LARGE SAMPLE BEHAVIOR OF SOME WELL-KNOWN ROBUST ESTIMATORS UNDER LONG-RANGE DEPENDENCE

C. LÉVY-LEDUC, H. BOISTARD, E. MOULINES, M. S. TAQQU, AND V. A. REISEN

Abstract. The paper concerns robust location and scale estimators under long-range dependence, focusing on the Hodges-Lehmann location estimator, on the Shamos-Bickel scale estimator and on the Rousseeuw-Croux scale estimator. The large sample properties of these estimators are reviewed. The paper includes computer simulation in order to examine how well the estimators perform at finite sample sizes.

1. Introduction

This paper has two parts. The first is a review of the asymptotic theory behind robust location and scale estimators under long-range dependence. The second part involves computer simulation to see how well the methods perform at finite sample sizes. We focus on the Hodges-Lehmann location estimator [Hodges and Lehmann (1963)], the Shamos-Bickel scale estimator [Shamos (1976), Bickel and Lehmann (1979)] and the Rousseeuw-Croux scale estimator [Rousseeuw and Croux (1993)]. All of these estimators share the following property: they can be written as empirical quantiles of U-processes defined by

$$U_n(r) = \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} \mathbb{1}_{\{G(X_i,X_j) \leq r\}} , \quad r \in I , \tag{1}$$

where I is an interval included in $\mathbb{R}$, G is a symmetric function *i.e.* $G(x,y) = G(y,x)$ for all x, y in $\mathbb{R}$, and the process $(X_i)_{i \geq 1}$ satisfies:

(A1) $(X_i)_{i\geq 1}$ is a stationary mean-zero Gaussian process with covariances $\rho(k) = \mathbb{E}(X_1 X_{k+1})$ satisfying:

$$\rho(0) = 1 \text{ and } \rho(k) = k^{-D} L(k),\ 0 < D < 1 ,$$

where L is slowly varying at infinity and is positive for large k.

For notational convenience, we shall denote by $h(\cdot, \cdot, r)$ the kernel on which the U-process U_n is based, that is,

$$h(x, y, r) = \mathbb{1}_{\{G(x,y) \leq r\}} , \forall x, y \in \mathbb{R} \text{ and } r \in I . \tag{2}$$

We are interested in the asymptotic behavior of the quantile $U_n^{-1}(p)$, $p \in [0, 1]$, suitably normalized. The limit may be Gaussian or not. This will depend on the range of D and on a parameter m called the Hermite rank. Theorem 1 below describes the case $D > 1/m$ and

Date: November 9, 2010.

Key words and phrases. Long-range dependence, U-process, Hodges-Lehmann location estimator, Bickel-Shamos scale estimator, Croux-Rousseeuw scale estimator, Ma-Genton autocovariance estimator.

$m = 2$ and Theorem 2 describes the case $D < 1/m$ and $m = 1$ or 2. The applications we consider involve only the cases $m = 1$ and 2.

2. Asymptotic behavior of empirical quantiles

Let us first review some classical tools for the study of the asymptotic behavior of U-statistics and specifically for U-statistics constructed from Gaussian observations. These are the Hermite rank of a class of functions and the Hoeffding decomposition given in (8) below.

We start by recalling the definition of the Hermite rank of the class of functions $\{h(\cdot,\cdot,r) - U(r), r \in I\}$ which plays a crucial role in understanding the asymptotic behavior of empirical quantiles of the U-process $U_n(\cdot)$. The function $\{U(r), r \in I\}$ is defined below. We shall expand the kernel function $(x,y) \mapsto h(x,y,r)$ in a Hermite polynomial basis. We use Hermite polynomials with leading coefficients equal to one which are: $H_0(x) = 1$, $H_1(x) = x$, $H_2(x) = x^2 - 1$, $H_3(x) = x^3 - 3x, \ldots$. We get

$$h(x,y,r) = \sum_{p,q \geq 0} \frac{\alpha_{p,q}(r)}{p!q!} H_p(x) H_q(y) \;, \text{ for all } x,y \text{ in } \mathbb{R} \;, \tag{3}$$

where

$$\alpha_{p,q}(r) = \mathbb{E}[h(X,Y,r) H_p(X) H_q(Y)] \;, \tag{4}$$

and where (X,Y) is a *standard Gaussian vector* that is X and Y are independent standard Gaussian random variables.

The constant term in the Hermite decomposition is given by $\alpha_{0,0}(r)$: as we shall see below, it is the non-random limit of $U_n(r)$, as n tends to infinity. This is why we set:

$$U(r) = \alpha_{0,0}(r) = \int_{\mathbb{R}^2} h(x,y,r)\varphi(x)\varphi(y) \mathrm{d}x \mathrm{d}y \;, \text{ for all } r \text{ in } I \;, \tag{5}$$

where φ denotes the p.d.f of a standard Gaussian random variable.

Consider now the terms with $p + q > 0$. The Hermite rank of the function $h(\cdot,\cdot,r)$ is the smallest positive integer $m(r)$ such that there exist p and q satisfying $p + q = m(r)$ and $\alpha_{p,q}(r) \neq 0$. Thus, (3) can be rewritten as

$$h(x,y,r) - U(r) = \sum_{\substack{p,q \geq 0 \\ p+q \geq m(r)}} \frac{\alpha_{p,q}(r)}{p!q!} H_p(x) H_q(y) \;. \tag{6}$$

The Hermite rank m of the class of functions $\{h(\cdot,\cdot,r) - U(r) \;, r \in I\}$ is the smallest index $m = p + q \geq 1$ such that $\alpha_{p,q}(r) \neq 0$ for at least one r in I, that is, $m = \inf_{r \in I} m(r)$.

In the sequel, we shall assume that

$$m = 1 \text{ or } 2$$

since this covers the specific estimators we are interested in.

Having defined the Hermite rank, we now turn to the so-called "Hoeffding's decomposition" [Hoeffding (1948)] which is one of the main tools used in the proof of Theorem 1 below. The

Hoeffding decomposition amounts to decomposing, for all r in I, the difference

$$U_n(r) - U(r) = \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} [h(X_i, X_j, r) - U(r)] \,, \tag{7}$$

into two parts, as

$$U_n(r) - U(r) = W_n(r) + R_n(r) \,, \tag{8}$$

where

(9)

$$W_n(r) = \frac{1}{n}\sum_{i=1}^{n}\{h_1(X_i, r) - U(r)\} + \frac{1}{n}\sum_{j=1}^{n}\{h_1(X_j, r) - U(r)\} = \frac{2}{n}\sum_{i=1}^{n}\{h_1(X_i, r) - U(r)\} \;,$$

and

$$R_n(r) = \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} \{h(X_i, X_j, r) - h_1(X_i, r) - h_1(X_j, r) + U(r)\} \;. \tag{10}$$

The function $h_1(x, r)$ which is added in (9) and subtracted in (10) is defined for all x in $\mathbb{R}$ and r in I as

$$h_1(x, r) = \int_{\mathbb{R}} h(x, y, r)\varphi(y)\mathrm{d}y \,. \tag{11}$$

We shall focus on the empirical quantile $U^{-1}(p)$, $p \in [0, 1]$. Recall that if $V : I \longrightarrow [0, 1]$ is a non-decreasing cadlag function, where I is an interval of $\mathbb{R}$, then its generalized inverse V^{-1} is defined by $V^{-1}(p) = \inf\{r \in I,\ V(r) \geq p\}$. This applies to both $U_n(r)$ and $U(r)$ since these are non-decreasing functions of r.

Theorem 1, stated below, gives the asymptotic behavior of the empirical quantile $U_n^{-1}(\cdot)$ in the case where

$$D > 1/m \text{ and } m = 2 \,.$$

For a proof of Theorem 1, we refer the reader to the proofs of Theorem 1 and Corollary 3 in Lévy-Leduc et al. (2009).

Theorem 1. *Let I be a compact interval of $\mathbb{R}$. Let p be a fixed real number in $[0, 1]$. Suppose that there exists some r in I such that $U(r) = p$, that U is differentiable at r and that $U'(r)$ is non null. Assume that the Hermite rank of the class of functions $\{h(\cdot, \cdot, r) - U(r)\ , r \in I\}$ as defined in (6) is $m = 2$ and that Assumption (A1) is satisfied with $D > 1/2$. Assume that h and h_1, defined in (2) and (11), satisfy the three following conditions:*

(i) There exists a positive constant C such that for all s, t in I, u, v in $\mathbb{R}$,

$$\mathbb{E}[|h(X + u, Y + v, s) - h(X + u, Y + v, t)|] \leq C|t - s| \,, \tag{12}$$

where (X, Y) is a standard Gaussian vector.

(ii) There exists a positive constant C such that for all $k \geq 1$,

$$\mathbb{E}[|h(X_1 + u, X_{1+k} + v, t) - h(X_1, X_{1+k}, t)|] \leq C(|u| + |v|) \,, \tag{13}$$

$$\mathbb{E}[|h(X_1, X_{1+k}, s) - h(X_1, X_{1+k}, t)|] \leq C|t - s| \,. \tag{14}$$

(iii) There exists a positive constant C such that for all t, s in I, and x, u, v in $\mathbb{R}$,

$$|h_1(x+u,t) - h_1(x+v,t)| \leq C(|u|+|v|) \;, \tag{15}$$

and

$$|h_1(x,s) - h_1(x,t)| \leq C|t-s| \;. \tag{16}$$

Then, as n tends to infinity,

$$\sqrt{n}(U_n^{-1}(p) - U^{-1}(p)) \xrightarrow{d} -W(U^{-1}(p))/U'(U^{-1}(p)) \;,$$

where $\{W(r), r \in I\}$ is a zero mean Gaussian process with covariance structure given by

$$\begin{aligned}\mathbb{E}[W(s)W(t)] = & \; 4\,\mathrm{Cov}(h_1(X_1,s),h_1(X_1,t)) \\ & + 4\sum_{\ell\geq 1}\{\mathrm{Cov}(h_1(X_1,s),h_1(X_{\ell+1},t)) + \mathrm{Cov}(h_1(X_1,t),h_1(X_{\ell+1},s))\} \;. \end{aligned} \tag{17}$$

To study the case $D < 1/m$ and $m = 1$ or 2, we do not use the Hoeffding decomposition as in Theorem 1. We use instead a different decomposition of $U_n(\cdot)$ based on the expansion of h in the basis of Hermite polynomials given by (3). Thus, $U_n(r)$ defined in (1) can be rewritten as follows

$$n(n-1)\{U_n(r) - U(r)\} = \widetilde{W}_n(r) + \widetilde{R}_n(r) \;, \tag{18}$$

where

$$\widetilde{W}_n(r) = \sum_{1\leq i\neq j\leq n}\sum_{\substack{p,q\geq 0\\ p+q\leq m}} \frac{\alpha_{p,q}(r)}{p!q!} H_p(X_i)H_q(X_j) \tag{19}$$

and

$$\widetilde{R}_n(r) = \sum_{1\leq i\neq j\leq n}\sum_{\substack{p,q\geq 0\\ p+q> m}} \frac{\alpha_{p,q}(r)}{p!q!} H_p(X_i)H_q(X_j) \;. \tag{20}$$

Introduce also the Beta function

$$\mathrm{B}(\alpha,\beta) = \int_0^\infty y^{\alpha-1}(1+y)^{-\alpha-\beta}\mathrm{d}y = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \;, \quad \alpha > 0,\ \beta > 0 \;. \tag{21}$$

The limit processes which appear in the next theorem are the standard fractional Brownian motion (fBm) $(Z_{1,D}(t))_{0\leq t\leq 1}$ and the Rosenblatt process $(Z_{2,D}(t))_{0\leq t\leq 1}$. They are defined through multiple Wiener-Itô integrals and given by

$$Z_{1,D}(t) = \int_{\mathbb{R}}\left[\int_0^t (u-x)_+^{-(D+1)/2}\mathrm{d}u\right]\mathrm{d}B(x), \quad 0 < D < 1 \;, \tag{22}$$

and

$$Z_{2,D}(t) = \int'_{\mathbb{R}^2}\left[\int_0^t (u-x)_+^{-(D+1)/2}(u-y)_+^{-(D+1)/2}\mathrm{d}u\right]\mathrm{d}B(x)\mathrm{d}B(y), \quad 0 < D < 1/2 \;, \tag{23}$$

where B is the standard Brownian motion, see Fox and Taqqu (1987). The symbol $\int'$ means that the domain of integration excludes the diagonal. The following theorem treats the case

$$D < 1/m \text{ where } m = 1 \text{ or } 2 \;.$$

Theorem 2. *Let I be a compact interval of $\mathbb{R}$. Let p be a fixed real number in $[0,1]$. Suppose that there exists some r in I such that $U(r) = p$, that U is differentiable at r and that $U'(r)$ is non null. Assume that Assumption (A1) holds with $D < 1/m$, where $m = 1$ or 2 is the Hermite rank of the class of functions $\{h(\cdot,\cdot,r) - U(r)\ , r \in I\}$ as defined in (6). Assume also the following:*

(i) There exists a positive constant C such that, for all $k \geq 1$ and for all s, t in I,

$$\mathbb{E}[|h(X_1, X_{1+k}, s) - h(X_1, X_{1+k}, t)|] \leq C|t - s| \ . \tag{24}$$

(ii) U is a Lipschitz function.

(iii) The function $\widetilde{\Lambda}$ defined, for all s in I, by

$$\widetilde{\Lambda}(s) = \mathbb{E}[h(X, Y, s)(|X| + |XY| + |X^2 - 1|)] \ , \tag{25}$$

where X and Y are independent standard Gaussian random variables, is also a Lipschitz function.

Then, as n tends to infinity,

$$\frac{n^{mD/2}}{L(n)^{m/2}}(U_n^{-1}(p) - U^{-1}(p))$$

converges in distribution to

$$-2k(D)^{-1/2}\frac{\alpha_{1,0}(U^{-1}(p))}{U'(U^{-1}(p))}Z_{1,D}(1) \ , \text{ if } m = 1 \ ,$$

and to

$$-k(D)^{-1}\left\{\alpha_{1,1}(U^{-1}(p))Z_{1,D}(1)^2 + \alpha_{2,0}(U^{-1}(p))Z_{2,D}(1)\right\}/U'(U^{-1}(p)) \ , \text{ if } m = 2 \ ,$$

where the fractional Brownian motion $Z_{1,D}(\cdot)$ and the Rosenblatt process $Z_{2,D}(\cdot)$ are defined in (22) and (23) respectively,

$$k(D) = \mathrm{B}((1 - D)/2, D) \ , \tag{26}$$

where B is the Beta function defined in (21), and $\alpha_{p,q}(\cdot)$ is defined in (4).

For a proof of Theorem 2, we refer the reader to the proofs of Theorem 2 and Corollary 4 in Lévy-Leduc et al. (2009).

3. Applications

We use the results established in Section 2 to study the asymptotic properties of several robust estimators based on empirical quantiles of U-processes in the long-range dependence setting.

3.1. **Hodges-Lehmann robust location estimator.** To estimate the location parameter θ of a long-range dependent Gaussian process $(Y_i)_{i\geq 1}$ satisfying $Y_i = \theta + X_i$ where $(X_i)_{i\geq 1}$ satisfy Assumption (A1), Hodges and Lehmann (1963) suggest using

$$\hat{\theta}_{HL} = \text{median}\left\{\frac{Y_i + Y_j}{2}; 1 \leq i < j \leq n\right\} = \theta + \text{median}\left\{\frac{X_i + X_j}{2}; 1 \leq i < j \leq n\right\} . \tag{27}$$

Thus, $\hat{\theta}_{HL}$ may be expressed as

$$\hat{\theta}_{HL} = \theta + U_n^{-1}(1/2) ,$$

where $U_n(\cdot)$ is defined by (1) with $G(x, y) = (x + y)/2$ and satisfies the following proposition.

Proposition 3. *Under Assumption (A1), the Hodges-Lehmann location estimator $\hat{\theta}_{HL}$ defined in (27) from $Y_1, \ldots, Y_n$ satisfies*

$$n^{D/2} L(n)^{-1/2} (\hat{\theta}_{HL} - \theta) \xrightarrow{d} k(D)^{-1/2} Z_{1,D}(1) , \tag{28}$$

where $k(D)^{-1/2} Z_{1,D}(1)$ is a zero-mean Gaussian random variable with variance $2(-D + 1)^{-1}(-D + 2)^{-1}$.

Moreover, it converges to θ at the same rate as the sample mean $\bar{Y}_n = n^{-1} \sum_{i=1}^n Y_i$ with the same limiting distribution. There is no loss of efficiency.

The proof of Proposition 3 is a consequence of Theorem 2. For further details, we refer the reader to Section 4.1 of Lévy-Leduc et al. (2009).

3.2. **Shamos-Bickel robust scale estimator.** To estimate the scale parameter σ of a long-range dependent Gaussian process $(Y_i)_{i\geq 1}$ satisfying $Y_i = \sigma X_i$ where $(X_i)_{i\geq 1}$ satisfy Assumption (A1), Shamos (1976) and Bickel and Lehmann (1979) propose to use

$$\hat{\sigma}_{SB} = b \,\text{median}\{|Y_i - Y_j|; 1 \leq i < j \leq n\} = b\, \sigma \,\text{median}\{|X_i - X_j|; 1 \leq i < j \leq n\} , \tag{29}$$

where $b = 1/(\sqrt{2}\Phi^{-1}(3/4)) = 1.0483$ to achieve consistency for σ in the case of Gaussian distributions.

Thus, $\hat{\sigma}_{SB}$ may be expressed as

$$\hat{\sigma}_{SB} = b\sigma U_n^{-1}(1/2) ,$$

where $U_n(\cdot)$ is defined by (1) with $G(x, y) = |x - y|$ and satisfies the following proposition.

Proposition 4. *Under Assumption (A1), the Shamos-Bickel robust scale estimator $\hat{\sigma}_{SB}$ defined in (29) from $Y_1, \ldots, Y_n$ has the following asymptotic behavior:*

(i) If $1/2 < D < 1$,

$$\sqrt{n}(\hat{\sigma}_{SB} - \sigma) \xrightarrow{d} \mathcal{N}(0, \bar{\sigma}^2) \tag{30}$$

where

$$\bar{\sigma}^2 = \frac{2b^2\sigma^2}{\varphi^2(1/(b\sqrt{2}))} \left[\text{Var}(h_1(Y_1/\sigma, 1/b)) + 2\sum_{k\geq 1} \text{Cov}(h_1(Y_1/\sigma, 1/b), h_1(Y_{k+1}/\sigma, 1/b))\right]$$

and h_1 is given by

$$h_1(x,r) = \int_{\mathbb{R}} \mathbb{1}_{\{|x-y|\le r\}}\varphi(y)\mathrm{d}y = \Phi(x+r) - \Phi(x-r) \;,$$

Φ being the c.d.f of a standard Gaussian random variable.

(ii) If $0 < D < 1/2$,

$$k(D)n^D L(n)^{-1}(\hat{\sigma}_{SB} - \sigma) \xrightarrow{d} \frac{\sigma}{2}(Z_{2,D}(1) - Z_{1,D}(1)^2) \;, \tag{31}$$

where $k(D)$ is defined in (26) and the processes $Z_{1,D}(\cdot)$ and $Z_{2,D}(\cdot)$ are defined in (22) and (23).

The proof of Proposition 4 is a consequence of Theorem 1 in the case (i) and of Theorem 2 in the case (ii). For further details, we refer the reader to Section 4.4 of Lévy-Leduc et al. (2009).

Let us now compare the asymptotic behavior of $\hat{\sigma}_{SB}$ with that of the square root of the sample variance estimator defined by

$$\hat{\sigma}_{n,Y} = \left(\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2\right)^{1/2} \;. \tag{32}$$

Corollary 5. *Under the assumptions of Proposition 4, $\hat{\sigma}_{SB}$ defined in (29) has the following properties. In the case (i), its asymptotic relative efficiency with respect to the classical scale estimator $\hat{\sigma}_{n,Y}$ defined in (32) is larger than 86.31% and in the case (ii), there is no loss of efficiency.*

Corollary 5 is proved in Section 5.

As claimed in (Rousseeuw and Croux, 1993, p. 1277), however, one of the main drawbacks of the Shamos-Bickel estimator is its very low finite sample breakdown point (around 29%). We recall that the finite sample breakdown point of an estimator $\hat{\theta}_n$ obtained from any observations $\mathbf{x} = \{x_1, \ldots, x_n\}$ is defined, see Rousseeuw and Croux (1993), by

$$\varepsilon_n^{\star}(\hat{\theta}_n, \mathbf{x}) = \min\{\varepsilon_n^{+}(\hat{\theta}_n, \mathbf{x}), \varepsilon_n^{-}(\hat{\theta}_n, \mathbf{x})\}$$

where

$$\varepsilon_n^{+}(\hat{\theta}_n, \mathbf{x}) = \min\{m/n : \sup_{\mathbf{x}'} \hat{\theta}_n(\mathbf{x}') = \infty\},\; \varepsilon_n^{-}(\hat{\theta}_n, \mathbf{x}) = \min\{m/n : \inf_{\mathbf{x}'} \hat{\theta}_n(\mathbf{x}') = 0\}$$

and $\mathbf{x}'$ is obtained by replacing any m observations of $\mathbf{x}$ by arbitrary values. Intuitively, following (Maronna et al., 2006, p. 61), the finite sample breakdown point of $\hat{\theta}_n$ at $\mathbf{x}$ is the largest proportion of data points that can be arbitrarily replaced by outliers without $\hat{\theta}_n$ diverging to 0 or infinity. Large breakdown points are desirable. In the context of estimation of the mean, for example, the sample mean has a breakdown point of 0 and the median has a breakdown point of 50% which is the highest breakdown point that one can expect.

In order to increase the value of the finite sample breakdown point, Rousseeuw and Croux (1993) propose another robust scale estimator which is presented and studied in the next section. Their estimator has the advantage of having a breakdown point of 50% (Rousseeuw and Croux, 1993, Theorem 5) which is the highest breakdown point that one can expect.

3.3. **Rousseeuw-Croux robust scale estimator.** To estimate the scale parameter σ in the framework described in Section 3.2, Rousseeuw and Croux (1993) suggest using

$$\hat{\sigma}_{RC} = c\,\{|Y_i - Y_j|;\ 1 \le i < j \le n\}_{(k_n)} = c\,\sigma\,\{|X_i - X_j|;\ 1 \le i < j \le n\}_{(k_n)}\ , \tag{33}$$

where $k_n = \lfloor n(n-1)/4 \rfloor$, $c = 1/(\sqrt{2}\Phi^{-1}(5/8)) = 2.21914$. That is, up to the multiplicative constant c, $\hat{\sigma}_{RC}$ is the k_nth order statistic of the $n(n-1)$ distances $|X_i - X_j|$ between all the pairs of observations such that $i < j$.

Proposition 6. *Under Assumption (A1), the Rousseeuw-Croux robust scale estimator $\hat{\sigma}_{RC}$ defined in (33) from $Y_1, \dots, Y_n$ has the following asymptotic behavior.*

(i) If $D > 1/2$,

$$\sqrt{n}(\hat{\sigma}_{RC} - \sigma) \xrightarrow{d} \mathcal{N}(0, \widetilde{\sigma}^2)\ ,$$

where

$$\widetilde{\sigma}^2 = \sigma^2 \mathbb{E}[\mathrm{IF}(Y_1/\sigma)^2] + 2\sigma^2 \sum_{k \ge 1} \mathbb{E}[\mathrm{IF}(Y_1/\sigma)\mathrm{IF}(Y_{k+1}/\sigma)]\ ,$$

with

$$\mathrm{IF}(x) = c\left(\frac{1/4 - \Phi(x + 1/c) + \Phi(x - 1/c)}{\int_{\mathbb{R}} \varphi(y)\varphi(y + 1/c)\mathrm{d}y}\right)\ ,$$

Φ and φ denoting the c.d.f and the p.d.f of the standard Gaussian random variable respectively.

(ii) If $D < 1/2$,

$$k(D)n^D L(n)^{-1}(\hat{\sigma}_{RC} - \sigma) \xrightarrow{d} \frac{\sigma}{2}(Z_{2,D}(1) - Z_{1,D}^2(1))\ ,$$

where $k(D)$ is defined in (26) and the processes $Z_{1,D}(\cdot)$ and $Z_{2,D}(\cdot)$ are defined in (22) and (23).

The proof of Proposition 4 is a consequence of Theorem 1 in the case (i) and of Theorem 2 in the case (ii). For further details, we refer the reader to Theorem 6 of Lévy-Leduc et al. (2010). The following corollary is proved in Section 5.

Corollary 7. *Under the assumptions of Proposition 6, $\hat{\sigma}_{RC}$ defined in (33) has the following properties. In the case (i), its asymptotic relative efficiency with respect to the classical scale estimator $\hat{\sigma}_{n,Y}$ defined in (32) is larger than 82.27% and in the case (ii), there is no loss of efficiency.*

4. Numerical experiments

In this section, we investigate the robustness properties of the previous estimators using Monte Carlo experiments. We shall regard the observations X_t, $t = 1, \dots, n$, as a stationary series Y_t, $t = 1, \dots, n$, corrupted by additive outliers of magnitude ω. Thus we set

$$X_t = Y_t + \omega W_t, \tag{34}$$

where W_t are i.i.d. random variables. In Section 4.1, W_t are Bernoulli($p/2$) random variables. In Section 4.2, W_t are such that $\mathbb{P}(W_t = -1) = \mathbb{P}(W_t = 1) = p/2$ and $\mathbb{P}(W_t = 0) = 1 - p$, hence $\mathbb{E}[W_t] = 0$ and $\mathbb{E}[W_t^2] = \mathrm{Var}(W_t) = p$. Observe that, in this case, W is the product of Bernoulli(p) and *Rademacher* independent random variables; the latter equals 1 or -1, both with probability 1/2. $(Y_t)_t$ is a stationary time series and it is assumed that Y_t and W_t are independent random variables. The empirical study is based on 5000 independent replications with $n = 600$, $p = 10\%$ and $\omega = 10$. We consider the cases where Y_t are Gaussian ARFIMA$(0, d, 0)$ processes, that is,

$$Y_t = (I - B)^{-d} Z_t = \sum_{j \geq 0} \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)} Z_{t-j} \tag{35}$$

where B denotes the backward operator and d= 0.2, 0.45, corresponding respectively to $D = 0.6$, 0.1, where D is defined in (A1) and $\{Z_t\}$ are i.i.d $\mathcal{N}(0, 1)$.

4.1. **Hodges-Lehmann robust location estimator.** In Figure 1 the empirical density functions of $n^{D/2}L(n)^{-1/2}\hat{\theta}_{HL}$ and $n^{D/2}L(n)^{-1/2}\bar{X}_n$ are displayed when X_t has no outliers with $d = 0.2$ (left) and $d = 0.45$ (right). Following (Doukhan et al., 2003, p. 21), we take $L(n) = \Gamma(1-2d)/(\Gamma(d)\Gamma(1-d))$. In these cases both shapes are similar to the limit indicated in Proposition 3, that is, a Gaussian density with mean zero and variance 3.5714 when $d = 0.2$ and 1.1696 when $d = 0.45$. The empirical variance of the Hodges-Lehmann estimator and of the sample mean are equal to 3.6397 and 3.6249, when $d = 0.2$ and to 1.2543, 1.2538 when $d = 0.45$. These results are thus a good illustration of Proposition 3.

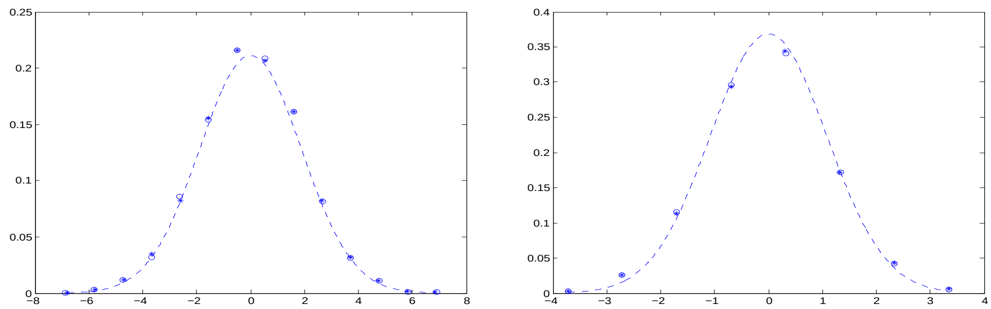

FIGURE 1. Empirical densities of the quantities $n^{D/2}L(n)^{-1/2}\hat{\theta}_{HL}$ ('*') and $n^{D/2}L(n)^{-1/2}\bar{X}_n$ ('o') for the ARFIMA$(0, d, 0)$ model with $d = 0.2$ (left), $d = 0.45$ (right), $n = 600$ without outliers and the p.d.f of a zero mean Gaussian random variable with a variance $2(-D+1)^{-1}(-D+2)^{-1}$ (dotted line).

Figure 2 displays the same quantities as in Figure 1 when X_t has outliers with $d = 0.2$ (left) and $d = 0.45$ (right). As expected, the sample mean is much more sensitive to the presence of outliers than the Hodges-Lehmann estimator. Observe that when the long-range dependence is strong (large d), the effect of outliers is less pronounced.

4.2. **Shamos-Bickel robust scale estimator.** In Figure 3, the empirical densities of $\sqrt{n}(\hat{\sigma}_{SB} - \sigma)$ and $\sqrt{n}(\hat{\sigma}_{n,X} - \sigma)$ are displayed when $d = 0.2$ without outliers (left) and with outliers (right). In the left part of this figure, we illustrate the results of part (i) of Corollary 5 since both shapes are similar to that of Gaussian density with mean zero. A lower bound for the

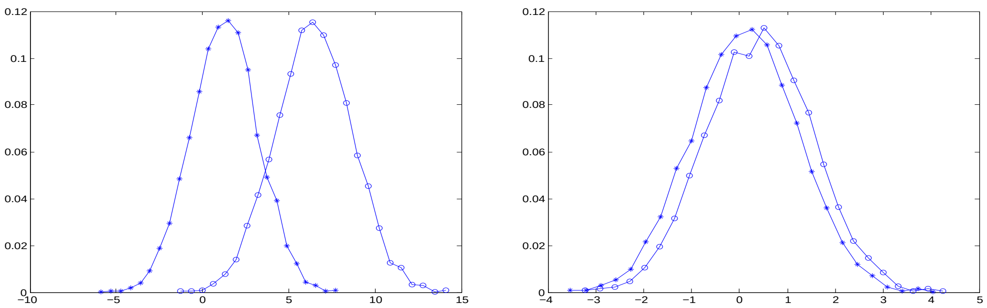

FIGURE 2. Empirical densities of the quantities $n^{D/2}L(n)^{-1/2}\hat{\theta}_{HL}$ ('*') and $n^{D/2}L(n)^{-1/2}\bar{X}_n$ ('o') for the ARFIMA$(0,d,0)$ model with $d=0.2$ (left), $d=0.45$ (right), $n=600$ with outliers ($p=10\%$ and $\omega=10$).

theoretical asymptotic relative efficiency stated in Corollary 5 is 86.31%. The empirical variances are equal to 0.7968 and 0.7135, respectively, corresponding to an asymptotic relative efficiency of 89.54%. On the right part of Figure 3, we can see that the classical scale estimator is much more sensitive to the presence of outliers than the Shamos-Bickel estimator.

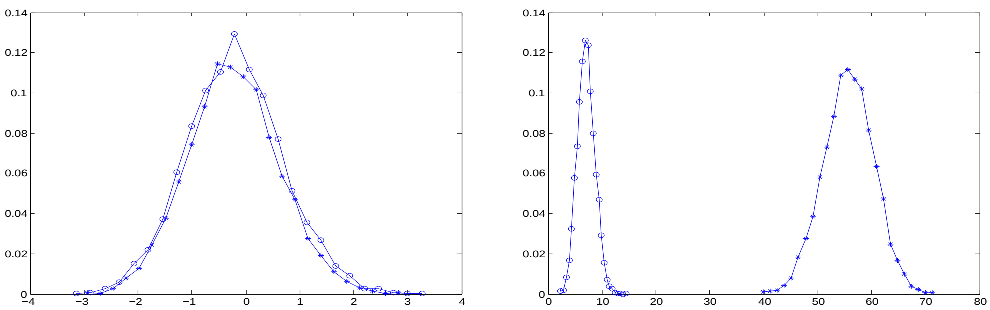

FIGURE 3. Empirical densities of the quantities $\sqrt{n}(\hat{\sigma}_{SB}-\sigma)$ ('*') and $\sqrt{n}(\hat{\sigma}_{n,X}-\sigma)$ ('o') for the ARFIMA$(0,d,0)$ model with $d=0.2$, $n=600$ without outliers (left) and with outliers $p=10\%$ and $\omega=10$ (right).

Figure 4 (left) illustrates the part (ii) of Corollary 5. $d=0.45$ corresponds to $D=0.1<1/2$. The right part of Figure 4 shows the robustness of the Shamos-Bickel estimator with respect to the classical scale estimator in the presence of outliers. In these numerical experiments, following (Doukhan et al., 2003, p. 21), we take $L(n)=\Gamma(1-2d)/(\Gamma(d)\Gamma(1-d))$.

4.3. **Rousseeuw-Croux robust scale estimator.** For numerical results associated to the Rousseeuw-Croux robust scale estimator (33), we refer the reader to Lévy-Leduc et al. (2010).

4.4. **Discussion on the construction of asymptotic confidence intervals.** We propose in this section to give some hints on how to build confidence intervals for the different estimators that we considered. There are mainly two situations: the limiting distribution is either Gaussian or a linear combination of the square of a Gaussian random variable and the Rosenblatt process evaluated at 1. For instance in the case of the Shamos-Bickel estimator

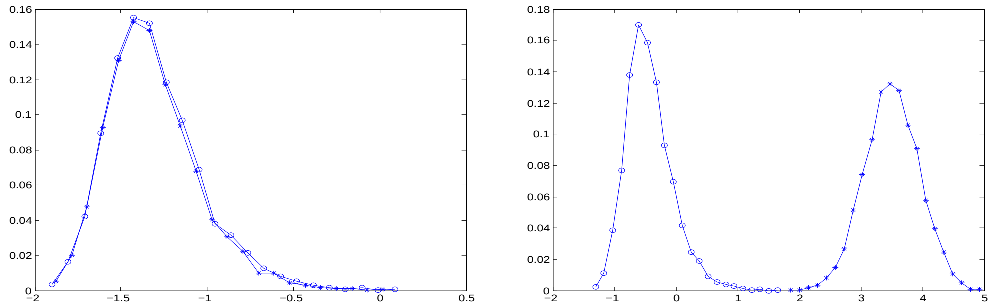

FIGURE 4. Empirical densities of the quantities $n^D L(n)^{-1}(\hat{\sigma}_{SB} - \sigma)$ ('*') and $n^D L(n)^{-1}(\hat{\sigma}_{n,X} - \sigma)$ ('o') for the ARFIMA$(0, d, 0)$ model with $d = 0.45$, $n = 600$ without outliers (left) and with outliers $p = 10\%$ and $\omega = 10$ (right).

when D is in $(1/2, 1)$, we give in Table 1 an empirical evaluation of $\bar{\sigma}^2/\sigma^2$ for different values of $D = 1 - 2d$ of an ARFIMA(0,d,0) process, where $\bar{\sigma}^2$ is defined in (30). These values have been obtained by using quasi-Monte Carlo approaches. An asymptotic confidence interval for σ can then be obtained by plug-in.

d	0.1	0.12	0.15	0.17	0.2
$\bar{\sigma}^2/\sigma^2$	0.6050	0.6232	0.6670	0.7121	0.8202

TABLE 1. Empirical values of $\bar{\sigma}^2/\sigma^2$ for an ARFIMA$(0, d, 0)$ process with different values of d.

The case where $D = 1 - 2d$ belongs to $(0, 1/2)$ is more involved. It requires an estimator of D as well as an estimation of $L(n)$. In Table 2, we give for different values of d the 95% empirical quantiles associated to $(Z_{2,D}(1) - Z_{1,D}(1)^2)/2$, namely the x's satisfying $\mathbb{P}(Z_{2,D}(1) - Z_{1,D}(1)^2) \leq 2x) = 0.95$. These values have been obtained by using that from Lemma 14 in Lévy-Leduc et al. (2009)

(36)
$$k(D)n^{D-2}L(n)^{-1}\left[n/2\sum_{i=1}^{n}(X_i^2 - 1) - 1/2\sum_{1\leq i,j\leq n} X_iX_j\right] \stackrel{d}{\to} 1/2\left[Z_{2,D}(1) - (Z_{1,D}(1))^2\right] ,$$

as n tends to infinity. More precisely, we simulated for $n = 1000$, 5000 replications of the l.h.s of (36) when X is an ARFIMA(0,d,0) process.

d	0.26	0.3	0.35	0.4	0.45
x	45.2103	29.5672	19.2967	15.0844	15.5393

TABLE 2. Values of x defined by $\mathbb{P}(Z_{2,D}(1) - Z_{1,D}(1)^2) \leq 2x) = 0.95$ for an ARFIMA$(0, d, 0)$ process with different values of d.

5. Proofs

Proof of Corollary 5. Consider $\hat{\sigma}_{n,Y}$ defined in (32).

Case (i). Let us first prove that, as n tends to infinity,

$$\sqrt{n}(\hat{\sigma}^2_{n,Y} - \sigma^2) \xrightarrow{d} \mathcal{N}(0, 2\sigma^4[1 + 2\sum_{k\geq 1}\rho(k)^2]) \; . \tag{37}$$

Using (32) and $X_i = Y_i/\sigma$, $n(n-1)(\hat{\sigma}^2_{n,Y} - \sigma^2) = \sigma^2[n\sum_{i=1}^n (X_i^2 - 1) - \sum_{1\leq i,j\leq n} X_i X_j + n]$. Since $H_2(X_i) = X_i^2 - 1$, we get

$$\sqrt{n}(\hat{\sigma}^2_{n,Y} - \sigma^2) = \sigma^2 \left[\frac{\sqrt{n}}{n-1}\sum_{i=1}^n H_2(X_i) - \frac{\sqrt{n}}{n(n-1)}\sum_{1\leq i,j\leq n} X_i X_j + \frac{\sqrt{n}}{n-1} \right] \; . \tag{38}$$

By Theorem 4 in Arcones (1994),

$$n^{-1/2}\sum_{i=1}^n H_2(X_i) \xrightarrow{d} \mathcal{N}(0, 2[1 + 2\sum_{k\geq 1}\rho(k)^2]) \; .$$

Using the same arguments as those used in the proof of (80) in Lemma 15 of Lévy-Leduc et al. (2009), $\sqrt{n}(n-1)^{-1}\sum_{i=1}^n H_2(X_i)$ is the leading term in (38), which gives (37). Using the Delta method to go from $\hat{\sigma}^2_{n,Y}$ to $\hat{\sigma}_{n,Y}$, setting $f(x) = \sqrt{x}$ so that $f'(\sigma^2) = 1/(2\sqrt{\sigma^2}) = 1/(2\sigma)$, we get

$$\sqrt{n}(\hat{\sigma}_{n,Y} - \sigma) \xrightarrow{d} \mathcal{N}(0, \frac{\sigma^2}{2}[1 + 2\sum_{k\geq 1}\rho(k)^2]) \; .$$

On the other hand, by Proposition 4 (case (i)),

$$\sqrt{n}(\hat{\sigma}_{SB} - \sigma) \xrightarrow{d} \mathcal{N}(0, \bar{\sigma}^2) \; ,$$

where

$$\bar{\sigma}^2 = \frac{2b^2\sigma^2}{\varphi^2(1/(b\sqrt{2}))}\left[\mathrm{Var}(h_1(Y_1/\sigma, 1/b)) + 2\sum_{k\geq 1}\mathrm{Cov}(h_1(Y_1/\sigma, 1/b), h_1(Y_{k+1}/\sigma, 1/b)) \right] \; .$$

Thus, the asymptotic relative efficiency (ARE) of $\hat{\sigma}_{SB}$ with respect to $\hat{\sigma}_{n,Y}$ satisfies

$$\mathrm{ARE} = \frac{\sigma^2/2[1 + 2\sum_{k\geq 1}\rho(k)^2]}{\bar{\sigma}^2} \; .$$

Using Lemma 1 in Arcones (1994) and the fact that h_1 is of Hermite rank greater than 2, we get that for any $k \geq 1$, $\mathrm{Cov}(h_1(Y_1/\sigma, 1/b), h_1(Y_{k+1}/\sigma, 1/b)) \leq \rho(k)^2\,\mathrm{Var}(h_1(Y_1/\sigma, 1/b))$, which gives

$$\bar{\sigma}^2 \leq \frac{2b^2\sigma^2}{\varphi^2(1/(b\sqrt{2}))}\mathrm{Var}(h_1(Y_1/\sigma, 1/b))[1 + 2\sum_{k\geq 1}\rho(k)^2] \; ,$$

and hence

$$\mathrm{ARE} \geq \frac{\varphi^2(1/(b\sqrt{2}))}{4b^2\,\mathrm{Var}(h_1(X_1, 1/b))} \approx 86.31\% \; .$$

The value of 86.31% has been obtained by approximating $\mathrm{Var}(h_1(X_1, 1/b))$ with some Monte-Carlo simulations.

Case (ii). When $0 < D < 1/2$, (38) and Lemma 14 in Lévy-Leduc et al. (2009) lead to

$$k(D)n^D L(n)^{-1}(\hat{\sigma}^2_{n,Y} - \sigma^2) \xrightarrow{d} \sigma^2(Z_{2,D}(1) - Z^2_{1,D}(1)) \ . \tag{39}$$

The result follows by applying the Delta method.

□

Proof of Corollary 7. The proof of (i) is given in the proof of Proposition 3 in Lévy-Leduc et al. (2010). The proof of (ii) comes from (39) and the statement (ii) in Proposition 6. □

Acknowledgements

Murad S. Taqqu would like to thank Telecom ParisTech for their hospitality. This research was partially supported by the NSF Grant DMS-0706786 at Boston University.

Valderio A. Reisen would like to thank Telecom ParisTech for their hospitality and CNPq-Brazil for financial support.

Dedication

We would like to dedicate this paper to Benoit Mandelbrot who led the way.

References

Arcones, M. (1994). Limit theorems for nonlinear functionals of a stationary Gaussian sequence of vectors. *Annals of Probability 22*(4), 2242–2274.

Bickel, P. J. and E. L. Lehmann (1979). Descriptive statistics for nonparametric models iv: Spread. In *Contributions to Statistics, Hájek Memorial Volume* (J. Jurečková ed.)., pp. 33–40. Prague: Academia.

Doukhan, P., G. Oppenheim, and M. S. Taqqu (Eds.) (2003). *Theory and applications of long-range dependence*. Boston, MA: Birkhäuser Boston Inc.

Fox, R. and M. S. Taqqu (1987). Multiple stochastic integrals with dependent integrators. *Journal of multivariate analysis 21*, 105–127.

Hodges, J. L. J. and E. L. Lehmann (1963). Estimates of location based on rank tests. *Annals of Mathematical Statistics 34*, 598–611.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics 19*, 293–325.

Lévy-Leduc, C., H. Boistard, E. Moulines, M. S. Taqqu, and V. A. Reisen (2009). Asymptotic properties of U-processes under long-range dependence. Technical report. arXiv:0912.4688.

Lévy-Leduc, C., H. Boistard, E. Moulines, M. S. Taqqu, and V. A. Reisen (2010). Robust estimation of the scale and of the autocovariance function of Gaussian short and long-range dependent processes. *To appear in Journal of Time Series Analysis*.

Maronna, R. A., R. D. Martin, and V. J. Yohai (2006). *Robust statistics*. Wiley Series in Probability and Statistics. Chichester: John Wiley & Sons Ltd. Theory and methods.

Rousseeuw, P. and C. Croux (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical Association 88*(424), 1273–1283.

Shamos, M. I. (1976). Geometry and statistics: Problems at the interface. In *New directions and Recent Results in Algorithms and Complexity* (J. F. Traub ed.)., pp. 251–280. New York: Academic Press.

CNRS/LTCI/TelecomParisTech - 46, rue Barrault, 75634 Paris Cédex 13, France.
E-mail address: celine.levy-leduc@telecom-paristech.fr

Toulouse School of Economics (GREMAQ, Université Toulouse 1) - Manufacture des Tabacs - Bât. F, aile J.J. Laffont - 21 allée de Brienne - 31000 Toulouse
E-mail address: helene@boistard.fr

CNRS/LTCI/TelecomParisTech - 46, rue Barrault, 75634 Paris Cédex 13, France.
E-mail address: eric.moulines@telecom-paristech.fr

Department of Mathematics and Statistics, Boston University, 111 Cumington Street, Boston, MA 02215, USA
E-mail address: murad@math.bu.edu

Departamento de Estatística, Universidade Federal do Espírito Santo, Vitória/ES, Brazil
E-mail address: valderioanselmoreisen@gmail.com