosen as $c = 1$, $\mu = 0.5$, $\sigma = 0.7$ in the O-U case and $c = 1$, $\mu = 0.2$, $\sigma = 0.5$ in the gBm case. The error introduced by using these approximated sample paths was negligible compared to the errors of the estimators themselves and will be ignored further on. All computations were performed using the `R` software environment (R Development Core Team, 2009).

1.1. *fBm Generation*

The algorithm to generate one sample path of the length n, using the circulant matrix embedding method, is as follows:

- Choose $M = 2^p \geqslant 2(n-1)$. Define the M-vector

$$V = \left(r(0), r(1), \ldots, r\left(\frac{M}{2} - 1\right), r\left(\frac{M}{2}\right), r\left(\frac{M}{2} - 1\right), \ldots, r(2), r(1)\right),$$

  where

$$r(k) = \frac{1}{2n^{2H}}\big[|k+1|^{2H} - 2k^{2H} + |k-1|^{2H}\big]$$

  is the autocovariance function of the fractional Gaussian noise.

- Compute $W = (w_1, \ldots, w_M)$, the fast Fourier transformation of V. All the coordinates of W must be non-negative. If this is not the case, the value of p must be increased until this requirement is met.
- Generate $U_j, V_j \sim \mathcal{N}(0,1)$ for all $1 \leqslant j < \frac{M}{2}$ and let $Z_1 = U_1$, $Z_{\frac{M}{2}+1} = V_1$,

$$Z_j = \frac{1}{\sqrt{2}}\big(U_j + iV_j\big), \quad Z_{M+2-j} = \frac{1}{\sqrt{2}}\big(U_j - iV_j\big), \quad 1 < j \leqslant \frac{M}{2}.$$

  Then, define the M-vector U as

$$U_k = \sqrt{w_k} Z_k, \quad k = 1, \ldots, M.$$

- Compute Y as an inverse fast Fourier transformation of the complex vector U and define X as

$$X_k = X_{k-1} + \Re(Y_k), \quad X_0 = 0,\ k = 1, \ldots, n,$$

  $\Re(Y)$ denoting the real part of the complex variable Y.

The obtained vector X is the desired sample path of the fractional Brownian motion with the Hurst index H.

1.2. *Estimators*

1.2.1. *Discrete Variation Estimators*

For a real-valued process $X = \{X_t, t \in [0,1]\}$, we define the first and second order quadratic variations as

$$V_n^{(1)}(X,2) = \sum_{k=1}^{n} \big(\Delta_k^{(1)} X\big)^2, \qquad V_n^{(2)}(X,2) = \sum_{k=1}^{n-1} \big(\Delta_k^{(2)} X\big)^2,$$

where

$$\Delta_k^{(1)} X = X(t_k^n) - X\big(t_{k-1}^n\big),$$
$$\Delta_k^{(2)} X = X\big(t_{k+1}^n\big) - 2X\big(t_k^n\big) + X\big(t_{k-1}^n\big), \quad t_k^n = \frac{k}{n}.$$

Let X be the solution of (1). It is known (see, Kubilius and Melichov, 2008, 2009, 2010) that

$$\widehat{H}_{\mathrm{dv1}}^n = \frac{1}{2} - \frac{1}{2\ln 2} \ln \frac{V_{2n}^{(1)}(X,2)}{V_n^{(1)}(X,2)}, \qquad \widehat{H}_{\mathrm{dv2}}^n = \frac{1}{2} - \frac{1}{2\ln 2} \ln \frac{V_{2n}^{(2)}(X,2)}{V_n^{(2)}(X,2)}$$

are strongly consistent estimators of the Hurst index H, i.e.,

$$\widehat{H}_{\mathrm{dv1}}^n - H \xrightarrow{\mathbf{a.s.}} 0 \quad \text{and} \quad \widehat{H}_{\mathrm{dv2}}^n - H \xrightarrow{\mathbf{a.s.}} 0 \quad \text{as } n \to \infty.$$

Here $V_{2n}^{(\cdot)}(X,2)$ corresponds to the quadratic variation of the whole sample path while $V_n^{(\cdot)}(X,2)$ is the variation of the subset $\{X_k : k = 2j,\ 0 \leqslant j \leqslant [n/2]\}$ where $[x]$ denotes the integer part of x.

1.2.2. *Gladyshev and η-Summing Oscillation Estimators*

The following estimators were described in Norvaiša and Salopek (2002). The ordinary least squares (OLS) Gladyshev and η-summing oscillation estimators require a sample path of the length $2^n + 1$, $n \in \mathbb{N}$, which dictated the length of our modeled sample paths. Define $\eta_M = \{N_m = 2^m \colon 1 \leqslant m \leqslant M\}$ and let

$$s(m) = \sum_{i=1}^{N_m} \left[X\left(\frac{i}{N_m}\right) - X\left(\frac{i-1}{N_m}\right)\right]^2.$$

The naive Gladyshev estimator of the Hurst index H is given by

$$\widehat{H}_{\text{gn}}^M = \frac{\log \sqrt{s(M)2^{-M}}}{\log 2^{-M}},$$

and the OLS Gladyshev estimator is given by

$$\widehat{H}_{\text{go}}^M = \frac{\sum_{m=1}^M \left(z_m - \bar{z}\right)^2}{\sum_{m=1}^M \left(z_m - \bar{z}\right) m},$$

where $z_m = \log_2 \sqrt{2^m / s(m)}$ for $m \in [1, \dots, M]$ and $\bar{z} = M^{-1} \sum_{m=1}^M z_m$.

For every $m \in [1, \dots, M]$, define

$$Q(m) = \sum_{i=1}^{N_m} \Big[\max_{t_k^n \in \Delta_{i,m}} \left\{X\left(t_k^n\right)\right\} - \min_{t_k^n \in \Delta_{i,m}} \left\{X\left(t_k^n\right)\right\}\Big],$$

where

$$\Delta_{i,m} = \left[\frac{i-1}{N_m}; \frac{i}{N_m}\right].$$

The naive oscillation estimator is defined by

$$\widehat{H}_{\text{osn}}^M = \frac{\log_2 \left(N_M / Q(M)\right)}{\log_2 N_M},$$

and the OLS oscillation estimator is defined by

$$\widehat{H}_{\text{oso}}^M = \frac{\sum_{m=1}^M \left(z_m - \bar{z}\right)^2}{\sum_{m=1}^M \left(z_m - \bar{z}\right) N_m},$$

where $z_m = \log_2 \sqrt{N_m / Q(m)}$ and $\bar{z} = M^{-1} \sum_{m=1}^M z_m$.

For $M = 14$ we simulate estimates defined above.

1.2.3. *Variogram Estimator*

The variogram of the process $X = \{X_t, t \in [0,1]\}$ for the lag l is defined (Chronopoulou and Viens, 2010) as

$$V(l) = \mathbb{E}\big[\big(X_t - X_{t-l}\big)^2\big].$$

In order to estimate the Hurst index H, we choose a set of lags, in our case, it was $\{l = 2^i, i = 0, \dots, 5\}$. Then $\widehat{H}^n_{\mathrm{var}} = b/2$, where b is the slope of the linear regression line of $\log(V(l))$ against $\log(l)$.

1.2.4. *IR Estimator*

This estimator was proposed by Bardet and Surgailis (2010). For the O-U or gBm process $X = \{X_t, t \in [0,1]\}$ given at points $t^n_k = k/n$, $k = 0, 1, \dots, n$, the IR estimator of H can be computed using the approximated formula below

$$\widehat{H}^n_{\mathrm{ir}} = \frac{1}{0.1468}\left(\frac{1}{n-2}\sum_{k=1}^{n-2}\frac{|\Delta^{(2)}_k X + \Delta^{(2)}_{k+1} X|}{|\Delta^{(2)}_k X| + |\Delta^{(2)}_{k+1} X|} - 0.5174\right),$$

where $\Delta^{(2)}_k X = X(t^n_{k+1}) - 2X(t^n_k) + X(t^n_{k-1})$.

2. Estimation for the Ornstein–Uhlenbeck Process

2.1. *Dependance on the Value of the Hurst Index*

The first goal of this paper is to compare the behavior of these estimators for different values of the Hurst index H. Table 1 presents the biases $\overline{H} - H = \mathbf{E}(\widehat{H} - H)$ as well as the mean squared errors defined as $\mathrm{MSE}(\widehat{H}) = \mathbf{E}(\widehat{H} - H)^2$ for the sample path lengths of, respectively, $2^{14} + 1$ and $2^{10} + 1$ points. Figure 1 illustrates this further presenting the boxplots of the considered estimators for the length of sample paths $n = 2^{14} + 1$ points. Here and further in this paper the figures related to the estimators $\widehat{H}_{\mathrm{gn}}$ and $\widehat{H}_{\mathrm{go}}$ are omitted, since their behavior does not significantly differ from the behavior of $\widehat{H}_{\mathrm{osn}}$ and $\widehat{H}_{\mathrm{oso}}$. The numbers printed in bold correspond to the estimators that performed better than the others for the specific value of H and the measure considered.

It can be seen that the estimators $\widehat{H}_{\mathrm{dv1}}$, $\widehat{H}_{\mathrm{var}}$, $\widehat{H}_{\mathrm{gn}}$ and $\widehat{H}_{\mathrm{osn}}$ exhibit increases of the biases and the mean squared errors for larger values of H. $\widehat{H}_{\mathrm{go}}$ and $\widehat{H}_{\mathrm{oso}}$ seem to be less dependant on that, however, they tend to slightly undervalue the Hurst index when it is close to 1. $\widehat{H}_{\mathrm{ir}}$ tends to slightly undervalue H when $H < 3/4$ and to overvalue it when $H > 3/4$; the most likely cause of this are the numerical constants in the formula used for this estimator. The behavior of $\widehat{H}_{\mathrm{dv2}}$ does not change noticeably for different values of H.

Another interesting observation is the, comparatively, very low mean squared errors of $\widehat{H}_{\mathrm{gn}}$ and $\widehat{H}_{\mathrm{osn}}$ which they display as long as the Hurst index is not too close to 1.

Table 1

Comparison of the estimators for the O-U process

H		0.55	0.7	0.8	0.95
MSE	dv1	0.008	**0.005**	**0.008**	0.020
	dv2	0.015	0.011	0.011	**0.010**
	var	**0.007**	0.009	0.014	0.024
	gn	**0.002**	**0.002**	**0.003**	**0.021**
	osn	**0.004**	**0.004**	**0.005**	0.025
	go	0.038	0.037	0.048	0.050
	oso	0.041	0.037	0.050	0.053
	ir	0.022	0.019	0.019	0.022
$\overline{H} - H$	dv1	**0.000**	**0.001**	$-$**0.001**	$-$**0.009**
	dv2	**0.000**	**0.001**	-0.002	$-$**0.001**
	var	**0.000**	$-$**0.001**	**0.000**	$-$**0.012**
	gn	0.037	0.037	0.037	0.041
	osn	0.060	0.060	0.060	0.063
	go	-0.013	-0.024	-0.019	-0.032
	oso	-0.004	0.017	-0.009	-0.020
	ir	-0.019	-0.009	-0.001	0.028

(a) $N = 2^{14} + 1$

H		0.55	0.7	0.8	0.95
MSE	dv1	0.027	0.024	**0.026**	**0.030**
	dv2	0.050	0.054	0.050	0.044
	var	0.029	0.031	0.038	**0.040**
	gn	**0.006**	**0.006**	**0.010**	**0.040**
	osn	**0.010**	**0.011**	**0.015**	0.048
	go	0.092	0.061	0.071	0.068
	oso	0.093	0.060	0.073	0.071
	ir	0.074	0.067	0.069	0.077
$\overline{H} - H$	dv1	**0.000**	$-$**0.004**	$-$**0.001**	$-$**0.016**
	dv2	**0.001**	**0.000**	$-$**0.002**	**0.008**
	var	**0.000**	-0.012	-0.007	-0.025
	gn	0.051	0.052	0.052	0.062
	osn	0.084	0.084	0.084	0.091
	go	-0.010	-0.035	-0.030	-0.044
	oso	**0.003**	$-$**0.023**	-0.014	$-$**0.028**
	ir	-0.009	$-$**0.004**	$-$**0.004**	0.039

(b) $N = 2^{10} + 1$

However these estimators also possess the largest bias. $\widehat{H}_{\text{go}}$ and $\widehat{H}_{\text{oso}}$, the OLS versions of these two estimators behave in a completely different way – they have smaller biases which are comparable to those of the other considered estimators, but this comes at the cost of a heavily increased MSE.

2.2. *Dependance on the Length of the Sample Path*

The second goal of this paper is to compare the behavior of these estimators for different lengths of sample paths as well as to illustrate how the estimators' variances fluctuate as the length of sample paths is increased. Table 2 shows the mean squared errors and the biases for the Hurst index values of 0.65 and 0.85, respectively. Figure 2 boxplots of the estimators for $H = 0.85$.

The first obvious observation is that the bias of $\widehat{H}_{\text{gn}}$ and $\widehat{H}_{\text{osn}}$ increases as the length of sample paths is decreased. $\widehat{H}_{\text{go}}$ and $\widehat{H}_{\text{oso}}$ do not share this property, however their mean squared errors display only minor decreases when longer sample path lengths are taken. The other estimators show a rather regular decrease of their mean squared errors which is further illustrated by Fig. 3 presenting the plots of $\log(SD)$ against $\log(n)$ for $H \in \{0.55, 0.6, \dots, 0.95\}$ where SD denotes the standard deviations.

Figure 3 shows the rate at which the standard deviation decreases as the sample path length is increased. It can be seen that this rate depends on the value of H for all the estimators except $\widehat{H}_{\text{dv2}}$ and $\widehat{H}_{\text{ir}}$. The general trend is that this rate is lower for higher

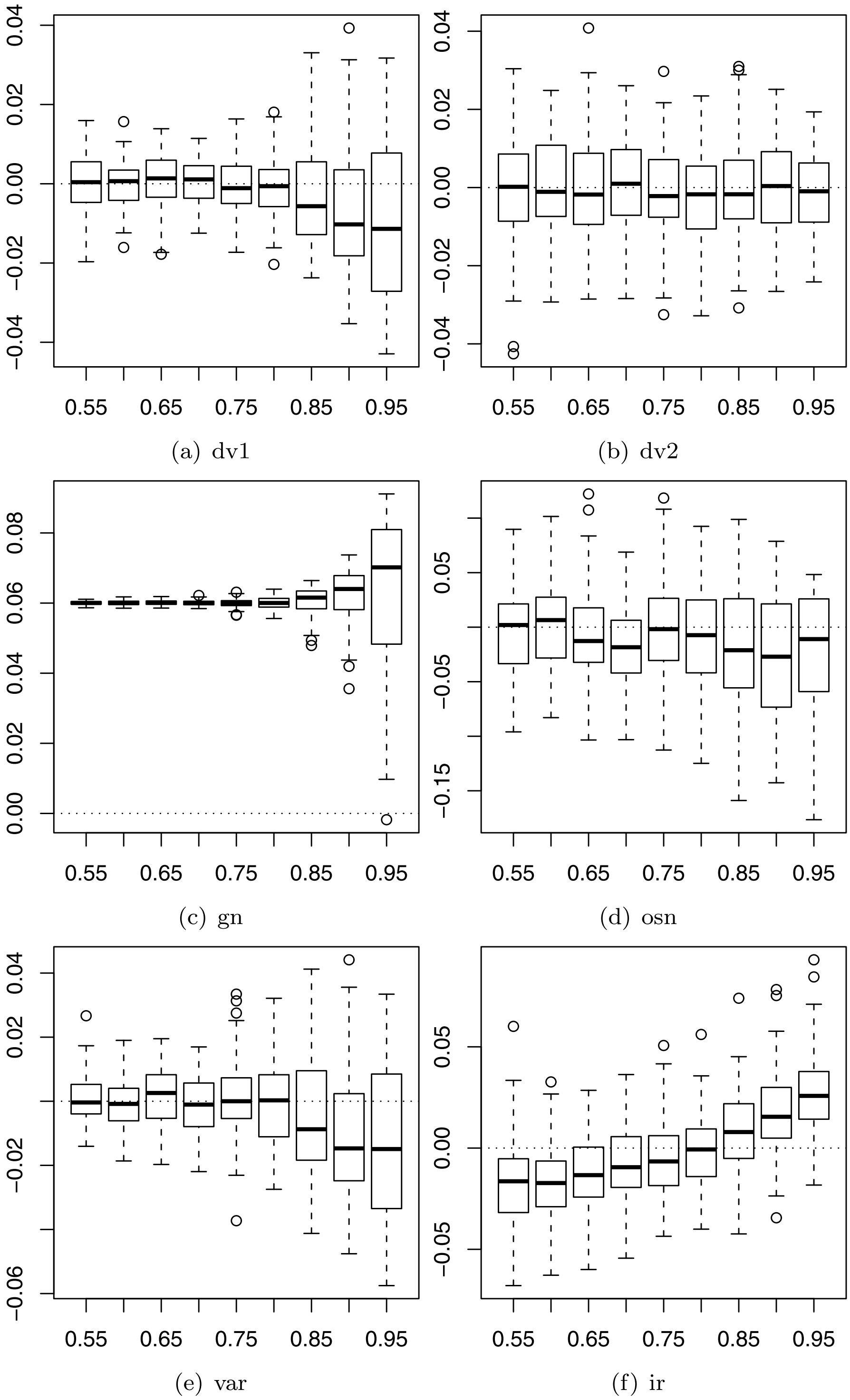
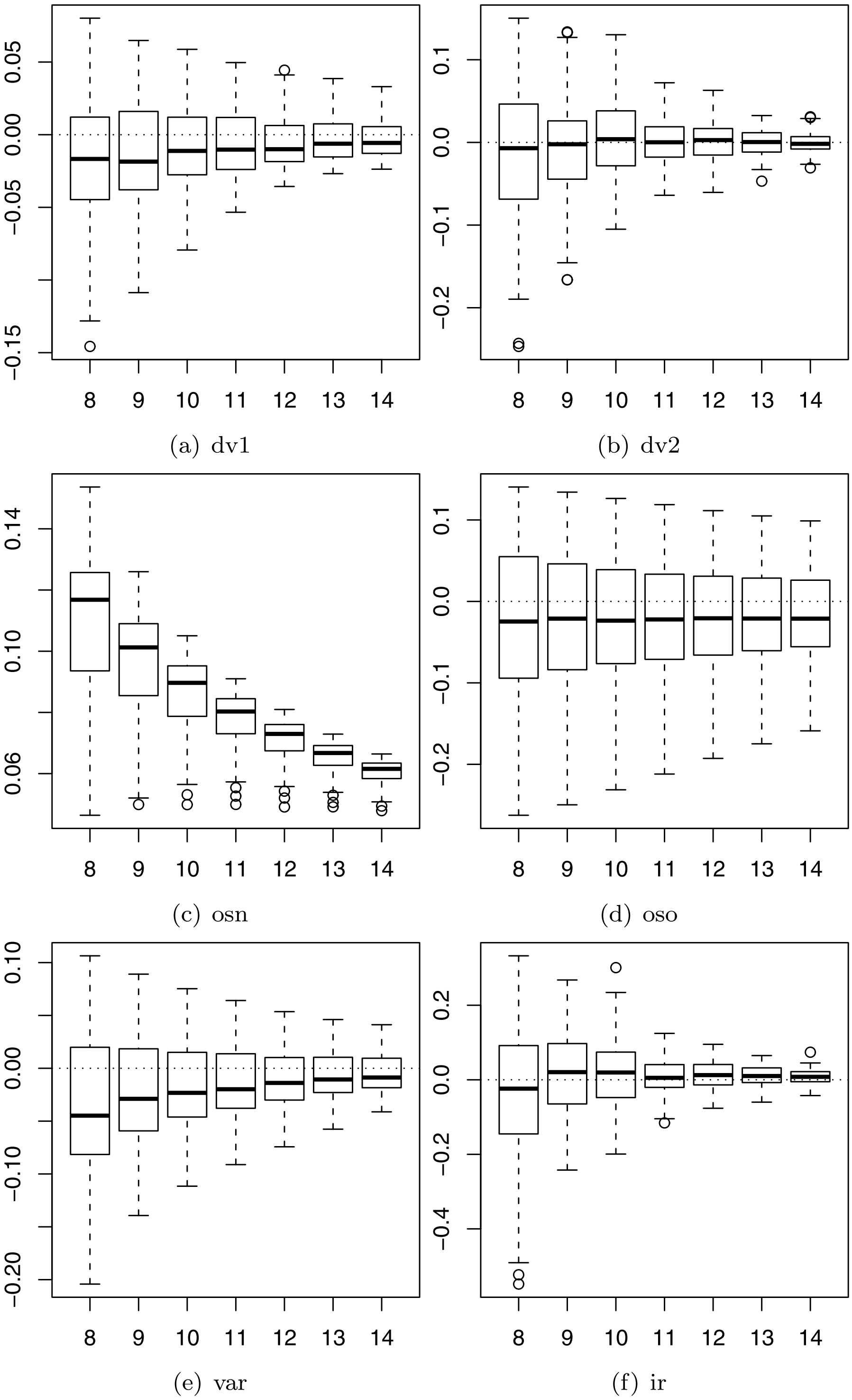


Fig. 1. Boxplots of the estimators for the O-U process of sample path length $n = 2^{14} + 1$.



(a) dv1
(b) dv2
(c) osn
(d) oso
(e) var
(f) ir

Fig. 2. Boxplots of the estimators for the O-U process for $H = 0.85$, $n = 2^k$, $k = 8, \dots, 14$.

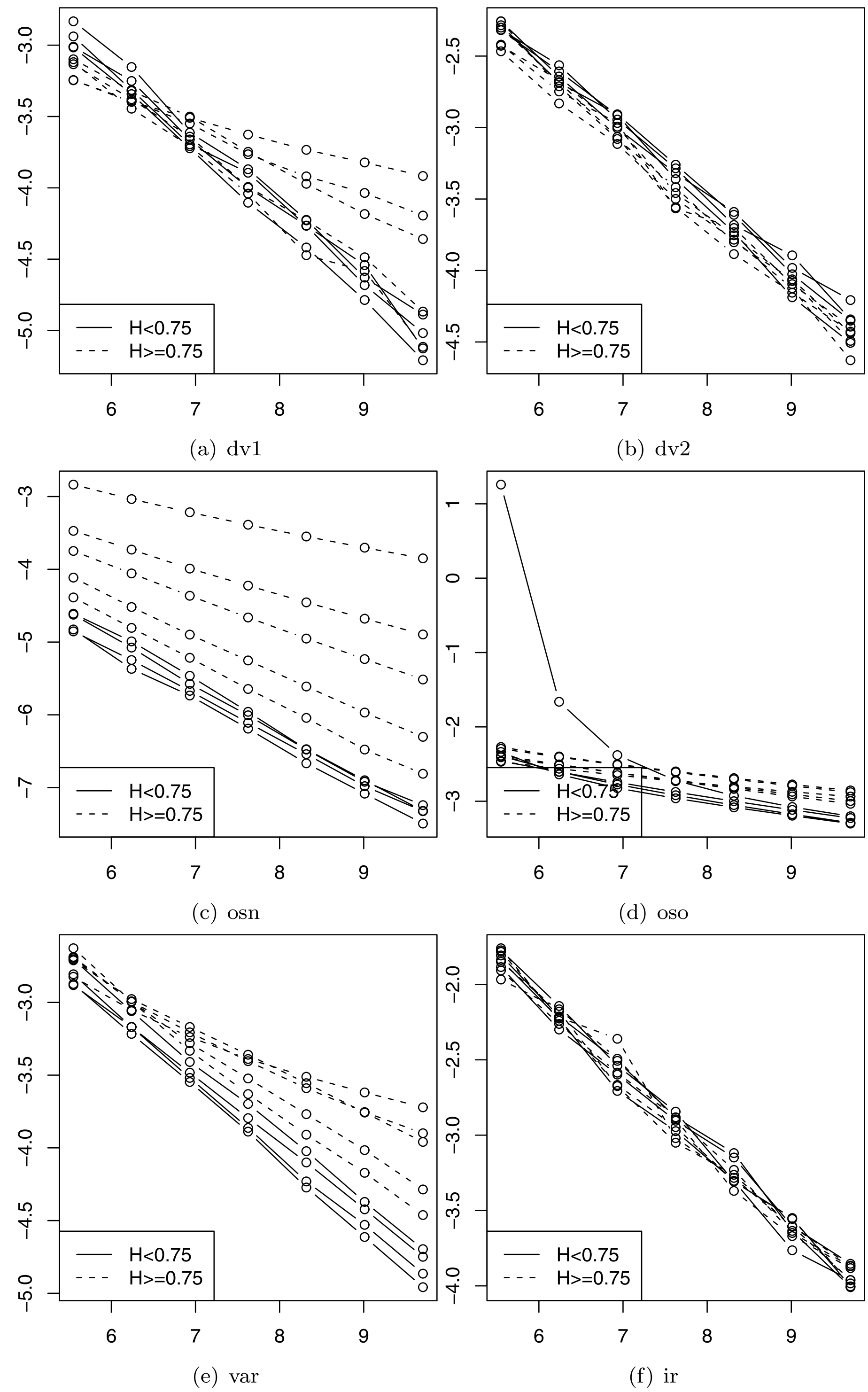


Fig. 3. Dependance of $\log(SD)$ against $\log(n)$ for the O-U process, $H \in \{0.55, 0.6, \ldots, 0.95\}$.

Table 2

Comparison of the estimators for the O-U process for sample path lengths $N = 2^k + 1$

	k	8	10	12	14
MSE	dv1	**0.053**	**0.025**	**0.014**	**0.007**
	dv2	0.102	0.055	0.027	0.013
	var	0.067	0.033	0.018	0.009
	gn	**0.013**	**0.006**	**0.003**	**0.002**
	osn	**0.021**	**0.011**	**0.006**	**0.004**
	go	0.084	0.063	0.049	0.039
	oso	0.086	0.064	0.050	0.040
	ir	0.172	0.079	0.037	0.018
$\overline{H} - H$	dv1	**−0.005**	**0.000**	**0.002**	**0.001**
	dv2	**−0.016**	**0.000**	**0.001**	**0.000**
	var	−0.021	**−0.003**	**−0.001**	**0.001**
	gn	0.064	0.051	0.043	0.037
	osn	0.105	0.084	0.070	0.060
	go	−0.036	−0.028	−0.021	−0.016
	oso	**−0.017**	−0.014	−0.011	−0.009
	ir	−0.043	−0.013	−0.014	−0.013

(a) $H = 0.65$

	k	8	10	12	14
MSE	dv1	**0.045**	**0.030**	**0.019**	0.013
	dv2	0.089	0.051	0.025	**0.012**
	var	0.068	0.042	0.029	0.019
	gn	**0.028**	**0.015**	**0.009**	**0.005**
	osn	**0.036**	**0.020**	**0.012**	**0.008**
	go	0.095	0.077	0.063	0.053
	oso	0.102	0.082	0.067	0.057
	ir	0.170	0.095	0.038	0.021
$\overline{H} - H$	dv1	**−0.016**	**−0.008**	**−0.005**	**−0.003**
	dv2	**−0.015**	**0.005**	**−0.002**	**−0.001**
	var	−0.038	−0.018	**−0.010**	**−0.005**
	gn	0.070	0.054	0.044	0.037
	osn	0.110	0.086	0.071	0.061
	go	−0.054	−0.043	−0.035	−0.028
	oso	−0.033	−0.027	−0.022	−0.018
	ir	−0.039	0.019	0.011	0.008

(b) $H = 0.85$

values of H which is most notable for $\widehat{H}_{\mathrm{gn}}$ and $\widehat{H}_{\mathrm{osn}}$. On the other hand $\widehat{H}_{\mathrm{dv2}}$ and $\widehat{H}_{\mathrm{ir}}$ display no dependance of this kind. Also, if we consider the linear regression $\log(SD) \sim \log(n)$ for these two estimators, its slope is -0.5003 for $\widehat{H}_{\mathrm{dv2}}$ and -0.5013 for $\widehat{H}_{\mathrm{ir}}$, which suggests that for both these estimators $SD(\widehat{H}_{(\cdot)}) \sim \mathcal{O}(n^{-1/2})$.

3. Estimation of the geometric Brownian motion

3.1. *Dependance on the Value of the Hurst Index*

Table 3 presents the mean squared errors and the biases for the sample path lengths of 2^{14} and 2^{10}. Boxplots of these estimators for the sample path length $n = 2^{14}$ can be found in Fig. 4. It can be seen that, for the non-Gaussian gBm process, the estimators $\widehat{H}_{\mathrm{gn}}$ and $\widehat{H}_{\mathrm{osn}}$ display higher biases for all the values of H.

An interesting observation is that, in the case of the O-U process the mean squared errors of $\widehat{H}_{\mathrm{gn}}$ and $\widehat{H}_{\mathrm{osn}}$ were the lowest of all the considered estimators, while for the gBm model their mean squared errors surpassed those of $\widehat{H}_{\mathrm{dv1}}$, $\widehat{H}_{\mathrm{dv2}}$ and $\widehat{H}_{\mathrm{var}}$.

The behavior of $\widehat{H}_{\mathrm{dv2}}$ and $\widehat{H}_{\mathrm{ir}}$ does not display notable differences for these two processes.

Table 3

Comparison of the estimators for the gBm process

H		0.55	0.7	0.8	0.95
MSE	dv1	**0.008**	**0.006**	**0.009**	**0.021**
	dv2	**0.016**	**0.012**	**0.011**	**0.010**
	var	**0.008**	**0.010**	0.015	0.025
	gn	0.032	0.030	0.034	0.049
	osn	0.035	0.032	0.037	0.054
	go	0.033	0.046	0.053	0.049
	oso	0.035	0.049	0.056	0.053
	ir	0.022	0.019	0.019	**0.022**
$\overline{H} - H$	dv1	**0.000**	**0.001**	**0.000**	**−0.006**
	dv2	**0.000**	**0.001**	**−0.002**	**0.000**
	var	**0.000**	**0.001**	**0.000**	**−0.009**
	gn	0.055	0.051	0.061	0.059
	osn	0.080	0.076	0.086	0.082
	go	−0.014	−0.016	−0.021	−0.026
	oso	**−0.003**	−0.006	−0.011	−0.014
	ir	−0.019	−0.009	**−0.001**	0.028

(a) $N = 2^{14} + 1$

H		0.55	0.7	0.8	0.95
MSE	dv1	**0.030**	**0.026**	**0.029**	**0.030**
	dv2	0.054	0.054	0.055	**0.046**
	var	**0.032**	0.039	0.044	0.038
	gn	0.047	0.044	0.051	0.079
	osn	0.053	0.048	0.058	0.088
	go	0.055	0.072	0.080	0.066
	oso	0.057	0.077	0.084	0.070
	ir	0.074	0.067	0.069	0.077
$\overline{H} - H$	dv1	**−0.001**	**−0.001**	**0.000**	**−0.012**
	dv2	**0.001**	**0.000**	**−0.002**	**0.007**
	var	**0.000**	−0.006	−0.008	−0.020
	gn	0.077	0.072	0.086	0.086
	osn	0.112	0.106	0.120	0.116
	go	−0.020	−0.027	−0.036	−0.038
	oso	**−0.001**	−0.009	−0.019	−0.020
	ir	−0.009	−0.002	**−0.004**	0.039

(b) $N = 2^{10} + 1$

3.2. *Dependance on the Length of the Sample Path*

Table 4 presents the mean squared errors and the biases for $H = 0.65$ and $H = 0.85$. Figure 5 shows the boxplots of the estimators considered for $H = 0.85$, while Fig. 6 presents the plots of $\log(SD)$ against $\log(n)$.

Compared to the O-U case the biases of $\widehat{H}_{\mathrm{gn}}$ and $\widehat{H}_{\mathrm{osn}}$ are higher for all sample path lengths. The mean squared errors of $\widehat{H}_{\mathrm{gn}}$, $\widehat{H}_{\mathrm{osn}}$, $\widehat{H}_{\mathrm{go}}$ and $\widehat{H}_{\mathrm{oso}}$ are higher for all sample path lengths. In the case of relatively short sample paths ($2^8 - 2^{10}$) and $H > 3/4$, the estimators $\widehat{H}_{\mathrm{go}}$ and $\widehat{H}_{\mathrm{oso}}$ have at times severely overestimated the Hurst index H with the estimated value being higher than 2. Those values were excluded from their boxplots. The slope of the linear regression $\log(SD) \sim \log(n)$ is -0.5015 for $\widehat{H}_{\mathrm{dv2}}$ and -0.5011 for $\widehat{H}_{\mathrm{ir}}$, which does not differ significantly from the O-U case.

4. Conclusions

- The estimators $\widehat{H}_{\mathrm{gn}}$ and $\widehat{H}_{\mathrm{osn}}$, despite showing the least mean squared errors in the O-U case, have also shown much higher biases than other estimators considered in this paper. This bias increases as the sample path length is decreased but shows no dependance on the value of the Hurst index H as long as H is not too close to 1. When $H > 0.9$, this bias increases further. In the gBm case the mean squared errors of these two estimators were greater than those of $\widehat{H}_{\mathrm{dv1}}$, $\widehat{H}_{\mathrm{dv2}}$ and $\widehat{H}_{\mathrm{var}}$.

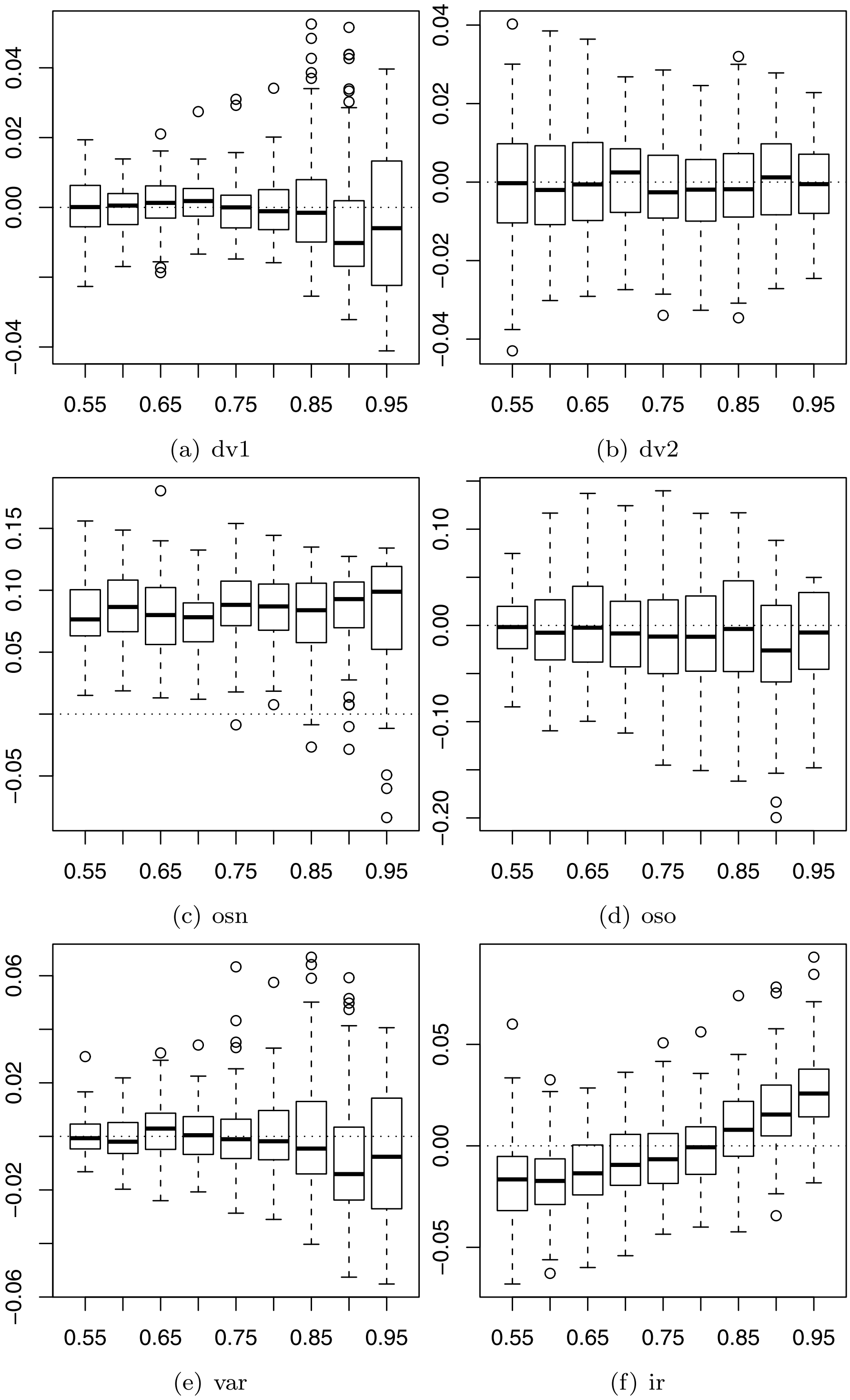


Fig. 4. Boxplots of the estimators for the gBm process for sample path length $n = 2^{14} + 1$.

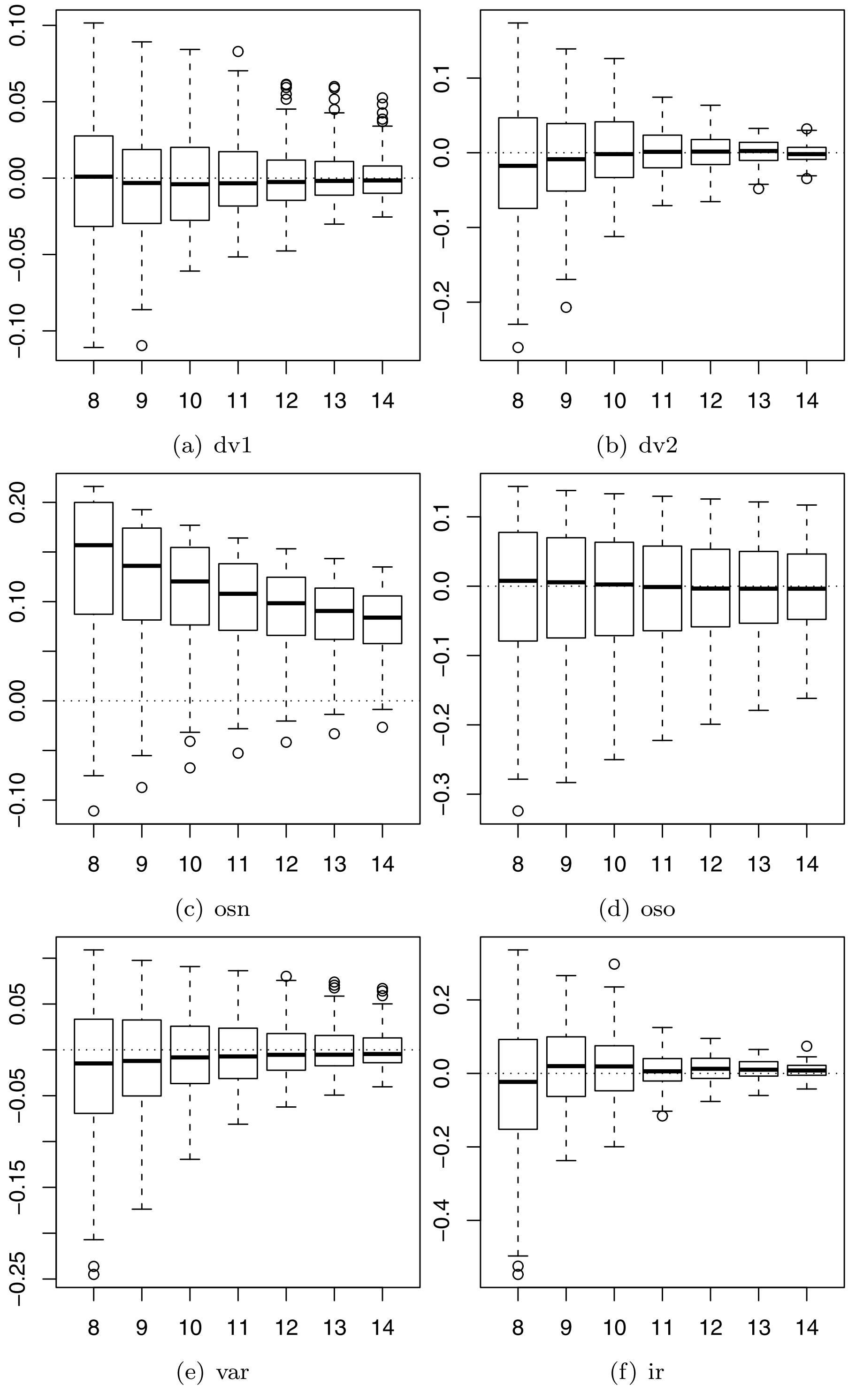
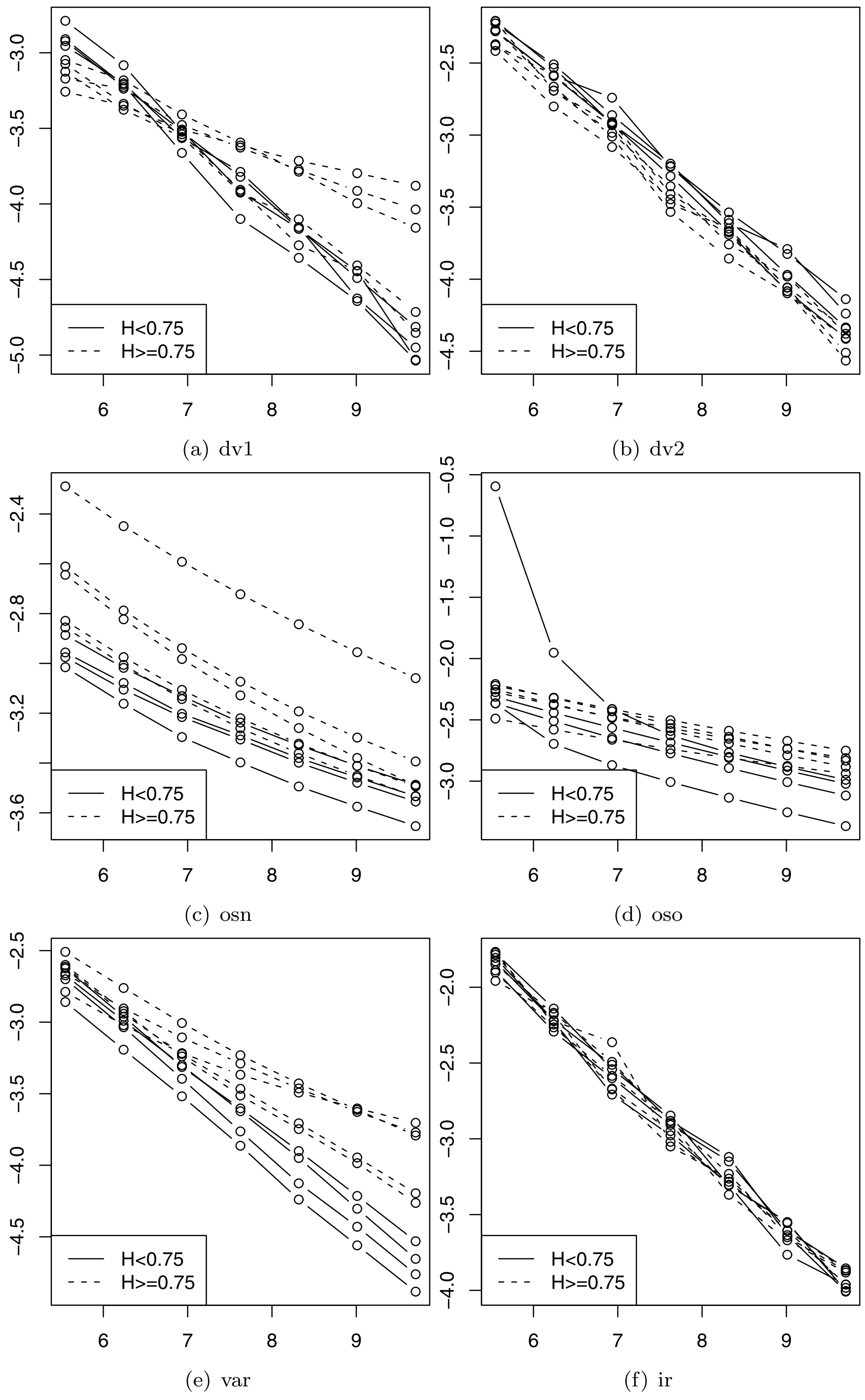


Fig. 5. Boxplots of the estimators for the gBm process for $H = 0.85$, $n = 2^k$, $k = 8, \dots, 14$.



(a) dv1 (b) dv2
(c) osn (d) oso
(e) var (f) ir

Fig. 6. Dependance of $\log(SD)$ against $\log(n)$ for the gBm process, $H \in \{0.55, 0.6, \ldots, 0.95\}$.

Table 4

Comparison of the estimators for the gBm process for sample path lengths $N = 2^k + 1$

(a) $H = 0.65$

	k	8	10	12	14
MSE	dv1	**0.052**	**0.028**	**0.016**	**0.007**
	dv2	0.103	0.065	**0.028**	0.013
	var	0.073	**0.036**	**0.020**	**0.011**
	gn	**0.066**	0.050	0.041	0.034
	osn	0.075	0.056	0.044	0.037
	go	2.411	0.097	0.061	0.048
	oso	0.556	0.090	0.063	0.050
	ir	0.172	0.079	0.037	0.018
$\overline{H} - H$	dv1	**−0.002**	**0.001**	**0.002**	**0.001**
	dv2	−0.020	**−0.002**	**0.000**	**0.000**
	var	−0.013	**0.001**	**0.001**	**0.002**
	gn	0.093	0.074	0.062	0.054
	osn	0.137	0.110	0.092	0.079
	go	0.223	−0.009	−0.011	−0.010
	oso	0.065	0.006	**0.001**	**0.000**
	ir	−0.043	−0.013	−0.014	−0.013

(b) $H = 0.85$

	k	8	10	12	14
MSE	dv1	**0.047**	**0.033**	**0.023**	**0.016**
	dv2	0.094	**0.053**	**0.026**	**0.013**
	var	0.082	**0.050**	**0.032**	0.023
	gn	0.082	0.059	0.046	0.037
	osn	0.092	0.065	0.049	0.040
	go	0.104	0.085	0.071	0.060
	oso	0.109	0.090	0.075	0.064
	ir	0.169	0.095	0.038	0.021
$\overline{H} - H$	dv1	**−0.005**	**−0.001**	**0.001**	**0.000**
	dv2	−0.016	**0.003**	**−0.002**	**−0.001**
	var	−0.025	−0.007	**−0.002**	**0.001**
	gn	0.094	0.075	0.062	0.054
	osn	0.136	0.108	0.091	0.078
	go	−0.038	−0.028	−0.022	−0.017
	oso	−0.011	−0.009	−0.006	−0.004
	ir	−0.039	0.019	0.011	0.008

- The estimators $\widehat{H}_{\text{go}}$ and $\widehat{H}_{\text{oso}}$, the ordinary least squares versions of the previous estimators, display totally different behavior – their biases are comparable to those of the other estimators. However, their mean squared errors are considerably higher than those of other estimators and tend to decrease only slightly as the sample path length is increased. Additionally, both of these estimators require the sample path length to be equal to $2^k + 1$, $k \in \mathbb{N}$, which means that, for sample paths of different length, some of the observations must be truncated.
- The estimators $\widehat{H}_{\text{dv1}}$ and $\widehat{H}_{\text{var}}$ behaved differently for "small" and "large" values of H. As $H \in (1/2; 3/4)$, they displayed the best characteristics while for higher values of H their performance was close to or worse than that of other estimators. $\widehat{H}_{\text{var}}$ displayed increased biases for shorter sample paths.
- The characteristics of $\widehat{H}_{\text{dv2}}$ were slightly worse than those of $\widehat{H}_{\text{dv1}}$ and $\widehat{H}_{\text{var}}$ for shorter sample paths and $H < 3/4$, and they were similar or better for longer sample paths and $H > 3/4$. Additionally, this estimator showed no notable dependance of its behavior on the value of H. $\widehat{H}_{\text{ir}}$ displayed such a dependance only for rather long sample paths, but its biases and mean squared errors were higher. Having considered the linear regression $\log(SD) \sim \log(n)$ for these two estimators, the results suggest that for both these estimators $SD(\widehat{H}_{(\cdot)}) \sim \mathcal{O}(n^{-1/2})$.
- Calculation times for the estimators $\widehat{H}_{\text{dv1}}$, $\widehat{H}_{\text{dv2}}$ and $\widehat{H}_{\text{osn}}$ were about $0.02s$ with 100 sample paths of the length $N = 2^8 + 1$ and about $0.4s$ with 100 sample paths

of the length $N = 2^{14} + 1$. Calculation times of $\widehat{H}_{\mathrm{gn}}$ were about twice lower and those of $\widehat{H}_{\mathrm{go}}$, $\widehat{H}_{\mathrm{oso}}$ and $\widehat{H}_{\mathrm{ir}}$ were 2–5 times higher.

As a conclusion, the results of this study suggest that when there's a reason to expect the Hurst index to be high or when the Hurst index is estimated from a sufficiently long sample path, the $\widehat{H}_{\mathrm{dv2}}$ estimator performs best. If either of these assumptions is not present, then the $\widehat{H}_{\mathrm{dv1}}$ and $\widehat{H}_{\mathrm{var}}$ would likely give a more precise estimate.

References

Bardet, J.-M. *et al.* (2003). Long-range dependence. In: Doukham, P. (Ed.), *Theory and Applications of Long-Range Dependence*, Birkhäuser, Basel, pp. 579–625.

Begyn, A. (2005). Quadratic variations along irregular subdivisions for Gaussian processes. *Electronic Journal of Probability*, 10, 691–717.

Begyn, A. (2006). *Generalized quadratic variations of gaussian processes: limit theorems and applications to fractional processes*. PhD thesis.

Berzin, C., León, J. (2008). Estimation in models driven by fractional Brownian motion. *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, 44(2), 191–213.

Coeurjolly, J.F. (2000). Simulation and identification of the fractional Brownian motion: a bibliographical and comparative study. *Journal of Statistical Software*, 50 (7).

Chronopoulou, A., Viens, F.G. (2010). Hurst index estimation for self-similar processes with long-memory. In: *Recent Advances in Stochastic Dynamics and Stochastic Analysis* (in press), World Scientific, Singapore.

Davis, G. (1996). Implicit image models in fractal image compression. In: *Proceedings of the SPIE Conference on Wavelet Applications in Signal and Image Processing IV*.

Dudley, R.M. (1999). Picard iteration and p-variation: the work of Lyons. In: *Mini-Proceedings: Workshop on Product Integrals and Pathwise Integration*, MaPhySto, Aarhus.

Gladyshev, E.G. (1963). A new limit theorem for stochastic processes with Gaussian increments. *Theory Probability and its Applications*, 6(1), 52–61.

Istas, J., Lang, G. (1997). Quadratic variations and estimation of the local Hölder index of a Gaussian process. *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, 33, 407–436.

Kubilius, K. (2000). The existence and uniqueness of the solution of the integral equation driven by the fractional Brownian motion. *Lithuanian Mathematical Journal*, 40, 104–110.

Kubilius, K., Melichov, D. (2008). On estimation of the Hurst index of solutions of stochastic integral equations. *Lithuanian Mathematical Journal*, 48–49, 401–406.

Kubilius, K., Melichov, D. (2009). Estimating the Hurst index of the solution of a stochastic integral equation. *Lithuanian Mathematical Journal*50, 24–29.

Kubilius, K., Melichov, D. (2010). Quadratic variations and estimation of the Hurst index of the solution of SDE driven by a fractional Brownian motion. *Lithuanian Mathematical Journal*, 50(4), 401–417.

Li, M. (2010). Essay on teletraffic models (I). In: *Proceedings of the 9th WSEAS International Conference on Applied Computer and Applied Computational Science*, pp. 130–135.

Lyons, T. (1994). Differential equations driven by rough signals (I): an extension of an inequality of L.C. Young. *Mathematical Research Letters*, 1, 451–464.

Norvaiša, R., Salopek, D.M. (2002). Estimating the p-variation index of a sample function: an application to financial data set. *Methodology and Computing in Applied Probability*, 4(1), 27–53.

Nualart, D., Răşcanu, A. (2002). Differential equations driven by fractional Brownian motion. *Collectanea Mathematica*, 53 (1), 55–81.

R Development Core Team. (2009). *R: A Language and Environment for Statistical Computing*. http://www.R-project.org.

Wenlu, X., Weixin, X. (1997). Image object detection based on fractional Brownian motion. *Journal of Electronics*, 14(4), 289–294.

K. Kubilius graduated from the Vilnius University in 1976. He received the PhD degree from Vilnius University (1981), the dr. habil. from the Vilnius University Institute of Mathematics and Informatics (1993). He was conferred the title of professor at Vilnius Gediminas Technical University in 2000. Currently, he is a senior researcher in the Stochastic Process Department in the Vilnius University Institute of Mathematics and Informatics. His research interests include stochastic processes, stochastic differential equations and their modeling, limit theorems.

D. Melichov graduated from the Vilnius Gediminas Technical University in 2007 receiving the master of science of statistics degree. He is a doctoral student at the Vilnius Gediminas Technical University. His PhD research topic is asymptotic properties and parameter estimation of the solutions of stochastic differential equations driven by the fractional Brownian motion.

Apie stochastinių diferencialinių lygčių, valdomų trupmeninio Brauno judesio, sprendinių Hursto indekso įvertinių palyginimą

Kęstutis KUBILIUS, Dmitrij MELICHOV

Straipsnyje pateikiamas Hursto indekso įvertinių tyrimas trupmeniniams Ornštein–Ulenbeck ir Black–Scholes modeliams. Nagrinėjama įvertinių skaitinių charakteristikų priklausomybė tiek nuo Hursto indekso reikšmės, tiek nuo trajektorijos ilgio.