Bernoulli **13**(2), 2007, 473–491
DOI: 10.3150/07-BEJ5123

Estimation of the memory parameter of the infinite-source Poisson process

GILLES FAŸ[1], FRANÇOIS ROUEFF[2] and PHILIPPE SOULIER[3]

[1]*Laboratoire Paul-Painlevé, Université de Lille 1, 59655 Villeneuve d'Ascq Cedex, France.*
E-mail: gilles.fay@univ-lille1.fr
[2]*École Nationale Supérieure des Télécommunications, 46, rue Barrault, 75634 Paris Cedex 13, France.*
E-mail: roueff@tsi.enst.fr
[3]*Modal'X, Université de Paris X, 200 avenue de la République, 92000 Nanterre, France.*
E-mail: philippe.soulier@u-paris10.fr

Long-range dependence induced by heavy tails is a widely reported feature of internet traffic. Long-range dependence can be defined as the regular variation of the variance of the integrated process, and half the index of regular variation is then referred to as the Hurst index. The infinite-source Poisson process (a particular case of which is the $M/G/\infty$ queue) is a simple and popular model with this property, when the tail of the service time distribution is regularly varying. The Hurst index of the infinite-source Poisson process is then related to the index of regular variation of the service times. In this paper, we present a wavelet-based estimator of the Hurst index of this process, when it is observed either continuously or discretely over an increasing time interval. Our estimator is shown to be consistent and robust to some form of non-stationarity. Its rate of convergence is investigated.

Keywords: heavy tails; internet traffic; long-range dependence; Poisson point processes; semiparametric estimation; wavelets

1. Introduction

We consider the infinite-source Poisson process with random transmission rate defined by

$$X(t) = \sum_{\ell \in \mathbb{N}} U_\ell \mathbb{1}_{\{t_\ell \le t < t_\ell + \eta_\ell\}}, \qquad t \ge 0, \tag{1.1}$$

where the arrival times $\{t_\ell\}_{\ell \ge 0}$ are the points of a unit-rate homogeneous Poisson process on the positive half-line, independent of the initial conditions; and the durations and transmission rates $\{(\eta_\ell, U_\ell)\}$ are independent and identically distributed random variables with values in $(0, \infty) \times \mathbb{R}$ and independent of the Poisson process and of the initial conditions. This process was considered by Resnick and Rootzén [12] and Mikosch *et al.* [9], among others. The $M/G/\infty$ queue is a special case, for $U_\ell \equiv 1$. An important motivation for the infinite-source Poisson process is to model the instantaneous rate of the workload going though an internet link. Although overly simple models are generally not relevant for internet traffic at the packet level, it is generally admitted that rather simple models can be used for higher-level (the so-called *flow level*) traffic such as TCP or HTTP sessions, one of them being the infinite-source Poisson process (see Barakat *et al.* [1]). One way to empirically analyse internet traffic at the flow level using the infinite-source

Poisson process would consist in retrieving all the variables $\{t_\ell, \eta_\ell, U_\ell\}$ involved in the observed traffic during a given period of time, but this would require the collection of all the relevant information in the packets headers (such as source and destination addresses) for the purpose of separating the aggregated workload into transmission rates at a pertinent level; see Duffield *et al.* [3] for many insights into this problem.

It is well known that heavy tails in the durations $\{\eta_k\}$ result in long-range dependence of the process $X(t)$. Long-range dependence can be defined by the regular variation of the autocovariance of the process or more generally by the regular variation of the variance of the integrated process:

$$\operatorname{var}\left(\int_0^t X(s)\,\mathrm{d}s\right) = L(t)t^{2H},$$

where L is a slowly varying function at infinity and $H > 1/2$ is often refered to as the *Hurst index* of the process. For the infinite-source Poisson process, the Hurst index H is related to the tail index α of the durations by the relation $H = (3-\alpha)/2$. The long-range dependence property has motivated many empirical studies of internet traffic and theoretical ones concerning its impact on queuing (these questions are studied in the $M/G/\infty$ case in Parulekar and Makowki [10]).

However, to the best of our knowledge, no statistical procedure to estimate H has been rigorously justified. It is the aim of this paper to propose an estimator of the Hurst index of the infinite-source Poisson process, and to derive its statistical properties. We propose to estimate H (or equivalently α) from a path of the process $X(t)$ over a finite interval $[0, T]$, observed either continuously or discretely. In practice this can be done by counting all the packets going through some point of the network and then collecting local traffic rate measurements. Our estimator is based on the so-called wavelet coefficients of a path. There is a wide literature on this methodology for estimating long-range dependence, starting as long ago as Wornell and Oppenheim [13], but we are not aware of rigorous results for non-Gaussian or non-stable processes. The main contribution of this paper is thus the proof of the consistency of our estimator. We also investigate the rate of convergence of the estimator in the case $\alpha > 1$. If the process is observed continuously, the rate of convergence is good. In the case of discrete observations, the rate is much smaller. Also, the choice of the tuning parameters of the estimators is much more restricted in the latter case, and practitioners should perhaps be aware of this; see Section 4.3 for details.

The process X is formally defined in Section 2. We state our assumptions and, using a point-process representation of X, we establish some of its main properties. The wavelet coefficients are defined and the scaling property of their variances is obtained in Section 3. The estimator is defined and its properties are established in Section 4. The Appendix contains technical lemmas.

2. Basic properties of the model

2.1. Assumptions

We now introduce the complete assumption on the joint distribution of the transmissions rates and durations.

Assumption 1. (i) *The random vectors* $\{(\eta, U),\ (\eta_\ell, U_\ell),\ \ell \in \mathbb{Z}\}$ *are independent with common distribution* ν *on* $(0,\infty) \times \mathbb{R}$ *and independent of the homogeneous Poisson point process on the real line with points* $\{t_\ell\}_{\ell\in\mathbb{Z}}$ *such that* $t_\ell < t_{\ell+1}$ *for all* ℓ *and* $t_{-1} < 0 \le t_0$.

(ii) *There exists a positive integer* p^* *such that* $\mathbb{E}[|U|^{p^*}] < \infty$.

(iii) *There exist a real number* $\alpha \in (0,2)$ *and positive functions* $L_0, \dots, L_{p^*}$ *slowly varying at infinity such that, for all* $t > 0$ *and* $p = 0, \dots, p^*$,

$$H_p(t) := \mathbb{E}\big[|U|^p \mathbb{1}_{\{\eta>t\}}\big] = L_p(t) t^{-\alpha}. \tag{2.1}$$

Since $\eta > 0$, the functions H_p are continuous at zero and $H_p(0) = \mathbb{E}[|U|^p]$. Condition (2.1) is equivalent to saying that the functions H_p, $p = 0, 1, \dots, p^*$, are regularly varying with index $-\alpha$. If $\alpha > 1$ and $p^* \ge 2$, Assumption 1 and Karamata's theorem imply the following asymptotic equivalence:

$$\mathbb{E}[U^2\{\eta - t\}_+] = \mathbb{E}\left[U^2 \int_{v=t}^{\infty} \mathbb{1}_{\{v<\eta\}}\, \mathrm{d}v\right] = \int_{v=t}^{\infty} H_2(v)\, \mathrm{d}v \sim \frac{1}{\alpha - 1} L_2(t) t^{1-\alpha}. \tag{2.2}$$

Remark 2.1. Assumption 1 will be used with $p^* = 2$ to prove the regular variation of the autocovariance function of the process X and with $p^* = 4$ to prove consistency of our estimators. It can be related to the theory of multivariate regular variation (see, for instance, Maulik *et al.* [7]). But the definitions of multivariate regular variation involve vague convergence and do not necessarily ensure the convergence of moments required here.

Remark 2.2. We do not assume that U is non-negative. This allows us to consider applications other than teletraffic modelling. For instance, the process X could be used to model the volatiliy of some financial time series.

Remark 2.3. We will often have to separate the cases $\mathbb{E}[\eta] = \infty$ and $\mathbb{E}[\eta] < \infty$. These cases are respectively implied by $\alpha < 1$ and $\alpha > 1$. If $\alpha = 1$, the finiteness of $\mathbb{E}[\eta]$ depends on the precise behaviour of L_0 at infinity.

Example 2.1. Assumption 1 implies in particular that the tail of the distribution of η is regularly varying with index α. This in turns implies Assumption 1 if U and η are independent and $\mathbb{E}[|U|^{p^*}] < \infty$, in which case the functions L_p differ by a multiplicative constant.

Example 2.2. Assumption 1 also holds in the following case which is of interest in teletraffic modelling. In a TCP/IP traffic context, η and U represent respectively the duration of a download session and its intensity (bit rate). Then $W := U\eta$ represents the amount of transmitted data. We assume that, for some $u_0 > 0$, there exist two regimes, $U \ge u_0$ (xDSL/LAN/cable connection) and $U \in (0, u_0)$ (RTC connection), such that the following statements hold:

- The distribution of W given $U = u \ge u_0$ is heavy-tailed and independent of u: $\mathbb{P}(W \ge w | U = u) = L(w) w^{-\alpha}$.
- The distribution of W given $U = u \in (0, u_0)$ is light-tailed uniformly with respect to u. For instance, we assume expontially decaying tails, $\mathbb{P}(W \ge w | U = u) \le \exp(-\beta w^{-\gamma})$, for some $\beta > 0$ and $\gamma > 0$.

An explicit example for two such regimes is obtained when the conditional density of W given $U = u$ is equal to $\alpha w^{-\alpha-1}\mathbb{1}_{\{w\geq 1\}}$ if $u \geq u_0$ and $\exp(-w)$ if $u < u_0$.

Concerning the distribution of U we only assume that:

- $\mathbb{P}(U \geq u_0) > 0$, $\mathbb{E}[|U|^{-\alpha-\epsilon}] < \infty$ for some $\epsilon > 0$, and $\mathbb{E}[|U|^{p^*}] < \infty$.

Then (2.1) holds for $p \leq p^*$. Indeed,

$$\begin{aligned}\mathbb{E}\big[U^p\mathbb{1}_{\{\eta>t\}}\mathbb{1}_{\{U\geq u_0\}}\big] &= \mathbb{E}\big[U^p\mathbb{1}_{\{W>Ut\}}\mathbb{1}_{\{U\geq u_0\}}\big]\\ &= \mathbb{E}\big[U^pL(Ut)(Ut)^{-\alpha}\mathbb{1}_{\{U\geq u_0\}}\big]\\ &= L(t)t^{-\alpha}\mathbb{E}\big[U^{p-\alpha}L(Ut)/L(t)\mathbb{1}_{\{U\geq u_0\}}\big].\end{aligned} \tag{2.3}$$

Since L is slowly varying at infinity, $\lim_{t\to\infty} L(ut)/L(t) = 1$, uniformly with respect to u in compact sets of $(0, +\infty)$, and there exists $t_0 > 0$ such that, for $u \geq u_0$, $t \geq t_0$,

$$\frac{L(ut)}{L(t)} \leq (1+\alpha)u^{\alpha/2};$$

see, for example, Resnick ([11], Proposition 0.8). Then, by the dominated convergence theorem,

$$\lim_{t\to\infty}\mathbb{E}\big[U^{p-\alpha}L(Ut)/L(t)\mathbb{1}_{\{U\geq u_0\}}\big] = \mathbb{E}\big[U^{p-\alpha}\mathbb{1}_{\{U\geq u_0\}}\big]. \tag{2.4}$$

Consider now the low-bit-rate regime. Since, for all $x > 0$, $\exp\{-\beta x^\gamma\} \leq Cx^{-\alpha-\epsilon}$ for some positive constant C, we have

$$\mathbb{E}\big[U^p\mathbb{1}_{\{\eta>t\}}\mathbb{1}_{\{U<u_0\}}\big] \leq \mathbb{E}\big[U^p\exp\{-\beta(Ut)^\gamma\}\mathbb{1}_{\{U<u_0\}}\big] \leq Ct^{-\alpha-\epsilon}\mathbb{E}\big[U^{p-\alpha-\epsilon}\mathbb{1}_{\{U<u_0\}}\big].$$

Using the assumption on U, since $p \geq 0$, the rightmost expectation in the previous display is finite and we obtain that

$$\lim_{t\to\infty} t^\alpha L^{-1}(t)\mathbb{E}\big[U^p\mathbb{1}_{\{\eta>t\}}\mathbb{1}_{\{U<u_0\}}\big] = 0.$$

Together with (2.3) and (2.4), this implies that, as $t \to \infty$, $\mathbb{E}[U^p\mathbb{1}_{\{\eta>t\}}]t^\alpha \sim L(t)\mathbb{E}[U^{p-\alpha} \times \mathbb{1}_{\{U\geq u_0\}}]$ hence is slowly varying and Assumption 1 holds.

2.2. Point-process representation and stationary version

Let $\mathcal{N}$ denote a Poisson point process on a set E endowed with a σ-field $\mathcal{E}$ with intensity measure μ, that is, a random measure such that for any disjoint $A_1, \ldots, A_p$ in $\mathcal{E}$, $\mathcal{N}(A_1), \ldots, \mathcal{N}(A_p)$ are independent random variables with Poisson law with respective parameters $\mu(A_i)$, $i = 1, \ldots, p$. The main property of Poisson point processes that we will use is the following cumulant formula (see, for instance, Resnick [11]: Chapter 3). For any positive integer p and functions $f_1, \ldots, f_p$ such that $\int |f_i|\,\mathrm{d}\mu < \infty$ and $\int |f_i|^p\,\mathrm{d}\mu < \infty$ for all $i = 1, \ldots, p$, the pth-order joint cumulant of $\mathcal{N}(f_1), \ldots, \mathcal{N}(f_p)$ exists and is given by

$$\operatorname{cum}(\mathcal{N}(f_1), \ldots, \mathcal{N}(f_p)) = \int f_1 \cdots f_p\,\mathrm{d}\mu. \tag{2.5}$$

Let N_S be the point processes on $\mathbb{R} \times (0, \infty) \times \mathbb{R}$ with points $(t_\ell, \eta_\ell, U_\ell)_{\ell \in \mathbb{Z}}$, that is $N_S = \sum_{\ell \in \mathbb{Z}} \delta_{t_\ell, \eta_\ell, U_\ell}$. Under Assumption 1(i), it is a Poisson point process with intensity measure $\mathrm{Leb} \otimes \nu$, where Leb is the Lebesgue measure on $\mathbb{R}$. For $t, u \in \mathbb{R}$, define

$$\begin{aligned} A_t &= \{(s, v) \in \mathbb{R} \times \mathbb{R}_+ \mid s \le t < s + v\}, \\ B_u &= \{\lambda u \mid \lambda \in [1, \infty)\}. \end{aligned}$$

We can now show that if $\mathbb{E}[\eta] < \infty$, then one can define a stationary version for X and provide its second-order properties.

Proposition 2.1. *If Assumption* 1(i) *holds and* $\mathbb{E}[\eta] < \infty$, *then the process*

$$X_S(t) = \sum_{\ell \in \mathbb{Z}} U_\ell \mathbb{1}_{\{t_\ell \le t < t_\ell + \eta_\ell\}} \tag{2.6}$$

is well defined and strictly stationary. It has the point-process representation

$$X_S(t) = \int_0^\infty N_S(A_t \times B_u)\,\mathrm{d}u - \int_{-\infty}^0 N_S(A_t \times B_u)\,\mathrm{d}u. \tag{2.7}$$

Let $K_0 = \sup\{\ell > 0 \mid t_{-\ell} + \eta_{-\ell} > 0\}$, $\tilde{U}_\ell = U_{-\ell}$ *and* $\tilde{\eta}_\ell = \eta_{-\ell} + t_{-\ell}$. *Then, for all* $t \ge 0$,

$$X_S(t) = \sum_{\ell=1}^{K_0} \tilde{U}_\ell \mathbb{1}_{\{t < \tilde{\eta}_\ell\}} + X(t). \tag{2.8}$$

If, moreover, $p^* \ge 2$, *then* X_S *has finite variance and*

$$\begin{aligned} \mathbb{E}[X_S(t)] &= \mathbb{E}[U\eta], \\ \operatorname{cov}(X_S(0), X_S(t)) &= \mathbb{E}[U^2(\eta - t)_+] = \int_t^\infty H_2(v)\,\mathrm{d}v. \end{aligned}$$

Remark 2.4. Note that if $\alpha > 1$, then $\mathbb{E}[\eta] < \infty$ and, by Karamata's theorem,

$$\operatorname{cov}(X_S(0), X_S(t)) \sim \frac{1}{\alpha - 1} L_2(t) t^{1-\alpha}, \qquad t \to +\infty.$$

Proof. The number of non-vanishing terms in the sum (2.6) is $N_S(A_t \times \mathbb{R})$ and has a Poisson distribution with mean $\mathbb{E} \int_{\mathbb{R}} \mathbb{1}_{A_t}(s, \eta)\,\mathrm{d}s = \mathbb{E}[\eta]$. Thus X_S is well defined and stationary since N_S is stationary. The number of indices $\ell > 0$ such that $t_{-\ell} + \eta_{-\ell} > 0$ is $N_S(A_0 \times \mathbb{R})$, hence if K_0 is the largest of those ℓs, it is almost surely finite and

$$\sum_{t_\ell < 0} U_\ell \mathbb{1}_{\{t_\ell \le t < t_\ell + \eta_\ell\}} = \sum_{\ell=1}^{K_0} \tilde{U}_\ell \mathbb{1}_{\{t < \tilde{\eta}_\ell\}}.$$

Hence (2.8).

The point-process representation (2.7) and formule (2.5) and (2.2) finally yield the given expressions for the mean and covariance. □

Relation (2.8) shows that the stationary version X_S can be defined by changing the initial condition of the system. More generally, one could consider *any* initial conditions, that is, any process defined as on the right-hand side of (2.8) with K_0 and $\tilde{\eta}_\ell$, $\ell > 0$ finite. Since the initial conditions almost surely vanish after a finite period, they have a negligible impact on the estimation procedure. Thus, our result on X easily generalizes to any such initial conditions, and, in particular, to the stationary version X_S, when it exists.

Applying similar arguments as those used for showing Proposition 2.1, we obtain:

Proposition 2.2. *The process X admits a point-process representation*

$$X(t) = \int_0^\infty N_S(A_t^+ \times B_u)\,\mathrm{d}u - \int_{-\infty}^0 N_S(A_t^+ \times B_u)\,\mathrm{d}u, \tag{2.9}$$

where $A_t^+ = A_t \cap \mathbb{R}_+^2$.

If Assumption 1 holds with $p^* \geq 2$, then the process X is non-stationary with expectation and autocovariance function given, for $s \leq t$, by

$$\mathbb{E}[X(t)] = \mathbb{E}[U(\eta \wedge t)],$$

$$\operatorname{cov}(X(s), X(t)) = \mathbb{E}[U^2\{s - (t - \eta)_+\}_+] = \int_{t-s}^t H_2(v)\,\mathrm{d}v.$$

By the uniform convergence theorem for slowly varying functions, the following asymptotic equivalence of the covariance holds. For any $\alpha \in (0, 2)$ and all $t > s > 0$, as $T \to \infty$,

$$\operatorname{cov}(X(Tt), X(Ts)) \sim C L_2(T) T^{1-\alpha} \tag{2.10}$$

with $C = \int_{t-s}^s v^{-\alpha}\,\mathrm{d}v$.

In accordance with the notation in use in the context of long-memory processes, we can define the Hurst index of the process X as $H = (3-\alpha)/2$, because the variance of the process integrated between 0 and T increases as T^{2H}. If $\alpha < 1$, then $H > 1$. This case has been considered, for instance, by Resnick and Rootzén [12].

3. Wavelet coefficients

3.1. Continuous observation

Let ψ be a bounded real-valued function with compact support in $[0, M]$ and such that

$$\int_0^M \psi(s)\,\mathrm{d}s = 0. \tag{3.1}$$

For integers $j \geq 0$ and $k \in \mathbb{Z}$, define

$$\psi_{j,k}(s) = 2^{-j/2}\psi(2^{-j}s - k). \tag{3.2}$$

The wavelet coefficients of the path are defined as

$$d_{j,k} = \int_0^\infty \psi_{j,k}(s)X(s)\,\mathrm{d}s \tag{3.3}$$

(see, for example, Cohen [2]). Assume that a path of the process X is observed continuously between times 0 and T. Since $\psi_{j,k}$ has support in $[k2^j, (k+M)2^j]$, the coefficients $d_{j,k}$ can be computed for all (j,k) such that $T2^{-j} \geq M$ and $k = 0, 1, \ldots, T2^{-j} - M$.

According to Lemma A.1, one may define, for all j and k,

$$d^S_{j,k} = \sum_{\ell \in \mathbb{Z}} U_\ell \int_{t_\ell}^{t_\ell + \eta_l} \psi_{j,k}(s)\,\mathrm{d}s. \tag{3.4}$$

As stated in Lemma A.1, if $\mathbb{E}[\eta] < \infty$, we have $d^S_{j,k} = \int_0^\infty \psi_{j,k}(s)X_S(s)\,\mathrm{d}s$. Nevertheless, even if $\mathbb{E}[\eta] = \infty$, the sequence of coefficients at a given scale j, $\{d^S_{j,k}, k \in \mathbb{Z}\}$, is stationary. Moreover, the definition (3.4) yields:

Lemma 3.1. *Let Assumption 1 hold with $p^* \geq 2$. We have*

$$\mathbb{E}[d^S_{j,k}] = 0, \qquad \operatorname{var}(d^S_{j,k}) = \mathcal{L}(2^j)2^{(2-\alpha)j}, \tag{3.5}$$

where

$$\mathcal{L}(z) := z^\alpha \int_0^\infty \int_{-\infty}^\infty \left(\int_{-\infty}^\infty \left\{ \int_t^{t+vz^{-1}} \psi(s)\,\mathrm{d}s \right\}^2 \mathrm{d}t \right) w^2 \nu(\mathrm{d}v, \mathrm{d}w) \tag{3.6}$$

is slowly varying as $z \to \infty$. More precisely, we have the asymptotic equivalence

$$\mathcal{L}(z) \sim C_{\mathcal{L}} L_2(z) \qquad \text{as } z \to \infty, \tag{3.7}$$

with $C_{\mathcal{L}} = \alpha \int_0^\infty \int_{-\infty}^\infty \{\int_x^{x+y} \psi(s)\,\mathrm{d}s\}^2 \,\mathrm{d}x\, t^{-\alpha-1}\,\mathrm{d}t > 0$.

The proof of (3.5) is a straightforward application of (2.5), and the proof of the asymptotic equivalence (3.7) is obtained by standard arguments on slowly varying functions. A detailed proof can be found in Faÿ *et al.* [4].

3.2. Wavelet coefficients in discrete time

Let ϕ be a bounded $\mathbb{R} \to \mathbb{R}$ function with compact support included in $[-M+1, 1]$ and such that

$$\sum_{k \in \mathbb{Z}} \phi(t-k) = 1, \qquad t \in \mathbb{R}. \tag{3.8}$$

Let I_ϕ denote the operator defined on the set of functions $x:\mathbb{R}\to\mathbb{R}$ by

$$\mathrm{I}_\phi[x](t)=\sum_{k\in\mathbb{Z}}x(k)\phi(t-k). \tag{3.9}$$

The wavelet coefficients of x are then defined as the wavelet coefficients of $\mathrm{I}_\phi[x]$.

From a computational point of view, it is convenient to chose ϕ and ψ to be the so-called father and mother wavelets of a multiresolution analysis; see, for instance, Meyer [8]. The simplest choice is to take ϕ and ψ to be associated with the Haar system, in which case $M=1$, $\phi=\mathbb{1}_{[0,1)}$ and $\psi=\mathbb{1}_{[0,1/2)}-\mathbb{1}_{[1/2,1)}$.

If the process X is observed discretely, we denote its wavelet coefficients by

$$d^D_{j,k}=\int\psi_{j,k}(s)\mathrm{I}_\phi[X](s)\,\mathrm{d}s. \tag{3.10}$$

If we observe $X(0),X(1),\dots,X(T-1)$, for some positive integer T, we can compute $d^D_{j,k}$ for all j,k such that $0\le k\le 2^{-j}(T-M+1)-M$. Roughly, for $2^j\ge T/M$, no coefficients can be computed and if $2^j<T/M$ the number of computable wavelet coefficients at scale 2^{-j} is of order $T2^{-j}+1-M$ for j and T large.

Remark 3.1. Observe that the choice of time units is unimportant here. Indeed, in Assumption 1, changing the time units simply amounts to adapting the slowly varying functions L_k and the rate of the arrival process $\{t_k\}$. Clearly these adaptations do not modify our results since precise multiplicative constants are not considered.

3.3. Averaged observations

We describe now a third observation scheme for which our results can easily be extended. Suppose that T is a positive integer and that we observe local averages of the trajectory

$$\overline{X}(k):=\int_k^{k+1}X(t)\,\mathrm{d}t=\int X(t)\phi_H(t-k)\,\mathrm{d}t,\qquad k=0,1,\dots,T-1,$$

where $\phi_H:=\mathbb{1}_{[0,1]}$ is the Haar wavelet. Let $\bar{\mathrm{I}}_\phi$ denote the operator on locally integrable functions x defined by

$$\bar{\mathrm{I}}_\phi[x](t)=\sum_{k\in\mathbb{Z}}\left(\int x(s)\phi_H(s-k)\,\mathrm{d}s\right)\phi(t-k).$$

For this observation scheme, as in Section 3.2, one may compute the wavelet coefficients of the function $\bar{\mathrm{I}}_\phi[X]$ at all scale and location indices (j,k) such that $0\le k\le 2^{-j}(T-M+1)-M$. If $\phi=\phi_H$ and ψ is the Haar mother wavelet, $\psi=\mathbb{1}_{[0,1/2)}-\mathbb{1}_{[1/2,1)}$, then the wavelet coefficients of $\bar{\mathrm{I}}_\phi[X]$ are precisely the continuous wavelet coefficients defined in (3.3). For any other choice of ϕ and ψ, this is no longer true. We will not treat this case, but all our results can be extended at the cost of further technicalities.

4. Estimation

Tail index estimation methods do not seem appropriate here for estimating the parameter α. Indeed, α is the tail index of the unobserved durations $\{\eta_k\}$, whereas the observed process $X(t)$ always has finite variance ($\mathbb{E}[|X(t)|^p] < \infty$ if and only if $\mathbb{E}[|U^p|] < \infty$ and the marginal distribution of $X(t)$ is Poisson if $U = 1$ almost surely). But as shown by Proposition 2.2, α is related to the second-order properties of the process: the coefficient $H = (3-\alpha)/2$ can be viewed as its Hurst index, that is, H governs the rate of decay of the autocovariance function of the process. Therefore it seems natural to use an estimator of the Hurst index.

4.1. The estimator

Lemma 3.1 provides the rationale for a minimum contrast estimator of α which is related to the local Whittle estimator; cf. Künsch [6]. Let $\mathbf{d}_{j,k}$ denote the wavelet coefficients which are actually available; these may be obtained from continuous-time ($\mathbf{d}_{j,k} = d_{j,k}$) or discrete-time ($\mathbf{d}_{j,k} = d^D_{j,k}$) observations. Let Δ be a set of indices (j,k) of available wavelet coefficients. Denote the mean scale index over Δ by

$$\delta := \frac{1}{\#\Delta} \sum_{(j,k)\in\Delta} j.$$

The reduced local Whittle contrast function is

$$\hat{W}(\alpha') = \log\left(\sum_{(j,k)\in\Delta} \frac{\mathbf{d}^2_{j,k}}{2^{(2-\alpha')j}} \right) + \delta \log(2)(2-\alpha'). \tag{4.1}$$

The local Whittle estimator of α is then defined as

$$\hat{\alpha} := \arg\min_{\alpha'\in(0,2)} \hat{W}(\alpha'). \tag{4.2}$$

In order to simplify the proof of our result, we henceforth take Δ to be of the form

$$\Delta = \{(j,k);\, J_0 < j \le J_1, 0 \le k \le n_j - 1\},$$

with $J = \max\{j; 2^j \le (T-M+1)/(M+1)\}$, $n_j = 2^{J-j}$ and integers J_0 and J_1 such that

$$0 < J_0 < J_1 \le J. \tag{4.3}$$

The sequence of integers J depends on T in such a way that $2^J \asymp T$. Note that the dependence of the sequences J, J_0, J_1, n_j etc. on T is suppressed in our notation.

4.2. Consistency

Our estimator is consistent in the potentially unstable case, that is when α is not assumed to be in $(1,2)$, provided that the assumptions on the functions ϕ and ψ are strengthened. We assume that

$$\int_{-\infty}^{\infty} s\psi(s)\,\mathrm{d}s = 0, \tag{4.4}$$

and there exist constants a and b such that, for all $t \in \mathbb{R}$,

$$\sum_{k\in\mathbb{Z}} k\phi(t-k) = a + bt. \tag{4.5}$$

These conditions are not satisfied by the Haar wavelet, but hold for any Daubechies wavelets; see Cohen [2].

Theorem 4.1. *Let Assumption 1 hold with $p^* \geq 4$. Assume that J_0 and J_1 depend on T in such a way that*

$$\lim_{T\to\infty} J_0 = \lim_{T\to\infty} (J_1 - J_0) = \infty, \tag{4.6}$$

$$\limsup_{T\to\infty} J_0/J < 1/\alpha, \tag{4.7}$$

$$\limsup_{T\to\infty} J_1/J < 1/(2-\alpha). \tag{4.8}$$

Then $\hat{\alpha}$ is a consistent estimator of α. Moreover if $\alpha \in (1,2)$, then conditions (4.4), (4.5) and (4.8) are not necessary for the same result to hold.

Remark 4.1. Conditions (4.6), (4.7) and (4.8) are satisfied by the choice $J_0 = \lfloor J/2 \rfloor$ and $J_1 = \lfloor J/2 + \log(J) \rfloor$.

Proof of Theorem 4.1. For clarity of notation, we denote $\sum_j = \sum_{j=J_0+1}^{J_1}$, $\Delta_j := \{k : (j,k) \in \Delta\}$ and $\#\Delta_j = n_j$. Elementary computations give

$$\delta = J_0 + 2 + (J_0 - J_1)/(2^{J_1 - J_0} - 1) \tag{4.9}$$

so that $\delta - (J_0 + 2) \to 0$ under (4.6). By Karamata's representation theorem, the slowly varying function $\mathcal{L}$ defined in (3.6) can be written as

$$\mathcal{L}(z) = c\big(1 + r(z)\big)\exp\left\{\int_1^z \frac{\ell(s)}{s}\,\mathrm{d}s\right\},$$

with $c > 0$ and $\lim_{z\to\infty} \ell(z) = \lim_{z\to\infty} r(z) = 0$. Define $\mathcal{L}_0(z) = c\exp\{\int_1^z s^{-1}\ell(s)\,\mathrm{d}s\}$, $r^*(z) = \sup_{z'\geq z} |r(z')|$ and $\ell^*(z) = \sup_{z'\geq z} |\ell(z')|$. The functions r^* and ℓ^* are non-increasing and tend to zero at infinity. We now introduce some notation that will be used throughout the proof:

$$
\begin{aligned}
W(\alpha') &= \log\left(\sum_j 2^{(\alpha'-\alpha)j} n_j \mathcal{L}(2^j)\right) + \delta \log(2)(2-\alpha'), \\
W_0(\alpha') &= \log\left(\sum_j 2^{(\alpha'-\alpha)j} n_j \mathcal{L}_0(2^j)\right) + \delta \log(2)(2-\alpha'), \\
w_{j,0}(\alpha') &:= \frac{2^{(\alpha'-\alpha)j} n_j \mathcal{L}_0(2^j)}{\sum_{j'} 2^{(\alpha'-\alpha)j'} n_{j'} \mathcal{L}_0(2^{j'})}, \qquad w_j(\alpha') := \frac{2^{(\alpha'-\alpha)j} n_j \mathcal{L}(2^j)}{\sum_{j'} 2^{(\alpha'-\alpha)j'} n_{j'} \mathcal{L}(2^{j'})}, \\
v_j &= \mathcal{L}(2^j) 2^{(2-\alpha)j}, \qquad \Lambda_j = n_j^{-1} \sum_{k=0}^{n_j-1} \{v_j^{-1} (\mathbf{d}_{j,k})^2 - 1\}, \\
E(\alpha') &:= \sum_j w_j(\alpha') \Lambda_j.
\end{aligned}
$$

We have

$$
W(\alpha') - W_0(\alpha') = \log\left(1 + \frac{\sum_j 2^{(\alpha'-\alpha)j} n_j \mathcal{L}_0(2^j) r(2^j)}{\sum_j 2^{(\alpha'-\alpha)j} n_j \mathcal{L}_0(2^j)}\right).
$$

Here the fraction inside the logarithm is bounded by $r^*(2^{J_0})$, thus, for J large enough,

$$
\sup_{\alpha'} |W(\alpha') - W_0(\alpha')| \le C r^*(2^{J_0}).
$$

Standard algebra yields

$$
\begin{aligned}
W_0'(\alpha') &= \log 2 \sum_j w_{j,0}(\alpha')(j - \delta) \\
W_0''(\alpha') &= \log^2(2) \sum_j w_{j,0}(\alpha') \left(j - \sum_{j'} w_{j',0}(\alpha') j'\right)^2.
\end{aligned}
$$

By Lemma A.8, under (4.6),

$$
\lim_{T\to\infty} W_0'(\alpha) = 0, \qquad \lim_{T\to\infty} W_0''(\alpha) = 2.
$$

Thus, there exist $\eta > 0$ and $\zeta > 0$ such that

$$
\liminf_{T\to\infty} \inf_{\alpha' \in (\alpha-\eta, \alpha+\eta)} W_0''(\alpha') > \zeta.
$$

This implies that, for large T and some positive constant c,

$$
W(\hat{\alpha}) - W(\alpha) \ge W_0'(\alpha) \log(2)(\hat{\alpha} - \alpha) + c(\hat{\alpha} - \alpha)^2 - 2r^*(2^{J_0}). \tag{4.10}
$$

Since $W_0'(\alpha) \to 0$ and $|\hat{\alpha} - \alpha| \le 2$, this implies that, for all $\epsilon > 0$,

$$\limsup_{T\to\infty} \mathbb{P}\big((\hat{\alpha} - \alpha)^2 > \epsilon\big) \le \limsup_{T\to\infty} \mathbb{P}\big(W(\hat{\alpha}) - W(\alpha) \ge c\epsilon\big). \tag{4.11}$$

Write

$$\begin{aligned} \hat{W}(\alpha') &= W(\alpha') + \log\{1 + E(\alpha')\}, \\ W(\hat{\alpha}) - W(\alpha) &= \hat{W}(\hat{\alpha}) - \hat{W}(\alpha) - \log\{1 + E(\hat{\alpha})\} + \log\{1 + E(\alpha)\} \\ &\le 2 \sup_{\alpha' \in (0,2)} |\log\{1 + E(\alpha')\}|. \end{aligned} \tag{4.12}$$

Consistency will follow from (4.11) and (4.12) provided that we can prove that $\sup_{\alpha' \in (0,2)} |E(\alpha')| = o_P(1)$. If $\alpha > 1$, take $\varepsilon \in (0, (\alpha - 1)/2)$ such that $\limsup J_0/J < 1/(\alpha + \varepsilon)$, which is possible by assumption (4.7). Define

$$J_2 = \begin{cases} J_1, & \text{if } \alpha \le 1, \\ J_1 \wedge [J/(\alpha + \varepsilon)], & \text{if } \alpha > 1, \end{cases} \tag{4.13}$$

so that, for T large enough, $J_0 < J_2 \le J_1$. Write

$$E(\alpha') = \sum_{j=J_0+1}^{J_2} w_j(\alpha')\Lambda_j + \sum_{j=J_2+1}^{J_1} w_j(\alpha')\Lambda_j =: E_1(\alpha') + E_2(\alpha'),$$

with the convention that $\sum_{j=J_2+1}^{J_1} = 0$ if $J_2 = J_1$. By Lemma A.6,

$$\sup_{\alpha' \in (0,2)} |E_1(\alpha')| = O_P(2^{-\xi_1 J}), \tag{4.14}$$

for some positive ξ_1. Now treat E_2 for $\alpha > 1$ and $J_2 = [J/(\alpha + \varepsilon)] > J_1$. For all $\alpha' \in (0,2)$, we have $\alpha' - \alpha - 1 < -2\varepsilon$. Since $\mathcal{L}$ is slowly varying, we obtain, for some positive constant C, for all $j = J_2 + 1, J_2 + 2, \ldots, J_1$, $w_j(\alpha') \le C 2^{-\varepsilon(J_2 - J_0)}$. Using Lemma A.5, it follows that

$$\mathbb{E}\Big[\sup_{\alpha' \in (0,2)} |E_2(\alpha')|\Big] \le C(J_1 - J_2) 2^{-\varepsilon(J_2 - J_0)} = O(2^{-\xi_2 J}), \tag{4.15}$$

for some $\xi_2 > 0$ because $\limsup J_0/J < 1/(\alpha + \varepsilon)$. This concludes the proof. □

4.3. Rate of convergence in the stable case

Theorem 4.2. *Let Assumption 1 hold with $\alpha \in (1,2)$ and $p^* = 4$. Assume, moreover, that L_4 is bounded and that $\mathcal{L}(z) = c + O(z^{-\beta})$ with $c > 0$ and $\beta > 0$.*

If X is observed continuously on $[0, T]$, that is, $\mathbf{d}_{j,k} = d_{j,k}$, then the rate of convergence in probability of $\hat{\alpha}$ is $T^{-\beta/(2\beta+\alpha)}$, obtained for $J_0 = \lfloor J/\{2\beta + \alpha\}\rfloor$ and $J_1 = J$.

If X is observed at discrete time points $1, 2, \ldots, T$, that is, $\mathbf{d}_{j,k} = d^D_{j,k}$, then the rate of convergence in probability of $\hat{\alpha}$ is $T^{-\gamma/(2\gamma+\alpha)}$ with $\gamma = \beta \wedge (2-\alpha)$, obtained for $J_0 = \lfloor J/\{2\gamma+\alpha\} \rfloor$ and $J_1 = J$.

Remark 4.2. Observe that the choice of J_0 corresponding to the best rate for $\hat{\alpha}$ depends both on the unknown smoothness parameter β and on the parameter α itself. The case of discrete observations is similar to that of continuous-time observations but with the smoothness parameter β replaced by $\gamma = \beta \wedge (2-\alpha)$, resulting in a slower rate of convergence. This can be explained by the aliasing induced by the interpolation step (3.9). It is clear that these rates of convergence are the best possible for our estimator under the assumption on $\mathcal{L}$, since this choice of J_0 makes the squared bias and the variance of the same order of magnitude. However, to our knowledge, the best possible rate of convergence for the estimation of α under these observations schemes is an open question. In other words, whether our estimator is rate optimal remains unknown.

The rate of convergence of our estimator is derived under assumptions on the function $\mathcal{L}$. The following lemma allows us to check them through conditions on the joint distribution of (U, η).

Lemma 4.3. *Let Assumption 1 hold.*

(i) *If there exist positive constants c and β such that, as $t \to \infty$,*

$$L_2(t) = c + O(t^{-\beta}),$$

then there exists a constant c' such that, as $z \to \infty$,

$$\mathcal{L}(z) = c' + \begin{cases} O(z^{-\beta}), & \text{if } \beta < 2-\alpha, \\ O(z^{\alpha-2}\log z), & \text{if } \beta = 2-\alpha, \\ O(z^{\alpha-2}), & \text{if } \beta > 2-\alpha. \end{cases} \tag{4.16}$$

(ii) *If there exist positive constants c and β such that, as $t \to 0$,*

$$\mathbb{E}[U^2\{1-\cos(\eta t)\}] = c|t|^{-\alpha}\{1 + O(|t|^{\beta})\},$$

then there exists a constant c' such that, as $z \to \infty$,

$$\mathcal{L}(z) = c' + O(z^{-\beta}), \tag{4.17}$$

provided that ψ belongs to the Sobolev space $W^{(\alpha+\beta)/2-1}$, that is,

$$\int_{-\infty}^{\infty} (1+|\xi|)^{(\alpha+\beta)-2} |\psi^*(\xi)|^2 \, \mathrm{d}\xi < \infty, \tag{4.18}$$

where ψ^* denotes the Fourier transform of ψ,

$$\psi^*(\xi) = \int_0^M \psi(t) \mathrm{e}^{-\mathrm{i}\xi t} \, \mathrm{d}t. \tag{4.19}$$

Example 4.1. Assume that η has a Pareto distribution, that is, $\mathbb{P}(\eta > t) = (1 \vee t)^{-\alpha}$, and is independent of U. This corresponds to Lemma 4.3(i) with $\beta = \infty$, and we can easily compute an exact expression for the $O(z^{\alpha-2})$ term:

$$\mathcal{L}(z) = c' + \frac{\alpha\mathbb{E}[U^2]}{2-\alpha} z^{\alpha-2} + o(z^{\alpha-2}).$$

The best possible rate of convergence of $\hat{\alpha}$ is thus $T^{-(2-\alpha)/(4-\alpha)}$, regardless of the observation scheme.

Example 4.2. Let $\alpha \in (1,2)$ and suppose that η is the absolute value of a symmetric α-stable random variable. Then Assumption 1 holds, say, if U is independent of η and has sufficiently many finite moments, and

$$\mathbb{E}[\cos(\eta t)] = \exp(-\sigma|t|^{\alpha}) = 1 - \sigma|t|^{\alpha} + O(|t|^{2\alpha}).$$

By Lemma 4.3, the best possible rate of convergence of $\hat{\alpha}$ is thus $T^{-\gamma/(2\gamma+\alpha)}$ with $\gamma = \alpha$ for continuous-time observations and $\gamma = 2-\alpha$ for discrete-time observations.

In the following, we give a decomposition of the error valid under the assumption

$$0 < \liminf_{T\to\infty} \frac{J_0}{J} \le \limsup_{T\to\infty} \frac{J_0}{J} < 1.$$

Optimizing J_0 in this decomposition will then give the result. We use the same notation as in the proof of Theorem 4.1 with $J_1 = J$. We first give a first rough rate of convergence for $\hat{\alpha}$ by adapting the proof of Theorem 4.1. Under the present assumptions, $\mathcal{L}_0(z) = c$, which implies $W_0'(\alpha) = 0$, and $r^*(z) = O(z^{-\beta})$ as $z \to \infty$. Then, (4.10), (4.12), (4.14) and (4.15) yield

$$(\hat{\alpha} - \alpha)^2 = O_P(2^{-\xi J} + 2^{-\beta J_0}). \tag{4.20}$$

Since $\hat{\alpha}$ is consistent and α is an interior point of the parameter set, the first derivative of the contrast function vanishes at $\hat{\alpha}$ with probability tending to one. Hence

$$0 = \frac{\sum_{(j,k)\in\Delta} j 2^{(\hat{\alpha}-2)j} \mathbf{d}_{j,k}^2}{\sum_{(j,k)\in\Delta} 2^{(\hat{\alpha}-2)j} \mathbf{d}_{j,k}^2} - \delta\log(2).$$

By the definition of δ, this yields

$$\begin{aligned} 0 &= \sum_{(j,k)\in\Delta} (j-\delta) 2^{(\hat{\alpha}-2)j} \mathbf{d}_{j,k}^2 \\ &= \sum_{(j,k)\in\Delta} (j-\delta) 2^{(\alpha-2)j} \mathbf{d}_{j,k}^2 + \log(2)(\hat{\alpha}-\alpha) \sum_{(j,k)\in\Delta} j(j-\delta) 2^{(\tilde{\alpha}-2)j} \mathbf{d}_{j,k}^2 \end{aligned} \tag{4.21}$$

for a random $\tilde{\alpha}$ between α and $\hat{\alpha}$. By the definition of Λ_j, (4.21) implies that

$$\hat{\alpha} - \alpha = -\frac{\sum_j (j-\delta)2^{-j}\mathcal{L}(2^j)(1+\Lambda_j)}{\log 2 \sum_j j(j-\delta)2^{(\tilde{\alpha}-\alpha-1)j}\mathcal{L}(2^j)(1+\Lambda_j)}.$$

Denote the sum in the denominator by D, and write

$$\begin{aligned} D &= \sum_j j(j-\delta)2^{-j}\mathcal{L}(2^j) + \sum_j j(j-\delta)\mathcal{L}(2^j)2^{-j}\big(2^{(\tilde{\alpha}-\alpha)j} - 1\big)(1+\Lambda_j) \\ &\quad + \sum_j j(j-\delta)2^{-j}\mathcal{L}(2^j)\Lambda_j \\ &=: S + R_1 + R_2. \end{aligned}$$

Using Lemma A.8 and (4.9), one easily obtains that $S \sim 2^{1-J_0}$ as $J \to \infty$.

Using Lemma A.5, and the fact that $|\tilde{\alpha}-\alpha| \le |\hat{\alpha}-\alpha| = o_P(J^{-2})$, one similarly obatins $R_1 = o_P(2^{-J_0})$. To bound R_2, we proceed as for bounding $E(\alpha')$ in the proof of Theorem 4.1 (here with $\alpha' = \alpha > 1$): we write $\sum_j = \sum_{j=J_0+1}^{J_2} + \sum_{j=J_2+1}^{J}$ and apply Lemmas A.5 and A.6 to obtain $R_2 = o_P(2^{-J_0})$. Hence, we finally obtain

$$\hat{\alpha} - \alpha = \frac{2^{J_0}}{2\log 2}\left\{\sum_j (j-\delta)2^{-j}\mathcal{L}(2^j) + \sum_j (j-\delta)2^{-j}\mathcal{L}(2^j)\Lambda_j\right\}\{1 + o_p(1)\}. \tag{4.22}$$

In (4.22), the terms inside the curly brackets are interpreted as a deterministic bias term and a stochastic fluctuation term. The bias is bounded as follows:

$$2^{J_0}\sum_j (j-\delta)2^{-j}\mathcal{L}(2^j) = 2^{J_0}\sum_j (j-\delta)2^{-j}(\mathcal{L}(2^j) - c) = O(2^{-\beta J_0}). \tag{4.23}$$

In the case of continuous-time observations, that is, $\mathbf{d}_{j,k} = d^S_{j,k}$ or $\mathbf{d}_{j,k} = d_{j,k}$, we have

$$\sum_j (j-\delta)2^{-j}\mathcal{L}(2^j)\Lambda_j = O_P\big(2^{-J/2+(\alpha/2-1)J_0}\big). \tag{4.24}$$

Gathering this bound with (4.22) and (4.23), and setting $J_0 = J/(2\beta+\alpha)$, yields the first claim of Theorem 4.2, that is, $\hat{\alpha} - \alpha = O_P(2^{-\beta/(2\beta+\alpha)})$.

We now prove (4.24). Define $\beta_j = n_j^{-1}\sum_{k=0}^{n_j-1}\{v_j^{-1}(d^S_{j,k})^2 - 1\}$. Then $\beta_j = \Lambda_j$ if $\mathbf{d}_{j,k} = d^S_{j,k}$. Since $\alpha > 1$, Lemmas 3.1 and A.2 yield, for some positive constant C,

$$\mathbb{E}[\beta_j^2] = \operatorname{var}(\beta_j) \le C\frac{L_4(2^j)}{\mathcal{L}^2(2^j)}2^{\alpha j - J}. \tag{4.25}$$

Since $\mathcal{L}$ is bounded away from zero and L_4 is bounded by assumption, the ratio $L_4/\mathcal{L}^2$ is also bounded. The Minkowski inequality then yields, for some constant $C > 0$,

$$\mathbb{E}\left[\left(\sum_j (j-\delta)2^{-j}\mathcal{L}(2^j)\beta_j\right)^2\right]^{1/2} \le C2^{-J/2}\sum_j |j-\delta|2^{(\alpha/2-1)j}$$
$$= O\big(2^{-J/2+(\alpha/2-1)J_0}\big). \tag{4.26}$$

If $\mathbf{d}_{j,k} = d_{j,k}$, we use (A.3) in Lemma A.5, and obtain $\mathbb{E}[|\Lambda_j - \beta_j|] \le Cn_j^{-1/2}$ for some constant $C > 0$. Hence, in this case, since $-1/2 < \alpha/2 - 1$,

$$\mathbb{E}\left[\left|\sum_j (j-\delta)2^{-j}\mathcal{L}(2^j)(\Lambda_j - \beta_j)\right|\right] \le C2^{-J/2}\sum_j |j-\delta|\mathcal{L}(2^j)2^{-j/2}$$
$$= o\big(2^{-J/2+(\alpha/2-1)J_0}\big). \tag{4.27}$$

Inequalities (4.26) and (4.27) imply (4.24).

We now briefly adapt the previous proof to the case of discrete observations. Define $v_j^D = \mathbb{E}[(d_{j,k}^{SD})^2]$. Lemma A.4(iii) implies $v_j^D = v_j + O(1)$. Thus we have $v_j^D = \mathcal{L}_D(2^j)2^{(2-\alpha)j}$ and

$$\mathcal{L}_D(z) = \mathcal{L}(z) + O(z^{\alpha-2}) = c + O(z^{-\gamma}),$$

with $\gamma = \beta \wedge (2-\alpha)$. Then, defining

$$\Lambda_j^D = n_j^{-1}\sum_{k=1}^{n_j}\{(v_j^D)^{-1}\mathbf{d}_{j,k}^2 - 1\},$$

we obtain that (4.22) still holds with $\mathcal{L}_D$ and Λ^D replacing $\mathcal{L}$ and Λ, respectively. Lemma A.7 implies that Λ_j^D has the same order of magnitude as Λ_j, so that the stochastic fluctuation term has the same order of magnitude as in the previous case. The difference comes from the bias term, which is $O(2^{-\gamma J_0})$. Thus, $\hat{\alpha} - \alpha = O_P(2^{-\gamma J_0} + 2^{-\gamma J_0})$, and setting $J_0 = J/(2\gamma + \alpha)$ yields the second claim of Theorem 4.2.

5. Concluding remarks

In this work, we have proved the validity of a wavelet method for the estimation of the long-memory parameter of an infinite-source Poisson traffic model, either in a stable or in an unstable state, that is, when it does or does not converge to a stationary process. We have shown that a suitable choice of the scales in the estimator (see Remark 4.1) yields a consistent estimator in both situations, and checked that the estimator is robust to discrete data sampling.

However, the study of the rates raises some questions concerning the optimality of this estimator. To draw a comparison, suppose that one *directly* observes the durations $\eta_1, \dots, \eta_n$ of clients arriving at times $t_1, \dots, t_n$ in $[0, T]$. Then one can use the Hill estimator for estimating the tail index α. Since T and n are asymptotically proportional and $\eta_1, \dots, \eta_n$ are independent and identically distributed, the rates of this estimator are those derived in Hall and Welsh [5]. In particular, if η has a Pareto distribution, then a parametric rate $\sqrt{T}$ can be obtained. On the other hand, in the same situation, our wavelet estimator defined on the observations $\{X(t), t \in [0, T]\}$ has a dramatically deteriorating rate for α close to 2. It remains to establish whether this discrepancy comes from the choice of the estimator or from the fact that the durations η_k are not directly observed.

Finally, let us draw a practical conclusion from our study. Care precaution should be taken with the choice of the scales used in the estimation, as shown by the conditions on J_0 and J_1. In particular, if only discrete observations are available, the best possible rate of convergence is obtained for a much larger value of J_0 than if continous observations are available. Too small a value of J_0 will induce an important bias for finite samples. Practitioners should be aware of this restriction and be careful in the interpretation of the results. These questions will be tackled numerically in a future work.

Appendix: Technical results

The following technical lemmas are proved in Faÿ *et al.* [4].

Lemma A.1. *Let Assumption* 1 *hold. Let f be a bounded measurable compactly supported function such that $\int f(s)\,\mathrm{d}s = 0$. Define*

$$\tilde{f}(t, v, w) = w \int_t^{t+v} f(s)\,\mathrm{d}s.$$

Then $\int |\tilde{f}(t, v, w)|^p \,\mathrm{d}t\, \nu(\mathrm{d}v, \mathrm{d}w) < \infty$, $\mathbb{E}[N_S(\tilde{f})] = 0$ *and* $\int_0^\infty X(s) f(s)\,\mathrm{d}s = N_S(\tilde{f} \mathbb{1}_{\mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}})$. *If, moreover,* $\mathbb{E}[\eta] < \infty$, *then* $N_S(\tilde{f}) = \int X_S(s) f(s)\,\mathrm{d}s$.

Lemma A.2. *Let Assumption* 1 *hold with $p^* \geq 4$. Then, there exists a positive constant $C > 0$ such that*

$$\operatorname{var}\left(\sum_{k=0}^{n-1} (d_{j,k}^S)^2\right) \leq Cn\left\{L_2^2(2^j) 2^{(4-2\alpha)j} + L_4(2^j) 2^{(3-\alpha)j}\right\}.$$

Note that the first term dominates for $\alpha < 1$ and the second dominates for $\alpha > 1$.

Lemma A.3. *Let Assumption* 1 *hold. Let f be a bounded measurable compactly supported function such that $\int f(s)\,\mathrm{d}s = 0$. Define*

$$\hat{f}(t, v, w) = w \int_{-\infty}^{\infty} g_{t,v}(s) f(s)\,\mathrm{d}s, \qquad \check{f}(t, v, w) = w \int_{-\infty}^{\infty} h_{t,v}(s) f(s)\,\mathrm{d}s.$$

Then, for $p = 1, \dots, p^*$, $\int |\hat{f}(t, v, w)|^p \,\mathrm{d}t\, \nu(\mathrm{d}v, \mathrm{d}w) < \infty$, $\int |\check{f}(t, v, w)|^p \,\mathrm{d}t\, \nu(\mathrm{d}v, \mathrm{d}w) < \infty$, $\int \mathrm{I}_\phi[X](s) f(s)\,\mathrm{d}s = N_S(\hat{f} \mathbb{1}_{\mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}})$, *and* $\mathbb{E}[N_S(\hat{f})] = \mathbb{E}[N_S(\check{f})] = 0$. *If, moreover,* $\mathbb{E}[\eta] < \infty$, *then* $N_S(\hat{f}) = \int \mathrm{I}_\phi[X_S](s) f(s)\,\mathrm{d}s$.

Applying Lemma A.3, we can extend the definition of $d_{j,k}^{SD}$ in (3.10) to the case $\mathbb{E}[\eta] = \infty$ by

$$d_{j,k}^{SD} = N_S(\hat{\psi}_{j,k}). \tag{A.1}$$

Lemma A.4. (i) *Let Assumption 1 hold with $p^* \geq 1$ and $\alpha \in (0, 2)$. Then $\mathbb{E}[d_{j,k}^{SD}] = 0$ for all $j \geq 0$ and $k \in \mathbb{Z}$.*

(ii) *Let Assumption 1 hold with $p^* \geq 2$ and $\alpha \in (0, 2)$. Then $\operatorname{var}(d_{j,k}^S - d_{j,k}^{SD})$ is bounded uniformly for $j \in \mathbb{N}$ and $k \in \mathbb{Z}$.*

(iii) *Let Assumption 1 hold with $p^* \geq 2$ and $\alpha \in (1, 2)$. Then $|\operatorname{var}(d_{j,k}^S) - \operatorname{var}(d_{j,k}^{SD})|$ is bounded uniformly for $j \in \mathbb{N}$ and $k \in \mathbb{Z}$.*

Lemma A.5. *Let Assumption 1 hold with $\alpha \in (1, 2)$ and $p^* \geq 2$. Then*

$$\sup_{0 \leq j \leq J} \mathbb{E}|\Lambda_j| = O(1); \tag{A.2}$$

$$\sup_{n \geq 1, j \geq 0} n^{-1/2} \mathbb{E}\left[\left| v_j^{-1} \sum_{k=0}^{n-1} \{(d_{j,k}^S)^2 - d_{j,k}^2\} \right|\right] < \infty. \tag{A.3}$$

Lemma A.6. *Let Assumption 1 hold with $\alpha \in (0, 2)$ and $p^* \geq 4$. If $\alpha \leq 1/2$, assume (4.4) and (4.5).*

Let J^* be a sequence depending on J such that $\limsup J^*/J < (1/\alpha) \wedge (1/(2 - \alpha))$. Then, there exists $\epsilon > 0$ such that

$$\sup_{u \in \mathcal{S}} \left| \sum_{j=J_0+1}^{J^*} u_j \Lambda_j \right| = O_P(2^{-\epsilon J}), \tag{A.4}$$

where $\mathcal{S}$ is the set of sequences $u = (u_0, \ldots)$ satisfying $\sum_{j \in \mathbb{N}} |u_j| \leq 1$.

Lemma A.7. *Let Assumption 1 hold with $p^* \geq 4$ and $\alpha \in (1, 2)$. Then, there exists a positive constant $C > 0$ such that*

$$\operatorname{var}\left(\sum_{k=0}^{n-1} (d_{j,k}^{SD})^2 \right) \leq C L_4(2^j) n 2^{(3-\alpha)j}. \tag{A.5}$$

Lemma A.8. *Let ρ be a positive real and $\rho' := (2^\rho - 1)^{-1}$. Let ℓ^* be a non-increasing function on $[1, \infty)$ such that $\lim_{s \to \infty} \ell^*(s) = 0$, and let ℓ be a function on $[1, \infty)$ such that $|\ell(s)| \leq \ell^*(s)$ for all $s \in [1, \infty)$. Define*

$$L(x) = c \exp\left\{ \int_1^x \frac{\ell(s)}{s} \, \mathrm{d}s \right\} \quad \text{and} \quad \omega_j = \frac{2^{-\rho j} L(2^j)}{\sum_{j'=J_0+1}^{J_1} 2^{-\rho j'} L(2^{j'})}.$$

Then, as $J_0 \to \infty$ *and for any* $\epsilon > 0$,

$$\sum_{j=J_0+1}^{J_1} \omega_j j = J_0 + 1 + \rho'\big(1 + O(\ell^*(2^{J_0}))\big) + O\big(J_1(2-\epsilon)^{J_0-J_1}\big), \tag{A.6}$$

$$\sum_{j=J_0+1}^{J_1} \omega_j j^2 = J_0^2 + 2J_0(1+\rho') + 2\rho'^2 + 3\rho' + 1 + \rho' O(\ell^*(2^{J_0})) + O\big(J_1^2(2-\epsilon)^{J_0-J_1}\big). \tag{A.7}$$

References

[1] Barakat, C., Thiran, P., Iannaccone, G., Diot, C. and Owezarski, P. (2002). A flow-based model for internet backbone traffic. In *Proceedings of the 2nd ACM SIGCOMM Workshop on Internet Measurement*, pp. 35–47. New York: ACM Press.

[2] Cohen, A. (2003). *Numerical Analysis of Wavelet Methods*. Amsterdam: North-Holland. MR1990555

[3] Duffield, N.G., Lund, C. and Thorup, M. (2002). Properties and prediction of flow statistics from sampled packet streams. In *Proceedings of the 2nd ACM SIGCOMM Workshop on Internet Measurement*, pp. 159–171. New York, ACM Press.

[4] Faÿ, G., Roueff, F. and Soulier, P. (2005). Estimation of the memory parameter of the infinite source Poisson process. arXiv:math.ST/0509371v2.

[5] Hall, P. and Welsh, A.H. (1984). Best attainable rates of convergence for estimates of regular variation. *Ann. Statist.* **3** 1079–1084. MR0751294

[6] Künsch, H.R. (1987). Statistical aspects of self-similar processes. In Yu.A. Prohorov and V.V. Sazonov (eds), *Proceedings of the First World Congres of the Bernoulli Society*, Vol. 1, pp. 67–74. Utrecht: VNU Science Press. MR1092336

[7] Maulik, K., Resnick, S. and Rootzén, H. (2002). Asymptotic independence and a network traffic model. *J. Appl. Probab.* **39** 671–699. MR1938164

[8] Meyer, Y. (1992). *Wavelets and Operators*. Cambridge: Cambridge University Press. MR1228209

[9] Mikosch, T., Resnick, S., Rootzén, H. and Stegeman, A. (2002). Is network traffic approximated by stable Lévy motion or fractional Brownian motion? *Ann. Appl. Probab.* **12** 23–68. MR1890056

[10] Parulekar, M. and Makowski, A.M. (1997). M/G/infinity input processes: A versatile class of models for network traffic. In *Proceedings of INFOCOM '97*, pp. 419–426. Los Alamitos, CA: IEEE Computer Society Press.

[11] Resnick, S. (1987). *Extreme Values, Regular Variation and Point Processes*. New York: Springer-Verlag. MR0900810

[12] Resnick, S. and Rootzén, H. (2000). Self-similar communication models and very heavy tails. *Ann. Appl. Probab.* **10** 753–778. MR1789979

[13] Wornell, G.A. and Oppenheim, A.V. (1992). Wavelet-based representations for a class of self-similar signals with application to fractal modulation. *IEEE Trans. Inform. Theory* **38** 785–800.

Received September 2005 and revised November 2006