Corrigendum to "Subsampling inference for the mean of heavy-tailed long-memory time series" by A. Jach, T. S. McElroy, and D.N. Politis

Agnieszka Jach[1] Tucker S. McElroy[2,3] Dimitris N. Politis[4]

1 Introduction

In Jach, McElroy, and Politis (2012), it was claimed in Theorem 4 of Appendix B that subsampling (Politis, Romano, and Wolf, 1999)) is valid under the θ-weak dependence condition. Both the statement and proof of Theorem 4 are incorrect. We provide a corrected statement and proof in what follows. The authors are indebted to S. Bai, M. Taqqu, and T. Zhang for alerting us to the errors in a personal communication (Bai et al., 2015a).

More specifically, Theorem 4 is not true without additional assumptions, such as Lipschitz continuity of the statistic of interest. This corrigendum has three objectives: (i) we discuss the gap in the general result; (ii) we add additional assumptions such that the general result holds; (iii) we make application to the subordinated Gaussian process of Jach et al. (2012), with the sample mean's particular studentization. Note that recent work of Bai, Taqqu, and Zhang (2015b) also establishes this last result, albeit using a different strategy: instead of utilizing θ-weak dependence, they proceed directly to a maximal correlation inequality that holds for subordinated Gaussian processes.

Consider a statistic $\widehat{\theta}_n$, which is an estimator of an unknown parameter θ, with a studentization $\widehat{\sigma}_n$ such that $\tau_n(\widehat{\theta} - \theta)/\widehat{\sigma}_n$ converges weakly to some limit random variable with cumulative distribution function (cdf) denoted by L. Let the sampling cdf be denoted by L_n, which is computed from a sample of size n drawn from a stationary θ-weakly dependent process (Doukhan and Louhichi, 1999)) with coefficients ε_r.

The θ-weak dependence condition is convenient for long-range dependent processes, because some of these processes (e.g., Gaussian long-range dependent processes) may not be strong mixing; see the discussion in Doukhan, Prohl, and Robert (2011). In order to establish the validity of subsampling, it is necessary to have some sort of covariance inequality,

[1]Department of Finance and Statistics, Hanken School of Economics, Arkadiankatu 22, 00100 Helsinki, Finland, agnieszka.jach@hanken.fi

[2]Center for Statistical Research and Methodology, U.S. Census Bureau, 4600 Silver Hill Road, Washington, D.C. 20233-9100, USA, tucker.s.mcelroy@census.gov

[3]This report is released to inform interested parties of research and to encourage discussion. The views expressed on statistical issues are those of the authors and not necessarily those of the U.S. Census Bureau.

[4]Department of Mathematics, University of California, San Diego, 9500 Gilman Drive, Mail Code 0112, La Jolla, CA 92093-0112, USA, dpolitis@ucsd.edu

which relates the covariance of certain functions of the data (e.g., statistics and pivots) to mixing coefficients. The strong mixing inequality is very generic, as it does not depend upon the structure of these functions; however, the inequality for θ-weak dependence requires a Lipschitz continuity condition, which somewhat limits the types of statistics that can be studied. This is the approach used in Ango-Nze, Dupoiron, and Rios (2003).

The proof of Theorem 4 in Jach et al. (2012) is flawed, because the studentized pivot $\tau_n(\widehat{\theta} - \theta)/\widehat{\sigma}_n$ is falsely presumed to be Lipschitz in the data. Moreover, that result was applied to the the studentized pivot of Jach et al. (2012), which is not Lipschitz in the data. In particular, the application involves θ being given as the mean of a subordinated Gaussian process $\{X_t\}$, and $\widehat{\theta}_n$ given by the sample mean $\overline{X}_n$. The studentization is given by

$$\begin{aligned}
\widehat{\sigma}_n &= \sqrt{M_n(2) + \widehat{\mathrm{LM}}_n(\rho)} \qquad (1)\\
M_n(2) &= n^{-1} \sum_{t=1}^{n} \left(X_t - \overline{X}_n\right)^2 \\
\widehat{\mathrm{LM}}_n(\rho) &= \left| \sum_{|h|=1}^{\lfloor n^\rho \rfloor} (n - |h|)^{-1} \sum_{t=1}^{n-|h|} (X_t X_{t+h} - \overline{X}_n^2) \right|^{1/\rho},
\end{aligned}$$

with $\tau_n = \sqrt{n}$. The studentized pivot is a function of θ and $X_1, X_2, \cdots, X_n$; when each $X_i = \theta + \kappa$ for any $\kappa \neq 0$, then the pivot equals $\sqrt{n}\kappa$ divided by zero. Thus, it is not Lipschitz continuous and the techniques of Ango-Nze et al. (2003) are not directly applicable. However, both $\overline{X}_n$ and $M_n(2)$ are Lipschitz continuous in the sample (shown below), and hence it is possible to establish subsampling consistency for the inference problem of Jach et al. (2012), once the impact of the statistic $\widehat{\mathrm{LM}}_n(\rho)$ is accounted for.

Below, we add additional assumptions to Theorem 4 of Jach et al. (2012) to extend the results of Ango-Nze et al. (2003) to studentized pivots. As we later demonstrate, nonlinear statistics can be entertained, but certain conditions relating the Lipschitz constants of estimators and studentizers to the rate of θ-mixing are needed. As a second result, we show that the Lipschitz conditions are satisfied for the inference problem of Jach et al. (2012).

2 Results

We first describe the subsampling distribution estimator. Let $\widehat{\theta}_{b,t}$ and $\widehat{\sigma}_{b,t} \geq 0$ denote statistics computed on the samples $X_t, \cdots, X_{t+b-1}$, where b is a sequence depending on n chosen such that $b/n + 1/b \to 0$ as $n \to \infty$. (In the case that $\widehat{\sigma}_n$ is given by (1), the subsampled studentization is computed via calculating $M_{b,t}(2)$, and $\widehat{\mathrm{LM}}_{b,t}(\rho)$.) Let $q = n - b + 1$ be the number of such available subsamples. Then the subsampling distribution estimator is

$$\widehat{L}_{n,b}(x) = q^{-1} \sum_{t=1}^{q} 1_{\{\tau_b(\widehat{\theta}_{b,t} - \widehat{\theta}_n)/\widehat{\sigma}_{b,t} \leq x\}}.$$

Let the lower $1-p$ quantile of this empirical distribution be denoted $\ell_{n,b}(1-p)$, and denote the lower $1-p$ quantile of $L(x)$ by $\ell(1-p)$.

A key assumption is that the sampling distribution converges. But we also require that the studentization converges jointly with the pivot's numerator.

Assumption A: As $n \to \infty$ assume that $L_n(x) \to L(x)$. In addition, suppose that $[\alpha_n(\widehat{\theta}_n - \theta), \delta_n \widehat{\sigma}_n] \stackrel{\mathcal{L}}{\Longrightarrow} [Z, W]$ jointly, for some positive sequences $\{\alpha_n\}$ and $\{\delta_n\}$ such that $\tau_n = \alpha_n/\delta_n$, and for Z and W nondegenerate random variables such that W is positive with probability one. Finally, if x is a continuity point of L, then it is assumed that the cdf of $Z - xW$ is continuous.

Regarding the last point of Assumption A, we note that $1 - L(x) = \mathbb{P}[Z - xW > 0]$, so that the cdf of Z/W is related to that of $Z - xW$, for any x. If we know that the joint characteristic function of Z and W is continuous in a neighborhood of $(0, 0)$, then the joint probability density function (pdf) $p_{Z,W}$ exists. Also, if the joint Fourier-Laplace transform (Fitzsimmons and McElroy, 2010) is continuous in some neighborhood of the origin, then the joint pdf exists. In such a case, for any x and y

$$\mathbb{P}[Z - xW \leq y] = \int_0^\infty \mathbb{P}[Z \leq y + xw | W = w] \, p_W(w) \, dw$$

(because W is nondegenerate, and the conditional pdf exists when the joint pdf does), which is differentiable with respect to y; in particular, the pdf of $Z - xW$ is

$$p_{Z-xW}(y) = \int_0^\infty p_{Z|W=w}(y + xw) \, p_W(w) \, dw = \int_0^\infty p_{Z,W}(y + xw, w) \, dw.$$

In summary, if the joint characteristic function (or joint Fourier-Laplace transform) of Z and W is continuous in a neighborhood of the origin, then Z, W, and $Z - xW$ are nondegenerate.

In order to state our result, we need to introduce the concept of Lipschitz continuity of a statistic. A statistic is some function $g(X_1, \cdots, X_n)$ of the data. Let $\| \cdot \|_1$ denote the $\mathbb{L}_1$ norm on $\mathbb{R}^n$. Then let

$$\mathrm{Lip}(g) = \sup_{x,y \in \mathbb{R}^n} \frac{g(x) - g(y)}{\|x - y\|_1},$$

and if this is finite we say that g is Lipschitz continuous (with respect to the data). The constant $\mathrm{Lip}(g)$ can depend upon n. Let $\mathcal{L}$ denote the set of Lipschitz continuous functions. Definition 2.1 of Ango-Nze et al. (2003) defines $(\theta, \mathcal{L}, \Psi_1)$-weak dependence via the covariance inequaliity

$$|\mathrm{Cov}\,(f(X_{i_1}, \cdots, X_{i_u}), g(X_{j_1}, \cdots, X_{j_v}))| \leq \Psi_1(f, g, u, v) \, \varepsilon_r, \tag{2}$$

for any $f, g \in \mathcal{L}$, $\{\varepsilon_r\}$ the mixing sequence, and $i_1 < \cdots < i_u \leq r + i_u \leq j_1 < \cdots < j_v$. Finally, Ψ_1 takes the form

$$\Psi_1(f, g, u, v) = u \, \mathrm{Lip}(f) + v \, \mathrm{Lip}(g).$$

With these concepts in place, we can state our general theorem.

Theorem 1 *Suppose that $\{X_t\}$ is strictly stationary and $(\theta, \mathcal{L}, \Psi_1)$-weakly dependent with rate $\varepsilon_r = O(r^{-a})$ for some $a > 0$. Assume Assumption A, with rates such that $\alpha_b/\alpha_n \to 0$, $\tau_b/\tau_n \to 0$, and $b/n + 1/b \to 0$ as $n \to \infty$. Suppose that $\widehat{\theta}_n$ and $\widehat{\sigma}_n$ are Lipschitz continuous,* *and set* $\kappa_n(x) = \tau_n \, Lip(\widehat{\theta}_n) + |x| \, Lip(\widehat{\sigma}_n)$ *for any* $x \in \mathbb{R}$. *Given some* $\zeta < a \wedge 1$, *if* $b\,\delta_b\,\kappa_b(x) = O(b^{\zeta})$ *for any* $x \in \mathbb{R}$, *then the following hold:*

(i) If x *is a continuity point of* L, *then* $\widehat{L}_{n,b}(x) \xrightarrow{P} L(x)$.

(ii) If L *is continuous, then* $\sup_x |\widehat{L}_{n,b}(x) - L(x)| \xrightarrow{P} 0$.

(iii) If L *is continuous at* $\ell(1-p)$, *then as* $n \to \infty$

$$\mathbb{P}[\tau_n(\widehat{\theta}_n - \theta)/\widehat{\sigma}_n \leq \ell_{n,b}(1-p)] \to 1-p.$$

Remark 1 A subsampling theorem for Lipschitz statistics (or pivots) is given in Doukhan, Prohl, and Robert (2011), and they consider both overlapping and non-overlapping blocks. Our theorem considers the Lipschitz conditions for test statistic and studentization separately, and thus allows for applications where the studentized statistic is not Lipschitz.

Remark 2 In the case of slowly-decaying weak dependence (e.g., long-range dependence) – say with a less than unity and close to zero – the Lipschitz continuity of both the studentization (through δ_b) and the estimator (through $\kappa_b(x)$) must be greater, corresponding to smaller Lipschitz constants, because ζ is required to be smaller. As seen in the proof, the size of ζ has a direct impact on the variability of the subsampling distribution estimator.

Proof of Theorem 1. The proof of (i) splices Theorem 11.3.1 of Politis, Romano, and Wolf (1999) with an extension of Lemma 3.1 of Ango-Nze et al. (2003); then (ii) and (iii) will follow as in the proof of Theorem 3.2.1 of Politis, Romano, and Wolf (1999), utilizing the assumed continuity of L at x. For any $\epsilon > 0$, and for all $x, t \in \mathbb{R}$, let $h_x(t) = 1_{\{t \leq x\}}$ and

$$h_{x,\epsilon}(t) = h_x(t) + \left(\frac{x + \epsilon - t}{\epsilon} \right) 1_{\{x < t \leq x + \epsilon\}}. \tag{3}$$

Hence $\widehat{L}_{n,b}(x) = q^{-1} \sum_{t=1}^{q} h_x \left(\tau_b(\widehat{\theta}_{b,t} - \widehat{\theta}_n)/\widehat{\sigma}_{b,t} \right)$. It is straightforward to show that the functions $h_{x,\epsilon}(t)$ are Lipschitz with respect to both x and t, with Lipschitz constant ϵ^{-1}. The method of proof proceeds by approximating $L_{n,b}(x)$ via replacing each h_x function by $h_{x,\epsilon}$, so that we can use the θ-weak dependence covariance inequality (which requires Lipschitz functions of the data). First, we follow the arguments in the proof of Theorem 11.3.1 of Politis, Romano, and Wolf (1999), to replace $\widehat{\theta}_n$ by θ in $\widehat{L}_{n,b}(x)$, with the resulting substitution denoted

$$\widetilde{L}_n(x) = q^{-1} \sum_{t=1}^{q} h_x \left(\tau_b(\widehat{\theta}_{b,t} - \theta)/\widehat{\sigma}_{b,t} \right).$$

This argument utilizes Assumption A, and does not rely on the mixing asssumptions or Lipschitz continuity. Next, observe that

$$\left\{ \frac{\tau_b(\widehat{\theta}_{b,t} - \theta)}{\widehat{\sigma}_{b,t}} \leq x \right\} = \left\{ \alpha_b(\widehat{\theta}_{b,t} - \theta) \leq x \, \delta_b \, \widehat{\sigma}_{b,t} \right\},$$

so that

$$h_x \left(\tau_b(\widehat{\theta}_{b,t} - \theta)/\widehat{\sigma}_{b,t} \right) = h_{x \, \delta_b \, \widehat{\sigma}_{b,t}} \left(\alpha_b(\widehat{\theta}_{b,t} - \theta) \right).$$

So $\widetilde{L}_n(x)$ is an average of such over $t = 1, \cdots, q$, and the expectation equals

$$\mathbb{P}\left[\frac{\tau_b(\widehat{\theta}_{b,1} - \theta)}{\widehat{\sigma}_{b,1}} \leq x\right] \to \mathbb{P}[Z/W \leq x] = L(x)$$

as $b \to \infty$ (i.e., as $n \to \infty$). So it remains to show that the variance of $\widetilde{L}_n(x)$ tends to zero. Letting $\gamma(k)$ denote the autocovariance function of the stationary sequence $h_x(\tau_b(\widehat{\theta}_{b,t} - \theta)/\widehat{\sigma}_{b,t})$, it suffices (because $b/q \to 0$ and $\gamma(k)$ is bounded) to show that $q^{-1}\sum_{k=b}^{q}\gamma(k) \to 0$ as $n \to \infty$. So for $k \geq b$

$$\begin{aligned}
\gamma(k) = \mathrm{Cov}&\left(h_x(\tau_b(\widehat{\theta}_{b,t} - \theta)/\widehat{\sigma}_{b,t}) - h_{x\,\delta_b\,\widehat{\sigma}_{b,t},\epsilon}(\alpha_b(\widehat{\theta}_{b,t} - \theta)), h_x(\tau_b(\widehat{\theta}_{b,t+k} - \theta)/\widehat{\sigma}_{b,t+k})\right) \\
&+ \mathrm{Cov}\left(h_{x\,\delta_b\,\widehat{\sigma}_{b,t},\epsilon}(\alpha_b(\widehat{\theta}_{b,t} - \theta)), h_x(\tau_b(\widehat{\theta}_{b,t+k} - \theta)/\widehat{\sigma}_{b,t+k}) - h_{x\,\delta_b\,\widehat{\sigma}_{b,t+k},\epsilon}(\alpha_b(\widehat{\theta}_{b,t+k} - \theta))\right) \\
&+ \mathrm{Cov}\left(h_{x\,\delta_b\,\widehat{\sigma}_{b,t},\epsilon}(\alpha_b(\widehat{\theta}_{b,t} - \theta)), h_{x\,\delta_b\,\widehat{\sigma}_{b,t+k},\epsilon}(\alpha_b(\widehat{\theta}_{b,t+k} - \theta))\right).
\end{aligned}$$

The first two terms are bounded by a constant times $\mathbb{P}[0 < Z - xW \leq \epsilon]$ (for n sufficiently large) using the Cauchy-Schwarz inequality and the following calculation: by (3), for any t

$$\begin{aligned}
&\mathrm{Var}\left[h_{x\,\delta_b\,\widehat{\sigma}_{b,t},\epsilon}\left(\alpha_b(\widehat{\theta}_{b,t} - \theta)\right) - h_{x\,\delta_b\,\widehat{\sigma}_{b,t}}\left(\alpha_b(\widehat{\theta}_{b,t} - \theta)\right)\right] \\
&\leq \mathbb{P}\left[x\,\delta_b\,\widehat{\sigma}_{b,t} < \alpha_b(\widehat{\theta}_{b,t} - \theta) \leq x\,\delta_b\,\widehat{\sigma}_{b,t} + \epsilon\right] \\
&= \mathbb{P}\left[0 < \alpha_b(\widehat{\theta}_{b,t} - \theta) - x\,\delta_b\,\widehat{\sigma}_{b,t} \leq \epsilon\right] \\
&\to \mathbb{P}[0 < Z - xW \leq \epsilon].
\end{aligned}$$

Because of Assumption A, this probability can be made as small as desired by shrinking ϵ. For the third term in the decomposition of $\gamma(k)$ above, we wish to use the covariance inequality of Definition 2.1 in Ango-Nze et al. (2003), recognizing in (2) we have $u, v = b$ and $r = k$, and with f and g given accordingly. To obtain $\mathrm{Lip}(f)$, where $f(X_t, \cdots, X_{b+t-1}) = h_{x\,\delta_b\,\widehat{\sigma}_{b,t},\epsilon}(\alpha_b(\widehat{\theta}_{b,t} - \theta))$, we compute

$$\begin{aligned}
|f(y_1, \cdots, y_b) - f(z_1, \cdots, z_b)| &\leq \left|h_{x\,\delta_b\,\widehat{\sigma}(y),\epsilon}(\alpha_b(\widehat{\theta}(y) - \theta)) - h_{x\,\delta_b\,\widehat{\sigma}(y),\epsilon}(\alpha_b(\widehat{\theta}(z) - \theta))\right| \\
&\quad + \left|h_{x\,\delta_b\,\widehat{\sigma}(y),\epsilon}(\alpha_b(\widehat{\theta}(z) - \theta)) - h_{x\,\delta_b\,\widehat{\sigma}(z),\epsilon}(\alpha_b(\widehat{\theta}(z) - \theta))\right| \\
&\leq \epsilon^{-1}\left(\alpha_b\,|\widehat{\theta}(y) - \widehat{\theta}(z)| + |x|\,\delta_b\,|\widehat{\sigma}(y) - \widehat{\sigma}(z)|\right),
\end{aligned}$$

where the notation for the statistics denotes that they are either evaluated at $y = [y_1, \cdots, y_b]'$ or $z = [z_1, \cdots, z_b]'$. It follows that $\mathrm{Lip}(f) \leq \epsilon^{-1}\,\delta_b\,\kappa_b(x)$. Clearly the same expression holds for $\mathrm{Lip}(g)$, and hence in this case $\Psi_1(f, g, u, v) = 2\epsilon^{-1}\,b\,\delta_b\,\kappa_b(x)$, so that the overall bound on the covariance is $2\epsilon^{-1}\,b\,\delta_b\,\kappa_b(x)\,\varepsilon_{k+1-b}$. Summing k from b to q and dividing by q yields a bound $C\,\epsilon^{-1}\,b^{\zeta}\,d_n$, where $C > 0$ and d_n is given by n^{-a}, $\log n/n$, or n^{-1} depending on whether $a < 1$, $a = 1$, or $a > 1$. It follows that $d_n = o(n^{-\zeta})$ for any value of a, and hence the covariance upper bound is $o(\epsilon^{-1}(b/n)^{\zeta})$; setting ϵ to be of order $(b/n)^{\zeta/2}$ implies that a bound of $o((b/n)^{\zeta/2})$ is obtained. Putting everything together, the variance of $\widetilde{L}_n(x)$ tends to zero as $n \to \infty$. (The final arguments draw on comments of Zhang (2016).) $\square$

We wish to apply Theorem 1 directly to the statistics and processes studied in Jach et al. (2012), but the self-normalization is not Lipschitz, due to the inclusion of $\widehat{LM}_n(\rho)$ – although the statistic $M_n(2)$ is indeed Lipschitz continuous. Using the convergence in probability of $\widehat{LM}_n(\rho)$ to a positive constant, we are able to extend the application of Theorem 1 to the studentization (1).

Let $\{X_t\}$ be the Heavy-Tailed Long Memory (HTLM) process given by

$$X_t = \sigma_t \, G_t + \eta,$$

where η is the mean, $\{\sigma_t\}$ and $\{G_t\}$ are independent, the $\{\sigma_t\}$ are i.i.d. positive random variables with cdf in the domain of attraction of an α-stable distribution with $\alpha \in (1, 2)$, and $\{G_t\}$ a long memory process of parameter $\beta \in (0, 1)$. Furthermore, $G_t = g(V_t)$ for some causal long memory Gaussian process $\{V_t\}$, where g is Lipschitz and has unit Hermite rank. Proposition 1 of Jach et al. (2012) establishes that $\{G_t\}$ and $\{X_t\}$ are θ-weakly dependent, while Theorem 3 of that work establishes the joint convergences required by Assumption A under some assumptions on the rates of convergence of the sample mean and the studentization.

Theorem 2 *Suppose that $\{X_t\}$ is the HTLM process described above, and that the conditions of Theorem 3 in Jach et al. (2012) hold. Also suppose that the mixing rate is $\varepsilon_r = O(r^{-a})$ for some $a > 0$ such that $1 - a < \alpha^{-1} \vee (\beta + 1)/2$ if $a < 1$. Then the conclusions of Theorem 1 hold.*

Remark 3 In the case of small a (a slower rate of decay for the weak-dependence coefficients), the theorem requires that either α be smaller or β be concomittantly greater. This means that processes with strong long memory are not precluded (especially if they are quite heavy-tailed), although the condition that $1 - a < \alpha^{-1} \vee (\beta + 1)/2$ is typically not verifiable – as a, α, and β will be unknown in practice.

Proof of Theorem 2. Before applying Theorem 1, we show that the subsampling distribution estimator $\widehat{L}_{n,b}(x)$ corresponding to the studentization (1) has asymptotically negligible discrepancy with subsampling based upon the alternative normalization

$$\widetilde{\sigma}_n = \sqrt{M_n(2) + \delta_n^{-2} \, D_\rho},$$

where $\widehat{LM}_n(\rho) \xrightarrow{P} D_\rho$, which is positive by assumption. Let $\widehat{Z}_{n,b,t} = \alpha_b(\widehat{\theta}_{b,t} - \widehat{\theta}_n)/[\delta_b \widehat{\sigma}_{b,t}]$, and $\widetilde{Z}_{n,b,t} = \alpha_b(\widehat{\theta}_{b,t} - \widehat{\theta}_n)/[\delta_b \widetilde{\sigma}_{b,t}]$. Then for any $\xi > 0$ and any $t = 1, \cdots, q$,

$$\begin{aligned} 1_{\{\widehat{Z}_{n,b,t} \leq x\}} &= 1_{\{\widetilde{Z}_{n,bt} \leq x \widehat{\sigma}_{n,b,t}/\widetilde{\sigma}_{n,b,t}\}} \cdot \left[1_{\left\{\frac{\widehat{\sigma}_{n,b,t}}{\widetilde{\sigma}_{n,b,t}} \leq 1+\xi\right\}} + 1_{\left\{\frac{\widehat{\sigma}_{n,b,t}}{\widetilde{\sigma}_{n,b,t}} > 1+\xi\right\}} \right] \\ &\leq 1_{\{\widetilde{Z}_{n,bt} \leq x(1+\xi)\}} + 1_{\left\{\frac{\widehat{\sigma}_{n,b,t}}{\widetilde{\sigma}_{n,b,t}} > 1+\xi\right\}}. \end{aligned}$$

Let the subsampling distribution estimator based upon the $\widetilde{Z}_{n,b,t}$ be denoted $\widetilde{L}_{n,b}(x)$. Setting $R_n(\xi)$ to be the average over t of the indicators $1_{\{\widehat{\sigma}_{n,b,t}/\widetilde{\sigma}_{n,b,t} \leq 1+\xi\}}$, we obtain

$$\widehat{L}_{n,b}(x) \leq \widetilde{L}_{n,b}(x[1+\xi]) + 1 - R_n(\xi). \tag{4}$$

Similarly,

$$1_{\{\widetilde{Z}_{n,b,t} \leq x/(1+\xi)\}} \leq 1_{\{\widehat{Z}_{n,bt} \leq x\}} + 1_{\left\{\frac{\widetilde{\sigma}_{n,b,t}}{\widehat{\sigma}_{n,b,t}} > 1+\xi\right\}},$$

which implies that

$$\widehat{L}_{n,b}(x) \geq \widetilde{L}_{n,b}(x/[1+\xi]) - R_n(-\xi/[1+\xi]). \tag{5}$$

Using the Markov inequality and the fact that $\widehat{\sigma}_{n,b,t}/\widetilde{\sigma}_{n,b,t} \xrightarrow{P} 1$, we see that $R_n(\xi) \xrightarrow{P} 1$ and $R_n(-\xi/[1+\xi]) \xrightarrow{P} 0$, for $\xi > 0$. Together with (4) and (5), for any $\eta > 0$ with probability tending to one

$$\widetilde{L}_{n,b}(x/[1+\xi]) - \eta \leq \widehat{L}_{n,b}(x) \leq \widetilde{L}_{n,b}(x[1+\xi]) + \eta.$$

So if we can show that $\widetilde{L}_{n,b}(y) \xrightarrow{P} L(y)$ (for y a continuity point of L), then we may apply this for $y = x[1+\xi]$ and $y = x/[1+\xi]$ (choosing ξ such that these are continuity points). Letting ξ tend to zero we obtain

$$L(x) - \nu \leq \widehat{L}_{n,b}(x) \leq L(x) + \nu$$

for arbitrary $\nu > \eta$, with probability tending to one; hence $\widehat{L}_{n,b}(x) \xrightarrow{P} L(x)$, if we can show that $\widetilde{L}_{n,b}(y) \xrightarrow{P} L(y)$ at continuity points y of L. To do this, we verify the assumptions of Theorem 1.

There exists a rate c_n such that $\alpha_n = n/c_n$ and $\delta_n = \sqrt{n}/c_n$, so that $\tau_n = \sqrt{n}$. The rate c_n, up to a slowly varying function, is the maximum of $n^{1/\alpha}$ and $n^{(\beta+1)/2}$. Assumption A follows from Theorem 3 of Jach et al (2012), with Z corresponding to a linear combination of an α-stable and a Gaussian random variables, whereas W corresponds to the square root of the sum of an $\alpha/2$-stable random variable plus D_ρ; the joint Fourier-Laplace transform is continuous in a neighborhood of the origin.

Next, observe that $\widehat{\theta}_n$ is the sample mean, and hence is Lipschitz with constant n^{-1}. The studentization is also Lipschitz (using the notation $y = [y_1, \cdots, y_n]'$ or $z = [z_1, \cdots, z_n]'$ for evaluations of the statistic):

$$
\begin{aligned}
|\widetilde{\sigma}(y)-\widetilde{\sigma}(z)| &= \frac{\left|n^{-1}\sum_{i=1}^n (y_i-\overline{y}_n)^2+\delta_n^{-2}D_\rho - n^{-1}\sum_{i=1}^n (z_i-\overline{z}_n)^2-\delta_n^{-2}D_\rho\right|}{\sqrt{n^{-1}\sum_{i=1}^n (y_i-\overline{y}_n)^2+\delta_n^{-2}D_\rho}+\sqrt{n^{-1}\sum_{i=1}^n (z_i-\overline{z}_n)^2+\delta_n^{-2}D_\rho}} \\
&\le \frac{\left|n^{-1}\sum_{i=1}^n (y_i-\overline{y}_n)^2 - n^{-1}\sum_{i=1}^n (z_i-\overline{z}_n)^2\right|}{\sqrt{n^{-1}\sum_{i=1}^n (y_i-\overline{y}_n)^2}+\sqrt{n^{-1}\sum_{i=1}^n (z_i-\overline{z}_n)^2}} \\
&\le \frac{n^{-1}\sum_{i=1}^n |y_i-z_i-\overline{y}_n+\overline{z}_n|\,|y_i-\overline{y}_n| + n^{-1}\sum_{i=1}^n |y_i-z_i-\overline{y}_n+\overline{z}_n|\,|z_i-\overline{z}_n|}{\sqrt{n^{-1}\sum_{i=1}^n (y_i-\overline{y}_n)^2}+\sqrt{n^{-1}\sum_{i=1}^n (z_i-\overline{z}_n)^2}} \\
&\le \frac{\sqrt{n^{-1}\sum_{i=1}^n (y_i-z_i-\overline{y}_n+\overline{z}_n)^2}\sqrt{n^{-1}\sum_{i=1}^n (y_i-\overline{y}_n)^2}}{\sqrt{n^{-1}\sum_{i=1}^n (y_i-\overline{y}_n)^2}+\sqrt{n^{-1}\sum_{i=1}^n (z_i-\overline{z}_n)^2}} \\
&+ \frac{\sqrt{n^{-1}\sum_{i=1}^n (y_i-z_i-\overline{y}_n+\overline{z}_n)^2}\sqrt{n^{-1}\sum_{i=1}^n (z_i-\overline{z}_n)^2}}{\sqrt{n^{-1}\sum_{i=1}^n (y_i-\overline{y}_n)^2}+\sqrt{n^{-1}\sum_{i=1}^n (z_i-\overline{z}_n)^2}} \\
&\le n^{-1/2}\sum_{i=1}^n |y_i-z_i-\overline{y}_n+\overline{z}_n| \\
&\le 2n^{-1/2}\,\|y-z\|_1,
\end{aligned}
$$

using respectively $D_\rho > 0$, the inequality $|y^2-z^2| \le |y-z|\,(|y|+|z|)$, the Cauchy-Schwarz inequality, and the triangle inequality. Hence setting $\kappa_n(x) = n^{-1/2}(1+2|x|)$, we obtain $b\,\delta_b\,\kappa_b(x) = b\,c_b^{-1}\,(1+2|x|) = O(b^\zeta)$ for any $\zeta > 1-(\alpha^{-1}\vee(\beta+1)/2)$. We can apply Theorem 1 provided that $\zeta < a\wedge 1$; this is automatically satisfied if $a\le 1$, but if $a<1$ we require the additional condition that $1-a<\alpha^{-1}\vee(\beta+1)/2$. $\square$

References

[1] Ango-Nze, P., Dupoiron, S., and Rios, R. (2003). Subsampling under weak dependence conditions. Technical Report DT2003-42, CREST.

[2] Bai, S., Taqqu, M.S., and Zhang, T. (2015a). Personal communication.

[3] Bai, S., Taqqu, M.S., and Zhang, T. (2015b). Self-normalized block sampling for light and heavy-tailed time series under short and long-range dependence. Working paper.

[4] Doukhan, P. and Louhichi, S. (1999). A new weak dependence and applications to moment inequalities. *Stochastic Processes and Their Applications* **84**, 313–343.

[5] Doukhan, P., Prohl, S., and Robert, C. (2011). Subsampling weakly dependent time series and applications to extremes. *Test* **20**, 447–479.

[6] Fitzsimmons, P. and McElroy, T. (2010). On joint Fourier-Laplace transforms. *Communications in Statistics – Theory and Methods* **39**, 1883-1885.

[7] Jach, A., McElroy, T.S., and Politis, D.N. (2012). Subsampling inference for the mean of heavy-tailed long-memory time series. *Journal of Time Series Analysis* **33**, 96–111.

[8] Politis, D.N., Romano, J. P., and Wolf, M. (1999). *Subsampling.* New York: Springer.

[9] Zhang, T. (2016). Personal communication.