
Long Memory in Nonlinear Processes

Rohit Deo¹, Mengchen Hsieh¹, Clifford M. Hurvich¹, and Philippe Soulier²

¹ New York University, 44 W. 4th Street, New York NY 10012, USA
 {rdeo,mhsieh,churvich}@stern.nyu.edu

² Université Paris X, 200 avenue de la République, 92001 Nanterre cedex, France
 philippe.soulier@u-paris10.fr

1 Introduction

It is generally accepted that many time series of practical interest exhibit strong dependence, i.e., long memory. For such series, the sample autocorrelations decay slowly and log-log periodogram plots indicate a straight-line relationship. This necessitates a class of models for describing such behavior. A popular class of such models is the autoregressive fractionally integrated moving average (ARFIMA) (see [Ade74], [GJ80]), [Hos81], which is a linear process. However, there is also a need for nonlinear long memory models. For example, series of returns on financial assets typically tend to show zero correlation, whereas their squares or absolute values exhibit long memory. See, e.g., [DGE93]. Furthermore, the search for a realistic mechanism for generating long memory has led to the development of other nonlinear long memory models. (Shot noise, special cases of which are Parke, Taqqu-Levy, etc). In this chapter, we will present several nonlinear long memory models, and discuss the properties of the models, as well as associated parametric and semiparametric estimators.

Long memory has no universally accepted definition; nevertheless, the most commonly accepted definition of long memory for a weakly stationary process $X = \{X_t, t \in \mathbb{Z}\}$ is the regular variation of the autocovariance function: there exist $H \in (1/2, 1)$ and a slowly varying function L such that

$$\text{cov}(X_0, X_t) = L(t)|t|^{2H-2}. \quad (1)$$

Under this condition, it holds that:

$$\lim_{n \rightarrow \infty} n^{-2H} L(n)^{-1} \text{var} \left(\sum_{t=1}^n X_t \right) = 1/(2H(2H-1)). \quad (2)$$

The condition (2) does not imply (1). Nevertheless, we will take (2) as an alternate definition of long memory. In both cases, the index H will be referred

to as the *Hurst index* of the process X . This definition can be expressed in terms of the parameter $d = H - 1/2$, which we will refer to as the *memory parameter*. The most famous long memory processes are fractional Gaussian noise and the *ARFIMA*(p, d, q) process, whose memory parameter is d and Hurst index is $H = 1/2 + d$. See for instance [Taq03] for a definition of these processes.

The second-order properties of a stationary process are not sufficient to characterize it, unless it is a Gaussian process. Processes which are linear with respect to an i.i.d. sequence (strict sense linear processes) are also relatively well characterized by their second-order structure. In particular, weak convergence of the partial sum process of a Gaussian or strict sense linear long memory processes $\{X_t\}$ with Hurst index H can be easily derived. Define $S_n(t) = \sum_{k=1}^{\lfloor nt \rfloor} (X_k - \mathbb{E}[X_k])$ in discrete time or $S_n(t) = \int_0^{nt} (X_s - \mathbb{E}[X_s]) ds$ in continuous time. Then $\text{var}(S_n(1))^{-1/2} S_n(t)$ converges in distribution to a constant times the fractional Brownian motion with Hurst index H , that is the Gaussian process B_H with covariance function

$$\text{cov}(B_H(s), B_H(t)) = \frac{1}{2} \{ |s|^{2H} - |t-s|^{2H} + t^{2H} \} .$$

In this paper, we will introduce nonlinear long memory processes, whose second order structure is similar to that of Gaussian or linear processes, but which may differ greatly from these processes in many other aspects. In Section 2, we will present these models and their second-order properties, and the weak convergence of their partial sum process. These models include conditionally heteroscedastic processes (Section 2.1) and models related to point processes (Section 2.2). In Section 3, we will consider the problem of estimating the Hurst index or memory parameter of these processes.

2 Models

2.1 Conditionally heteroscedastic models

These models are defined by

$$X_t = \sigma_t v_t , \tag{3}$$

where $\{v_t\}$ is an independent identically distributed series with finite variance and σ_t^2 is the so-called volatility. We now give examples.

LMSV and LMSD

The Long Memory Stochastic Volatility (LMSV) and Long Memory Stochastic Duration (LMSD) models are defined by Equation (3), where $\sigma_t^2 = \exp(h_t)$ and $\{h_t\}$ is an unobservable Gaussian long memory process with memory parameter $d \in (0, 1/2)$, independent of $\{v_t\}$. The multiplicative innovation series

$\{v_t\}$ is assumed to have zero mean in the LMSV model, and positive support with unit mean in the LMSD model. The LMSV model was first introduced by [BCdL98] and [Har98] to describe returns on financial assets, while the LMSD model was proposed by [DHH05] to describe durations between transactions on stocks.

Using the moment generating function of a Gaussian distribution, it can be shown (see [Har98]) for the LMSV/LMSD model that for any real s such that $\mathbb{E}[|v_t|^s] < \infty$,

$$\rho_s(j) \sim C_s j^{2d-1} \quad j \rightarrow \infty,$$

where $\rho_s(j)$ denotes the autocorrelation of $\{|x_t|^s\}$ at lag j , with the convention that $s = 0$ corresponds to the logarithmic transformation. As shown in [SV02], the same result holds under more general conditions without the requirement that $\{h_t\}$ be Gaussian.

In the LMSV model, assuming that $\{h_t\}$ and $\{v_t\}$ are functions of a multivariate Gaussian process, [Rob01] obtained similar results on the autocorrelations of $\{|X_t|^s\}$ with $s > 0$ even if $\{h_t\}$ is not independent of $\{v_t\}$. Similar results were obtained in [SV02], allowing for dependence between $\{h_t\}$ and $\{v_t\}$.

The LMSV process is an uncorrelated sequence, but powers of LMSV or LMSD may exhibit long memory. [SV02] proved the convergence of the centered and renormalized partial sums of any absolute power of these processes to fractional Brownian motion with Hurst index $1/2$ in the case where they have short memory.

FIEGARCH

The weakly stationary FIEGARCH model was proposed by [BM96]. The FIEGARCH model, which is observation-driven, is a long-memory extension of the EGARCH (exponential GARCH) model of [Nel91]. The FIEGARCH model for returns $\{X_t\}$ takes the form 2.1 innovation series $\{v_t\}$ are i.i.d. with zero mean and a symmetric distribution, and

$$\log \sigma_t^2 = \omega + \sum_{j=1}^{\infty} a_j g(v_{t-j}) \tag{4}$$

with $g(x) = \theta x + \gamma(|x| - \mathbb{E}|v_t|)$, $\omega > 0$, $\theta \in \mathbb{R}$, $\gamma \in \mathbb{R}$, and real constants a_j such that the process $\{\log \sigma_t^2\}$ has long memory with memory parameter $d \in (0, 1/2)$. If θ is nonzero, the model allows for a so-called leverage effect, whereby the sign of the current return may have some bearing on the future volatility. In the original formulation of [BM96], the $\{a_j\}$ are the $AR(\infty)$ coefficients of an $ARFIMA(p, d, q)$ process.

As was the case for the LMSV model, here we can once again express the log squared returns as in (18) with $\mu = \mathbb{E}[\log v_t^2] + \omega$, $u_t = \log v_t^2 - \mathbb{E}[\log v_t^2]$, and $h_t = \log \sigma_t^2 - \omega$. Here, however, the processes $\{h_t\}$ and $\{u_t\}$ are not mutually independent. The results of [SV02] also apply here, and in particular, the processes $\{|X_t|^u\}$, $\{\log(X_t^2)\}$ and $\{\sigma_t\}$ have the same memory parameter d .

ARCH(∞) and FIGARCH

In ARCH(∞) models, the innovation series $\{v_t\}$ is assumed to have zero mean and unit variance, and the conditional variance is taken to be a weighted sum of present and past squared returns:

$$\sigma_t^2 = \omega + \sum_{k=1}^{\infty} a_k X_{t-k}^2, \quad (5)$$

where $\omega, a_j, j = 1, 2, \dots$ are nonnegative constants. The general framework leading to (3) and (5) was introduced by [Rob91]. [KL03] have shown that $\sum_{j=1}^{\infty} a_j \leq 1$ is a necessary condition for existence of a strictly stationary solution to equations (3), (5), while [GKL00] showed that $\sum_{j=1}^{\infty} a_j < 1$ is a sufficient condition for the existence of a strictly stationary solution. If $\sum_{j=1}^{\infty} a_j = 1$, the existence of a strictly stationary solution has been proved by [KL03] only in the case where the coefficients a_j decay exponentially fast. In any case, if a stationary solution exists, its variance, if finite, must be equal to $\omega(1 - \sum_{k=1}^{\infty} a_k)^{-1}$, so that it cannot be finite if $\sum_{k=1}^{\infty} a_k = 1$ and $\omega > 0$. If $\omega = 0$, then the process which is identically equal to zero is a solution, but it is not known whether a nontrivial solution exists.

In spite of a huge literature on the subject, the existence of a strictly or weakly stationary solution to (3), (5) such that $\{\sigma_t^2\}$, $\{|X_t|^u\}$ or $\{\log(X_t^2)\}$ has long memory is still an open question. If $\sum_{j=1}^{\infty} a_j < 1$, and the coefficients a_j decay sufficiently slowly, [GKL00] found that it is possible in such a model to get hyperbolic decay in the autocorrelations $\{\rho_r\}$ of the squares, though the rates of decay they were able to obtain were proportional to $r^{-\theta}$ with $\theta > 1$. Such autocorrelations are summable, unlike the autocorrelations of a long-memory process with positive memory parameter. For instance, if the weights $\{a_j\}$ are proportional to those given by the AR(∞) representation of an ARFIMA(p, d, q) model, then $\theta = -1 - d$. If $\sum_{j=1}^{\infty} a_j = 1$, then the process has infinite variance so long memory as defined here is irrelevant.

Let us mention for historical interest the FIGARCH (fractionally integrated GARCH) model which appeared first in [BBM96]. In the FIGARCH model, the weights $\{a_j\}$ are given by the AR(∞) representation of an ARFIMA(p, d, q) model, with $d \in (0, 1/2)$, which implies that $\sum_{j=1}^{\infty} a_j = 1$, hence the very existence of FIGARCH series is an open question, and in any case, if it exists, it cannot be weakly stationary. The lack of weak stationarity of the FIGARCH model was pointed out by [BBM96]. Once again, at the time of writing this paper, we are not aware of any rigorous result on this process or on any ARCH(∞) process with long memory.

LARCH

Since the ARCH structure (apparently) fails to produce long memory, an alternative definition of heteroskedasticity has been considered in which long memory can be proved rigorously. [GS02] considered models which satisfy the

equation $X_t = \zeta_t A_t + B_t$, where $\{\zeta_t\}$ is a sequence of i.i.d. centered random variables with unit variance and A_t and B_t are linear in $\{X_t\}$ instead of quadratic as in the ARCH specification. This model nests the LARCH model introduced by [Rob91], obtained for $B_t \equiv 0$. The advantage of this model is that it can exhibit long memory in the conditional mean B_t and/or in the conditional variance A_t , possibly with different memory parameters. See [GS02, Corollary 4.4]. The process $\{X_t\}$ also exhibits long memory with a memory parameter depending on the memory parameters of the mean and the conditional variance [GS02, Theorem 5.4]. If the conditional mean exhibits long memory, then the partial sum process converges to the fractional Brownian motion, and it converges to the standard Brownian motion otherwise. See [GS02, Theorem 6.2]. The squares $\{X_t^2\}$ may also exhibit long memory, and their partial sum process converge either to the fractional Brownian motion or to a non Gaussian self-similar process. This family of processes is thus very flexible. An extension to the multivariate case is given in [DTW05].

We conclude this section by the following remark. Even though these processes are very different from Gaussian or linear processes, they share with weakly dependent processes the Gaussian limit and the fact that weak limits and L^2 limits have consistent normalisations, in the sense that, if ξ_n denotes one of the usual statistics computed on a time series, there exists a sequence v_n such that $v_n \xi_n$ converges weakly to a non degenerate distribution and $v_n^2 \mathbb{E}[\xi_n^2]$ converges to a positive limit (which is the variance of the asymptotic distribution). In the next subsection, we introduce models for which this is no longer true.

2.2 Shot noise processes

General forms of the shot-noise process have been considered for a long time; see for instance [Tak54], [Dal71]. Long memory shot noise processes have been introduced more recently; an early reference seems to be [GMS93]. We present some examples of processes related to shot noise which may exhibit long memory. For simplicity and brevity, we consider only stationary processes.

Let $\{t_j, j \in \mathbb{Z}\}$ be the points of a stationary point process on the line, numbered for instance in such a way that $t_{-1} < 0 \leq t_0$, and for $t \geq 0$, let $N(t) = \sum_{j \geq 0} \mathbb{1}_{\{t_j \leq t\}}$ be the number of points between time zero and t . Define then

$$X_t = \sum_{j \in \mathbb{Z}} \epsilon_j \mathbb{1}_{\{t_j \leq t < t_j + \eta_j\}}, \quad t \geq 0. \tag{6}$$

In this model, the shocks $\{\epsilon_j\}$ are an i.i.d. sequence; they are generated at birth times $\{t_j\}$ and have durations $\{\eta_j\}$. The observation at time t is the sum of all surviving present and past shocks. In model (6), we can take time to be continuous, $t \in \mathbb{R}$ or discrete, $t \in \mathbb{Z}$. This will be made precise later for

each model considered. We now describe several well known special cases of model (6).

1. Renewal-reward process; [TL86], [Liu00].

The durations are exactly the interarrival times of the renewal process: $\eta_0 = t_0$, $\eta_j = t_{j+1} - t_j$, and the shocks are independent of their birth times. Then there is exactly one surviving shock at time t :

$$X_t = \epsilon_{N(t)}. \quad (7)$$

2. ON-OFF model; [TWS97].

This process consists of alternating ON and OFF periods with independent durations. Let $\{\eta_k\}_{k \geq 1}$ and $\{\zeta_k\}_{k \geq 1}$ be two independent i.i.d. sequences of positive random variables with finite mean. Let t_0 be independent of these sequences and define $t_j = t_0 + \sum_{k=1}^j (\eta_k + \zeta_k)$. The shocks ϵ_j are deterministic and equal to 1. Their duration is η_j . The η_j s are the ON periods and the ζ_j s are the OFF periods. The first interval t_0 can also be split into two successive ON and OFF periods η_0 and ζ_0 . The process X can be expressed as

$$X_t = \mathbb{1}_{\{t_{N(t)} \leq t < t_{N(t)} + \eta_{N(t)}\}}. \quad (8)$$

3. Error duration process; [Par99].

This process was introduced to model some macroeconomic data. The birth times are deterministic, namely $t_j = j$, the durations $\{\eta_j\}$ are i.i.d. with finite mean and

$$X_t = \sum_{j \leq t} \epsilon_j \mathbb{1}_{\{t < j + \eta_j\}}. \quad (9)$$

4. Infinite Source Poisson model.

If the t_j are the points of a homogeneous Poisson process, the durations $\{\eta_j\}$ are i.i.d. with finite mean and $\epsilon_j \equiv 1$, we obtain the infinite source Poisson model or M/G/ ∞ input model considered among others in [MRRS02].

[MRR02] have considered a variant of this process where the shocks (referred to as transmission rates in this context) are random, and possibly contemporaneously dependent with durations.

In the first two models, the durations satisfy $\eta_j \leq t_{j+1} - t_j$, hence are not independent of the point process of arrivals (which is here a renewal process). Nevertheless η_j is independent of the past points $\{t_k, k \leq j\}$. The process can be defined for all $t \geq 0$ without considering negative birth times and shocks. In the last two models, the shocks and durations are independent of the renewal process, and any past shock may contribute to the value of the process at time t .

Stationarity and second order properties

- The renewal-reward process (7) is strictly stationary since the renewal process is stationary and the shocks are i.i.d. It is moreover weakly stationary if the shocks have finite variance. Then $\mathbb{E}[X_t] = \mathbb{E}[\epsilon_1]$ and

$$\text{cov}(X_0, X_t) = \mathbb{E}[\epsilon^2] \mathbb{P}(\eta_0 > t) = \lambda \mathbb{E}[\epsilon_1^2] \mathbb{E}[(\eta_1 - t)_+] , \quad (10)$$

where η_0 is the delay distribution and $\lambda = \mathbb{E}[(t_1 - t_0)]^{-1}$ is intensity of the stationary renewal process. Note that this relation would be true for a general stationary point process. Cf. for instance [TL86] or [HHS04].

- The stationary version of the ON-OFF was studied in [HRS98]. The first On and OFF period η_0 and ζ_0 can be defined in such a way that the process X is stationary. Let F_{on} and F_{off} be the distribution functions of the ON and OFF periods η_1 and ζ_1 . [HRS98, Theorem 4.3] show that if $1 - F_{\text{on}}$ is regularly varying with index $\alpha \in (1, 2)$ and $1 - F_{\text{off}}(t) = o(F_{\text{on}}(t))$ as $t \rightarrow \infty$, then

$$\text{cov}(X_0, X_t) \sim c \mathbb{P}(\eta_0 > t) = c \lambda \mathbb{E}[(\eta_1 - t)_+] , \quad (11)$$

- Consider now the case when the durations are independent of the birth times. To be precise, assume that $\{(\eta_j, \epsilon_j)\}$ is an i.i.d. sequence of random vectors, independent of the stationary point process of points $\{t_j\}$. Then the process $\{X_t\}$ is strictly stationary as long as $\mathbb{E}[\eta_1] < \infty$, and has finite variance if $\mathbb{E}[\epsilon_1^2 \eta_1] < \infty$. Then $\mathbb{E}[X_t] = \lambda \mathbb{E}[\epsilon_1 \eta_1]$ and

$$\begin{aligned} \text{cov}(X_0, X_t) &= \lambda \mathbb{E}[\epsilon_1^2 (\eta_1 - t)_+] \\ &\quad + \{ \text{cov}(\epsilon_1 N(-\eta_1, 0], \epsilon_2 N(t - \eta_2, t]) - \lambda \mathbb{E}[\epsilon_1 \epsilon_2 (\eta_1 \wedge (\eta_2 - t)_+)] \} , \end{aligned}$$

where λ is the intensity of the stationary point process, i.e. $\lambda^{-1} = \mathbb{E}[t_0]$. The last term has no known general expression for a general point process, but it vanishes in two particular cases:

- if N is a homogeneous Poisson point process;
- if ϵ_1 is centered and independent of η_1 .

In the latter case (10) holds, and in the former case, we obtain a formula which generalizes (10):

$$\text{cov}(X_0, X_t) = \lambda \mathbb{E}[\epsilon_1^2 (\eta_1 - t)_+] . \quad (12)$$

We now see that second order long memory can be obtained if (10) holds and the durations have regularly varying tails with index $\alpha \in (1, 2)$ or,

$$\mathbb{E}[\epsilon_1^2 \mathbb{1}_{\{\eta_1 > t\}}] = \ell(t) t^{-\alpha} . \quad (13)$$

Thus, if (13) and either (11) or (12) hold, then X has long memory with Hurst index $H = (3 - \alpha)/2$ since

$$\text{cov}(X_0, X_t) \sim \frac{\lambda}{\alpha - 1} \ell(t)t^{1-\alpha}. \quad (14)$$

Examples of interest in teletraffic modeling where ϵ_1 and η_1 are not independent but (13) holds are provided in [MRR02] and [FRS05].

We conjecture that (14) holds in a more general framework, at least if the interarrival times of the point process have finite variance.

Weak convergence of partial sums

This class of long memory process exhibits a very distinguishing feature. Instead of converging weakly to a process with finite variance, dependent stationary increments such as the fractional Brownian motion, the partial sums of some of these processes have been shown to converge to an α -stable Levy process, that is, an α -stable process with independent and stationary increment. Here again there is no general result, but such a convergence is easy to prove under restrictive assumptions. Define

$$S_T(t) = \int_0^{Tt} \{X_s - \mathbb{E}[X_s]\} ds.$$

Then it is known in the particular cases described above that the finite dimensional distributions of the process $\ell(T)T^{-1/\alpha}S_T$ (for some slowly varying function ℓ) converge weakly to those of an α -stable process. This was proved in [TL86] for the renewal reward process, in [MRRS02] for the ON-OFF and infinite source Poisson processes when the shocks are constant. A particular case of dependent shocks and durations is considered in [MRR02]. [HHS04] proved the result in discrete time for the error duration process; the adaptation to the continuous time framework is straightforward. It is also probable that such a convergence holds when the underlying point process is more general.

Thus, these processes are examples of second order long memory process with Hurst index $H \in (1/2, 1)$ such that $T^{-H}S_T(t)$ converges in probability to zero. This behaviour is very surprising and might be problematic in statistical applications, as illustrated in Section 3.

It must also be noted that convergence does not hold in the space \mathcal{D} of right-continuous, left-limited functions endowed with the J_1 topology, since a sequence of processes with continuous path which converge in distribution in this sense must converge to a process with continuous paths. It was proved in [RvdB00, Theorem 4.1] that this convergence holds in the M_1 topology for the infinite source Poisson process. For a definition and application of the M_1 topology in queuing theory, see [Whi02].

Slow growth and fast growth

Another striking feature of these processes is the slow growth versus fast growth phenomenon, first noticed by [TL86] for the renewal-reward process and more rigorously investigated by [MRRS02] for the ON-OFF and infinite

source Poisson process³. Consider M independent copies $X^{(i)}$, $1, \leq i \leq M$ of these processes and denote

$$A_{M,T}(t) = \sum_{i=1}^M \int_0^{Tt} \{X_s^{(i)} - \mathbb{E}[X_s]\} ds .$$

If M depends on T , then, according to the rate growth of M with respect to T , a stable or Gaussian limit can be obtained. More precisely, the slow growth and fast growth conditions are, up to slowly varying functions $MT^{1-\alpha} \rightarrow 0$ and $MT^{1-\alpha} \rightarrow \infty$, respectively. In other terms, the slow and fast growth conditions are characterized by $\text{var}(A_{M,T}(1)) \ll b(MT)$ and $\text{var}(A_{M,T}(1)) \gg b(MT)$, respectively, where b is the inverse of the quantile function of the durations.

Under the slow growth condition, the finite dimensional distributions of $L(MT)(MT)^{-1/\alpha}A_{M,T}$ converge to those of a Levy α -stable process, where L is a slowly varying function. Under the fast growth condition, the sequence of processes $T^{-H}\ell^{-1/2}(T)M^{-1/2}A_{M,T}$ converges, in the space $\mathcal{D}(\mathbb{R}_+)$ endowed with the J_1 topology, to the fractional Brownian motion with Hurst index $H = (3 - \alpha)/2$. It is thus seen that under the fast growth condition, the behaviour of a Gaussian long memory process with Hurst index H is recovered.

Non stationary versions

If the sum defining the process X in (6) is limited to non negative indices j , then the sum has always a finite number of terms and there is no restriction on the distribution of the interarrival times $t_{j+1} - t_j$ and the durations η_j . These models can then be nonstationary in two ways: either because of initialisation, in which case a suitable choice of the initial distribution can make the process stationary; or because these processes are non stable and have no stationary distribution. The latter case arises when the interarrival times and/or the durations have infinite mean. These models were studied by [RR00] and [MR04] in the case where the point process of arrivals is a renewal process. contrary to the stationry case, where heavy tailed durations imply non Gaussian limits, the limiting process of the partial sums has non stationary increments and can be Gaussian in some cases.

2.3 Long Memory in Counts

The time series of counts of the number of transactions in a given fixed interval of time is of interest in financial econometrics. Empirical work suggests that such series may possess long memory. See [DHH05]. Since the counts are

³ Actually, in the case of the Infinite Source Poisson process, [MRRS02] consider a single process but with an increasing rate λ depending on T , rather than superposition of independent copies. The results obtained are nevertheless of the same nature.

induced by the durations between transactions, it is of interest to study the properties of durations, how these properties generate long memory in counts, and whether there is a connection between potential long memory in durations and long memory in counts.

The event times determine a counting process $N(t) = \text{Number of events in } (0, t]$. Given any fixed clock-time spacing $\Delta t > 0$, we can form the time series $\{\Delta N_{t'}\} = \{N(t'\Delta t) - N[(t' - 1)\Delta t]\}$ for $t' = 1, 2, \dots$, which counts the number of events in the corresponding clock-time intervals of width Δt . We will refer to the $\{\Delta N_{t'}\}$ as the *counts*. Let $\tau_k > 0$ denote the waiting time (duration) between the $k - 1$ 'st and the k 'th transaction.

We give some preliminary definitions taken from [DVJ03].

Definition 1. *A point process $N(t) = N(0, t]$ is stationary if for every $r = 1, 2, \dots$ and all bounded Borel sets A_1, \dots, A_r , the joint distribution of $\{N(A_1 + t), \dots, N(A_r + t)\}$ does not depend on $t \in [0, \infty)$.*

A second order stationary point process is long-range count dependent (LRcD) if

$$\lim_{t \rightarrow \infty} \frac{\text{var}(N(t))}{t} = \infty .$$

A second order stationary point process $N(t)$ which is LRcD has Hurst index $H \in (1/2, 1)$ given by

$$H = \sup\{h : \limsup_{t \rightarrow \infty} \frac{\text{var}(N(t))}{t^{2h}} = \infty\} .$$

Thus if the counts $\{\Delta N_{t'}\}_{t'=-\infty}^{\infty}$ on intervals of any fixed width $\Delta t > 0$ are LRD with memory parameter d then the counting process $N(t)$ must be LRcD with Hurst index $H = d + 1/2$. Conversely, if $N(t)$ is an LRcD process with Hurst index H , then $\{\Delta N_{t'}\}$ cannot have exponentially decaying autocorrelations, and under the additional assumption of a power law decay of these autocorrelations, $\{\Delta N_{t'}\}$ is LRD with memory parameter $d = H - 1/2$.

There exists a probability measure P^0 under which the doubly infinite sequence of durations $\{\tau_k\}_{k=-\infty}^{\infty}$ are a stationary time series, i.e., the joint distribution of any subcollection of the $\{\tau_k\}$ depends only on the lags between the entries. On the other hand, the point process N on the real line is stationary under the measure P . A fundamental fact about point processes is that in general (a notable exception is the Poisson process) there is no single measure under which both the point process N and the durations $\{\tau_k\}$ are stationary, i.e., in general P and P^0 are not the same. Nevertheless, there is a one-to-one correspondence between the class of measures P^0 that determine a stationary duration sequence and the class of measures P that determine a stationary point process. The measure P^0 corresponding to P is called the *Palm distribution*. The counts are stationary under P , while the durations are stationary under P^0 .

We now present an important theoretical result obtained by [Dal99].

Theorem 1. *A stationary **renewal** point process is LRcD and has Hurst index $H = (1/2)(3 - \alpha)$ under P if the interarrival time has tail index $1 < \alpha < 2$ under P^0 .*

Theorem 1 establishes a connection between the tail index of a duration process and the persistence of the counting process. According to the theorem, the counting process will be LRcD if the duration process is *iid* with infinite variance. Here, the memory parameter of the counts is completely determined by the tail index of the durations.

This prompts the question as to whether long memory in the counts can be generated solely by dependence in finite-variance durations. An answer in the affirmative was given by [DRV00], who provide an example outside of the framework of the popular econometric models. We now present a theorem on the long-memory properties of counts generated by durations following the LMSD model. The theorem is a special case of a result proved in [DHSW05], who give sufficient conditions on durations to imply long memory in counts.

Theorem 2. *If the durations $\{\tau_k\}$ are generated by the LMSD process with memory parameter d , then the induced counting process $N(t)$ has Hurst index $H = 1/2 + d$, i.e. satisfies $\text{var}(N(t)) \sim Ct^{2d+1}$ under P as $t \rightarrow \infty$ where $C > 0$.*

3 Estimation of the Hurst index or memory parameter

A weakly stationary process with autocovariance function satisfying (1) has a spectral density f defined by

$$f(x) = \frac{1}{2\pi} \sum_{t \in \mathbb{Z}} \gamma(t) e^{itx} . \tag{15}$$

This series converges uniformly on the compact subsets of $[-\pi, \pi] \setminus \{0\}$ and in $L^1([-\pi, \pi], dx)$. Under some strengthening of condition (1), the behaviour of the function f at zero is related to the rate of decay of γ . For instance, if we assume in addition that L is ultimately monotone, we obtain the following Tauberian result [Taq03, Proposition 4.1], with $d = H - 1/2$.

$$\lim_{x \rightarrow 0} L(x)^{-1} x^{2d} f(x) = \pi^{-1} \Gamma(2d) \cos(\pi d) . \tag{16}$$

Thus, a natural idea is to estimate the spectral density in order to estimate the memory parameter d . The statistical tools are the discrete Fourier transform (DFT) and the periodogram, defined for a sample U_1, \dots, U_n , as

$$J_{n,j}^U = (2\pi n)^{-1/2} \sum_{t=1}^n U_t e^{itw_j}, \quad I_U(\omega_j) = |J_{n,j}^U|^2,$$

where $\omega_j = 2j\pi/n$, $1 \leq j < n/2$ are the so-called Fourier frequencies. (Note that for clarity the index n is omitted from the notation). In the classical weakly stationary short memory case (when the autocovariance function is absolutely summable), it is well known that the periodogram is an asymptotically unbiased estimator of the spectral density f_U defined in (15). This is no longer true for second order long memory processes. [HB93] showed (in the case where the function L is continuous at zero but the extension is straightforward) that for any fixed positive integer j , there exists a positive constant $c(k, H)$ such that

$$\lim_{n \rightarrow \infty} \mathbb{E}[I_U(\omega_j)/f_U(\omega_j)] = c(j, H).$$

The previous results are true for any second order long memory process. Nevertheless, spectral method of estimation of the Hurst parameter, based on the heuristic (but incorrect) assumption that the renormalised DFTs $f_U^{-1/2}(\omega_j)J_{n,j}^U$ are i.i.d. standard complex Gaussian have been proposed and theoretically justified in some cases. The most well known is the GPH estimator of the Hurst index, introduced by [GPH83] and proved consistent and asymptotically Gaussian for Gaussian long memory processes by [Rob95b] and for a restricted class of linear processes by [Vel00]. Another estimator, often referred to as the local Whittle or GSE estimator was introduced by [Kün87] and again proved consistent asymptotically Gaussian by [Rob95a] for linear long memory processes.

These estimators are built on the m first log-periodogram ordinates, where m is an intermediate sequence, i.e. $1/m + m/n \rightarrow 0$ as $n \rightarrow \infty$. The choice of m is irrelevant to consistency of the estimator but has an influence on the bias. The rate of convergence of these estimators, when known, is typically slower than \sqrt{n} . Trimming of the lowest frequencies, which means taking the ℓ first frequencies out is sometimes used, but there is no theoretical need for this practice, at least in the Gaussian case. See [HDB98]. For nonlinear series, we are not sure yet if trimming may be needed in general.

In the following subsections, we review what is known, both theoretically and empirically, about these and related methods for the different types of nonlinear processes described previously.

We start by describing the behaviour of the renormalized DFTs at low frequencies, that is, when the index j of the frequency ω_j remains fixed as $n \rightarrow \infty$.

3.1 Low-Frequency DFTs of Counts from Infinite-Variance Durations

To the best of our knowledge there is no model in the literature for long memory processes of counts. Hence the question of parametric estimation has not arisen so far in this context. However, one may still be interested in semiparametric estimation of long memory in counts. We present the following result on the behavior of the Discrete Fourier Transforms (DFTs) of processes

of counts induced by infinite-variance durations that will be of relevance to us in understanding the behavior of the GPH estimator. Let n denote the number of observations on the counts, $\omega_j = 2\pi j/n$, and define

$$J_{n,j}^{\Delta N} = \frac{1}{\sqrt{2\pi n}} \sum_{t'=1}^n \Delta N_{t'} e^{it'\omega_j} .$$

Assume that the distribution of the durations satisfies

$$P(\tau_k \geq x) \sim \ell(x)x^{-\alpha} \quad x \rightarrow \infty \tag{17}$$

where $\ell(x)$ is a slowly varying function with $\lim_{x \rightarrow \infty} \frac{\ell(kx)}{\ell(x)} = 1 \ \forall k > 0$ and $\ell(x)$ is ultimately monotone at ∞ .

Theorem 3. *Let $\{\tau_k\}$ be i.i.d. random variables which satisfy (17) with $\alpha \in (1, 2)$ and mean μ_τ . Then for each fixed j , $\ell(n)^{-1}n^{1/2-1/\alpha}J_{n,j}^{\Delta N}$ converges in distribution to a complex α -stable distribution. Moreover, for each fixed j , $\omega_j^d J_{n,j}^{\Delta N} \xrightarrow{p} 0$, where $d = 1 - \alpha/2$.*

The theorem implies that when j is fixed, the normalized periodogram of the counts, $\omega_j^{2d} I_{\Delta N}(\omega_j)$ converges in probability to zero. The degeneracy of the limiting distribution of the normalized DFTs of the counts suggests that the inclusion of the very low frequencies may induce negative finite-sample bias in semiparametric estimators. In addition, the fact that the suitably normalized DFT has an asymptotic stable distribution could further degrade the finite-sample behavior of semiparametric estimators, more so perhaps for the Whittle-likelihood-based estimators than for the GPH estimator since the latter uses the logarithmic transformation.

By contrast, for linear long-memory processes, the normalized periodogram has a nondegenerate positive limiting distribution. See, for example, [TH94].

3.2 Low-Frequency DFTs of Counts from LMSD Durations

We now study the behavior of the low-frequency DFTs of counts generated from finite-variance LMSD durations.

Theorem 4. *Let the durations $\{\tau_k\}$ follow an LMSD model with memory parameter d . Then for each fixed j , $\omega_j^d J_{n,j}^{\Delta N}$, converges in distribution to a zero-mean Gaussian random variable.*

This result is identical to what would be obtained if the counts were a linear long-memory process, and stands in stark contrast to Theorem 3. The discrepancy between these two theorems suggests that the low frequencies will contribute far more bias to semiparametric estimates of d based on counts if the counts are generated by infinite-variance durations than if they were generated from LMSD durations.

3.3 Low and High Frequency DFTs of Shot-Noise Processes

Let X be either the renewal-reward process defined in (7) or the error duration process (9). [HHS04], Theorem 4.1, have proved that Theorem 3 still holds, i.e. $n^{1/2-1/\alpha} J_{n,j}^X$ converges in distribution to an α -stable law, where α is the tail index of the duration. This result can probably be extended to all the shot-noise process for which convergence in distribution of the partial sum process can be proved.

The DFTs of these processes have an interesting feature, related to the slow growth/fast growth phenomenon. The high frequency DFTs, i. e. the DFT $J_{n,j}^X$ computed at a frequency ω_j whose index j increases as n^ρ for some $\rho > 1 - 1/\alpha$, renormalized by the square root of the spectral density computed at ω_j , have a Gaussian weak limit. This is proved in Theorem 4.2 of [HHS04].

3.4 Estimation of the memory parameter of the LMSV and LMSD models

We now discuss parametric and semiparametric estimation of the memory parameter for the LMSV/LMSD models. Note that in both the LMSV and LMSD models, $\log x_t^2$ can be expressed as the sum of a long memory signal and *iid* noise. Specifically, we have

$$\log X_t^2 = \mu + h_t + u_t, \quad (18)$$

where $\mu = E(\log v_t^2)$ and $u_t = \log v_t^2 - E(\log v_t^2)$ is a zero-mean *iid* series independent of $\{h_t\}$. Since all the extant methodology for estimation for the LMSV model exploits only the above signal plus noise representation, the methodology continues to hold for the LMSD model.

Assuming that $\{h_t\}$ is Gaussian, [DH01] derived asymptotic theory for the log-periodogram regression estimator (GPH; [GPH83]) of d based on $\{\log X_t^2\}$. This provides some justification for the use of GPH for estimating long memory in volatility. Nevertheless, it can also be seen from Theorem 1 of [DH01] that the presence of the noise term $\{u_t\}$ induces a negative bias in the GPH estimator, which in turn limits the number m of Fourier frequencies which can be used in the estimator while still guaranteeing \sqrt{m} -consistency and asymptotic normality. This upper bound, $m = o[n^{4d/(4d+1)}]$, where n is the sample size, becomes increasingly stringent as d approaches zero. The results in [DH01] assume that $d > 0$ and hence rule out valid tests for the presence of long memory in $\{h_t\}$. Such a test based on the GPH estimator was provided and justified theoretically by [HS02].

[SP03] proposed a nonlinear log-periodogram regression estimator \hat{d}_{NLP} of d , using Fourier frequencies $1, \dots, m$. They partially account for the noise term $\{u_t\}$ through a first-order Taylor expansion about zero of the spectral density of the observations, $\{\log X_t^2\}$. They establish the asymptotic normality of $m^{1/2}(\hat{d}_{\text{NLP}} - d)$ under assumptions including $n^{-4d} m^{4d+1/2} \rightarrow \text{Const}$. Thus,

\hat{d}_{NLP} , with a variance of order $n^{-4d/(4d+1/2)}$, converges faster than the GPH estimator, but still arbitrarily slowly if d is sufficiently close to zero. [SP03] also assumed that the noise and signal are Gaussian. This rules out most LMSV/LMSD models, since $\{\log v_t^2\}$ is typically non-Gaussian.

For the LMSV/LMSD model, results analogous to those of [DH01] were obtained by [Art04] for the GSE estimator, based once again on $\{\log X_t^2\}$. The use of GSE instead of GPH allows the assumption that $\{h_t\}$ is Gaussian to be weakened to linearity in a Martingale difference sequence. [Art04] requires the same restriction on m as in [DH01]. A test for the presence of long memory in $\{h_t\}$ based on the GSE estimator was provided by [HMS05].

[HR03] proposed a local Whittle estimator of d , based on log squared returns in the LMSV model. The local Whittle estimator, which may be viewed as a generalized version of the GSE estimator, includes an additional term in the Whittle criterion function to account for the contribution of the noise term $\{u_t\}$ to the low frequency behavior of the spectral density of $\{\log X_t^2\}$. The estimator is obtained from numerical optimization of the criterion function. It was found in the simulation study of [HR03] that the local Whittle estimator can strongly outperform GPH, especially in terms of bias when m is large.

Asymptotic properties of the local Whittle estimator were obtained by [HMS05], who allowed $\{h_t\}$ to be a long-memory process, linear in a Martingale difference sequence, with potential nonzero correlation with $\{u_t\}$. Under suitable regularity conditions on the spectral density of $\{h_t\}$, [HMS05] established the \sqrt{m} -consistency and asymptotic normality of the local Whittle estimator, under certain conditions on m . If we assume that the short memory component of the spectral density of $\{h_t\}$ is sufficiently smooth, then their condition on m reduces to

$$\lim_{n \rightarrow \infty} (m^{-4d-1+\delta} n^{4d} + n^{-4} m^5 \log^2(m)) = 0 \quad (19)$$

for some arbitrarily small $\delta > 0$.

The first term in (19) imposes a lower bound on the allowable value of m , requiring that m tend to ∞ faster than $n^{4d/(4d+1)}$. It is interesting that [DH01], under similar smoothness assumptions, found that for $m^{1/2}(\hat{d}_{\text{GPH}} - d)$ to be asymptotically normal with mean zero, where \hat{d}_{GPH} is the GPH estimator, the bandwidth m must tend to ∞ at a rate *slower* than $n^{4d/(4d+1)}$. Thus for any given d , the optimal rate of convergence for the local Whittle estimator is faster than that for the GPH estimator.

Fully parametric estimation in LMSV/LMSD models once again is based on $\{\log X_t^2\}$ and exploits the signal plus noise representation (18). When $\{h_t\}$ and $\{u_t\}$ are independent, the spectral density of $\{\log X_t^2\}$ is simply the sum of the spectral densities of $\{h_t\}$ and $\{u_t\}$, viz.

$$f_{\log X^2}(\lambda) = f_h(\lambda) + \sigma_u^2/(2\pi), \quad (20)$$

where $f_{\log X^2}$ is the spectral density of $\{\log X_t^2\}$, f_h is the spectral density of $\{h_t\}$ and $\sigma_u^2 = \text{var}(u_t)$, all determined by the assumed parametric model. This

representation suggests the possibility of estimating the model parameters in the frequency domain using the Whittle likelihood. Indeed, [Hos97] claims that the resulting estimator is \sqrt{n} -consistent and asymptotically normal. We believe that though the result provided in [Hos97] is correct, the proof is flawed. [Deo95] has shown that the quasi-maximum likelihood estimator obtained by maximizing the Gaussian likelihood of $\{\log X_t^2\}$ in the time domain is \sqrt{n} -consistent and asymptotically normal.

One drawback of the latent-variable LMSV/LMSD models is that it is difficult to derive the optimal predictor of $|X_t|^s$. In the LMSV model, $\{|X_t|^s\}$ for $s > 0$ serves as a proxy for volatility, while in the LMSD model, $\{X_t\}$ represents durations. A computationally efficient algorithm for optimal linear prediction of such series was proposed in [DHL05], exploiting the Preconditioned Conjugate Gradient (PCG) algorithm. In [CHL05], it is shown that the computational cost of this algorithm is $O(n \log^{5/2} n)$, in contrast to the much more expensive Levinson algorithm, which has cost of $O(n^2)$.

3.5 Simulations on the GPH Estimator for Counts

We simulated i.i.d. durations from a positive stable distribution with tail index $\alpha = 1.5$, with an implied d for the counts of .25. We also simulated durations from an LMSD $(1, d, 0)$ model with Weibull innovations, $AR(1)$ parameter of $-.42$, and $d = .3545$, as was estimated from actual tick-by-tick durations in [DHH05]. The stable durations were multiplied by a constant $c = 1.21$ so that the mean duration matches that found in actual data. For the LMSD durations, we used $c = 1$. One unit in the rescaled durations is taken to represent one second. Tables 1 and 2, for the stable and LMSD cases respectively, present the GPH estimates based on the resulting counts for different values of Δt , using $n = 10,000$, $m = n^{0.5}$ and $m = n^{0.8}$. For the stable case, the bias was far more strongly negative for the smaller value of m , whereas for the LMSD case, the bias did not change dramatically with m . This is consistent with the discussion in Section 3.2, and also with the averaged log – log periodogram plots presented in Figure 1, where the averaging is taken over a large number of replications, and all positive Fourier frequencies are considered, $j = 1, \dots, n/2$. The plot for the stable durations (upper panel) shows a flat slope at the low frequencies. For this process, using more frequencies in the regression seems to mitigate the negative bias induced by the flatness in the lower frequencies as indicated by the less biased estimates of d when $m = n^{0.8}$.

For the LMSD process, if the conjecture is correct then the counts should have the same memory parameter as the durations, $d = .3545$. Assuming that this is the case, we did not find severe negative bias in the GPH estimators on the counts, though the estimate of d seems to increase with Δt in the case when $m = n^{0.5}$. The averaged log – log periodogram plot presented in the lower panel of Figure 1 shows a near-perfect straight line across all frequencies, which is quite different from the pattern we observed in the case of counts

based on stable durations. The straight-line relationship here is consistent with the bias results in our LMSD simulations, and with the discussion in Section 3.2.

Statistical properties of \hat{d}_{GPH} and the choice of m for Gaussian long-memory time series have been discussed in recent literature. [Rob95b] showed for Gaussian processes that the GPH estimator is $m^{1/2}$ -consistent and asymptotically normal if an increasing number of low frequencies L is trimmed from the regression of the log periodogram on log frequency. [HDB98] showed that trimming can be avoided for Gaussian processes. In our simulations, we did not use any trimming. There is as yet no theoretical justification for the GPH estimator in the current context since the counts are clearly non-Gaussian, and presumably constitute a nonlinear process. It is not clear whether trimming would be required for such a theory, but our simulations and theoretical results suggest that in some situations trimming may be helpful, while in others it may not be needed.

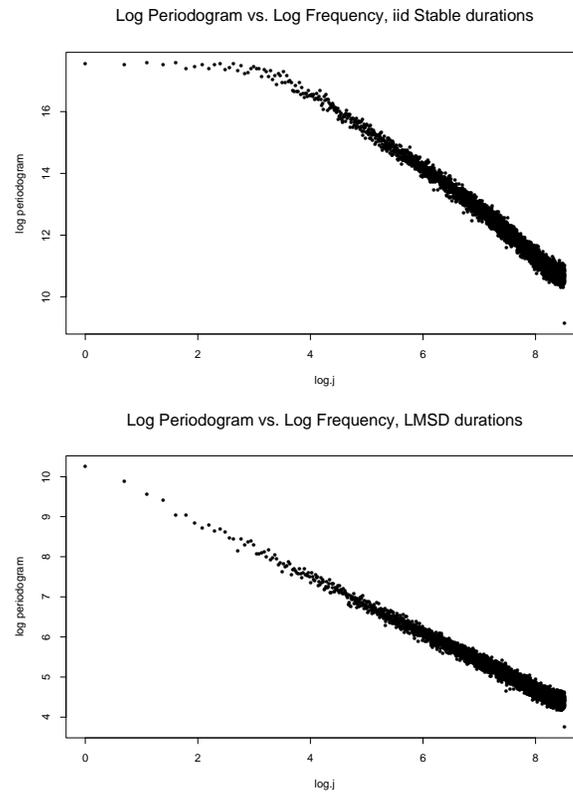
Table 1. GPH estimators for counts with different Δt . Counts generated from *iid* stable durations with skewness parameter $\beta = 0.8$ and tail index $\alpha = 1.5$. The corresponding memory parameter for counts is $d = .25$. We generated 500 replications each with sample size $n = 10,000$. The number of frequencies in the log periodogram regression was $m = n^{0.8} = 1585$ and $m = \sqrt{n} = 100$. t -values marked with * reject the null hypothesis, $d = 0.25$ in favor of $d < 0.25$.

| Δt | $m = n^{0.5}$ | | $m = n^{0.8}$ | |
|------------|-----------------------|---------|-----------------------|---------|
| $c = 1.21$ | $Mean(\hat{d}_{GPH})$ | t-Value | $Mean(\hat{d}_{GPH})$ | t-Value |
| 5 min | 0.1059 | -17.65* | 0.2328 | -5.77* |
| 10 min | 0.0744 | -23.08* | 0.2212 | -8.31* |
| 20 min | 0.0715 | -23.23* | 0.2186 | -7.75* |

Table 2. Mean of the GPH estimators for counts with different Δt . Counts generated from LMSD durations with Weibull $(1, \gamma)$ shocks. The number of frequencies in the log periodogram regression was $m = \sqrt{n}$ and $m = n^{0.8}$. We used $d = .3545$ and $\gamma = 1.3376$ for our simulations. We simulated 200 replications of the counts, each with sample size $n = 10,000$. t -values marked with * reject the null hypothesis, $d = 0.3545$ in favor of $d < 0.3545$.

| Δt | $m = n^{0.5}$ | | $m = n^{0.8}$ | |
|------------|-----------------------|---------|-----------------------|---------|
| $c = 1$ | $Mean(\hat{d}_{GPH})$ | t-Value | $Mean(\hat{d}_{GPH})$ | t-Value |
| 5 min | 0.3458 | -1.76* | 0.3471 | -6.49* |
| 30 min | 0.3873 | 3.45* | 0.3469 | -3.59* |
| 60 min | 0.3923 | 4.05* | 0.3478 | -3.20* |

Fig. 1. Averaged log – log periodogram plots for the counts generated from *iid* Stable and LMSD durations.



3.6 Estimation of the memory parameter of the Infinite Source Poisson process

Due to the underlying Poisson point process, the Infinite Poisson Source process is a very mathematically tractable model. Computations are very easy and in particular, convenient formulas for cumulants of integrals along paths of the process are available. This allows to derive the theoretical properties of estimators of the Hurst index or memory parameter. [FRS05] have defined an estimator of the Hurst index of the Infinite Poisson source process (with random transmission rate) related to the GSE and proved its consistency and rate of convergence. Instead of using the DFTs of the process, so-called wavelets coefficients are defined as follows. Let ψ be a measurable compactly supported function on \mathbb{R} such that $\int \psi(s) ds = 0$. For $j \in \mathbb{N}$ and $k = 0, \dots, 2^j - 1$, define

$$w_{j,k} = \int \phi(s) X_s ds .$$

If (13) holds, then $\mathbb{E}[w_{j,k}] = 0$ and $\text{var}(w_{j,k}) = L(2^j)2^{(2-\alpha)j} = L(2^j)2^{2dj}$, where α is the tail index of the durations, $d = 1 - \alpha/2$ is the memory parameter and L is a slowly varying function at infinity. This scaling property makes it natural to define a contrast function

$$\hat{W}(d') = \log \left(\sum_{(j,k) \in \Delta} 2^{-2d'j} w_{j,k}^2 \right) + \delta d' \log(2) ,$$

where Δ is the admissible set of coefficients, which depends on the interval of observation and the support of the function ψ . The estimator of d is then $\hat{d} = \arg \min_{d' \in (0, 1/2)} \hat{W}(d')$. [FRS05] have proved under some additional technical assumptions that this estimator is consistent. The rate of convergence can be obtained, but the asymptotic distribution is not known, though it is conjectured to be Gaussian, if the set Δ is properly chosen.

Note in passing that here again, the slow growth/fast growth phenomenon arises. It can be shown, if the shocks and durations are independent, that for fixed k , $2^{(1-\alpha)j/2} w_{j,k}$ converges to an α -stable distribution, but if k tends to infinity at a suitable rate, $2^{-dj} w_{j,k}$ converges to a complex Gaussian distribution. This slow growth/fast growth phenomenon is certainly a very deep property of these processes that should be understood more deeply.

References

- [Ade74] Rolf K. Adenstedt. On large-sample estimation for the mean of a stationary random sequence. *The Annals of Statistics*, 2:1095–1107, 1974.
- [Art04] Josu Arteche. Gaussian semiparametric estimation in long memory in stochastic volatility and signal plus noise models. *Journal of Econometrics*, 119(1):131–154, 2004.

- [BBM96] Richard T. Baillie, Tim Bollerslev, and Hans Ole Mikkelsen. Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 74(1):3–30, 1996.
- [BCdL98] F. Jay Breidt, Nuno Crato, and Pedro de Lima. The detection and estimation of long memory in stochastic volatility. *Journal of Econometrics*, 83(1-2):325–348, 1998.
- [BM96] Tim Bollerslev and Hans Ole Mikkelsen. Modeling and pricing long memory in stock market volatility. *Journal of Econometrics*, 73(1):151–184, 1996.
- [CHL05] Willa Chen, Clifford M. Hurvich, and Yi Lu. On the correlation matrix of the discrete fourier transform and the fast solution of large toeplitz systems for long-memory time series. To appear in *Journal of the American Statistical Association*, 2005.
- [Dal71] Daryl J. Daley. Weakly stationary point processes and random measures. *Journal of the Royal Statistical Society. Series B. Methodological*, 33:406–428, 1971.
- [Dal99] Daryl J. Daley. The Hurst index of long-range dependent renewal processes. *The Annals of Probability*, 27(4):2035–2041, 1999.
- [Deo95] Rohit Deo. On GMM and QML estimation for the long memory stochastic volatility model. Working paper, 1995.
- [DGE93] Zhuangxin Ding, Clive W.J. Granger, and Robert F. Engle. A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1:83–106, 1993.
- [DH01] Rohit Deo and Clifford M. Hurvich. On the log periodogram regression estimator of the memory parameter in long memory stochastic volatility models. *Econometric Theory*, 17(4):686–710, 2001.
- [DHH05] Rohit Deo, Mengchen Hsieh, and Clifford M. Hurvich. Tracing the source of memory in volatility. Preprint, 2005.
- [DHL05] Rohit Deo, Clifford M. Hurvich, and Yi Lu. Forecasting realized volatility using a long-memory stochastic volatility model: estimation, prediction and seasonal adjustment. To appear in *Journal of Econometrics*, 2005.
- [DHSW05] Rohit Deo, Clifford M. Hurvich, Philippe Soulier, and Yi Wang. Propagation of memory parameter from durations to counts. Preprint, available on <http://www.tsi.enst.fr/soulier/dhsw.pdf>, 2005.
- [DRV00] Daryl J. Daley, Tomasz Rolski, and Rein Vesilo. Long-range dependent point processes and their Palm-Khinchin distributions. *Advances in Applied Probability*, 32(4):1051–1063, 2000.
- [DTW05] Paul Doukhan, Gilles Teyssiere, and Pablo Winant. A larch(∞) vector valued process. In Patrice Bertail, Paul Doukhan and Philippe Soulier (eds), *Dependence in Probability and Statistics*. Springer, New York, 2005.
- [DVJ03] Daryl J. Daley and David Vere-Jones. *An introduction to the theory of point processes. Vol. I: Elementary theory and methods. 2nd ed.* Probability and Its Applications. New York, NY: Springer., 2003.
- [EKM97] Paul Embrechts, Claudia Klüppelberg, and Thomas Mikosch. *Modelling Extremal Events for Insurance and Finance*. Number 33 in *Stochastic modelling and applied probability*. Berlin: Springer, 1997.
- [FRS05] Gilles Faÿ, François Roueff, and Philippe Soulier. Estimation of the memory parameter of the infinite source poisson process. Prépublication 05-12 de l’Université Paris 10, 2005.

- [GJ80] Clive W.J. Granger and Roselyne Joyeux. An introduction to long memory time series and fractional differencing. *Journal of Time Series Analysis*, 1:15–30, 1980.
- [GKL00] Liudas Giraitis, Piotr Kokoszka, and Remigijus Leipus. Stationary ARCH models: dependence structure and central limit theorem. *Econometric Theory*, 16(1):3–22, 2000.
- [GMS93] Liudas Giraitis, Stanislas A. Molchanov, and Donatas Surgailis. Long memory shot noises and limit theorems with application to Burgers' equation. In *New directions in time series analysis, Part II*, volume 46 of *IMA Vol. Math. Appl.*, pages 153–176. Springer, New York, 1993.
- [GPH83] John Geweke and Susan Porter-Hudak. The estimation and application of long memory time series models. *Journal of Time Series Analysis*, 4(4):221–238, 1983.
- [GS02] Liudas Giraitis and Donatas Surgailis. ARCH-type bilinear models with double long memory. *Stochastic Processes and their Applications*, 100:275–300, 2002.
- [Har98] Andrew C. Harvey. Long memory in stochastic volatility. In J. Knight and S. Satchell (eds), *Forecasting volatility in financial markets*. Butterworth-Heinemann, London, 1998.
- [HB93] Clifford M. Hurvich and Kaizô I. Beltrão. Asymptotics for the low-frequency ordinates of the periodogram of a long-memory time series. *Journal of Time Series Analysis*, 14(5):455–472, 1993.
- [HDB98] Clifford M. Hurvich, Rohit Deo, and Julia Brodsky. The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long-memory time series. *Journal of Time Series Analysis*, 19, 1998.
- [HHS04] Mengchen Hsieh, Clifford M. Hurvich, and Philippe Soulier. Asymptotics for duration driven long memory processes. Available on <http://www.tsi.enst.fr/soulier>, 2004.
- [HMS05] Clifford M. Hurvich, Eric Moulines, and Philippe Soulier. Estimating long memory in volatility. *Econometrica*, 73(4):1283–1328, 2005.
- [Hos81] J. R. M. Hosking. Fractional differencing. *Biometrika*, 60:165–176, 1981.
- [Hos97] Yuzo Hosoya. A limit theory for long-range dependence and statistical inference on related models. *The Annals of Statistics*, 25(1):105–137, 1997.
- [HR03] Clifford M. Hurvich and Bonnie K. Ray. The local whittle estimator of long memory stochastic volatility. *Journal of Financial Econometrics*, 1:445–470, 2003.
- [HRS98] David Heath, Sidney Resnick, and Gennady Samorodnitsky. Heavy tails and long range dependence in ON/OFF processes and associated fluid models. *Mathematics of Operations Research*, 23(1):145–165, 1998.
- [HS02] Clifford M. Hurvich and Philippe Soulier. Testing for long memory in volatility. *Econometric Theory*, 18(6):1291–1308, 2002.
- [IW71] D. L. Iglehart and Ward Whitt. The equivalence of central limit theorems for counting processes and associated partial sums. *Annals of Mathematical Statistics*, 42:1372–1378, 1971.
- [KL03] Vytautas Kazakevičius and Remigijus Leipus. A new theorem on the existence of invariant distributions with applications to ARCH processes. *Journal of Applied Probability*, 40(1):147–162, 2003.

- [Kün87] H. R. Künsch. Statistical aspects of self-similar processes. In Yu.A. Prohorov and V.V. Sazonov (eds), *Proceedings of the first World Congress of the Bernoulli Society*, volume 1, pages 67–74. Utrecht, VNU Science Press, 1987.
- [Liu00] Ming Liu. Modeling long memory in stock market volatility. *Journal of Econometrics*, 99:139–171, 2000.
- [MR04] Thomas Mikosch and Sidney Resnick. Activity rates with very heavy tails. Technical Report 1411, Cornell University; to appear in *Stochastic Processes and Their Applications*, 2004.
- [MRR02] Krishanu Maulik, Sidney Resnick, and Holger Rootzén. Asymptotic independence and a network traffic model. *Journal of Applied Probability*, 39(4):671–699, 2002.
- [MRRS02] Thomas Mikosch, Sidney Resnick, Holger Rootzén, and Alwin Stegeman. Is network traffic approximated by stable Lévy motion or fractional Brownian motion? *The Annals of Applied Probability*, 12(1):23–68, 2002.
- [Nel91] Daniel B. Nelson. Conditional heteroskedasticity in asset returns: a new approach. *Econometrica*, 59(2):347–370, 1991.
- [Par99] William R. Parke. What is fractional integration? *Review of Economics and Statistics*, pages 632–638, 1999.
- [Rob91] Peter M. Robinson. Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *Journal of Econometrics*, 47(1):67–84, 1991.
- [Rob95a] P. M. Robinson. Gaussian semiparametric estimation of long range dependence. *The Annals of Statistics*, 23(5):1630–1661, 1995.
- [Rob95b] P. M. Robinson. Log-periodogram regression of time series with long range dependence. *The Annals of Statistics*, 23(3):1048–1072, 1995.
- [Rob01] P. M. Robinson. The memory of stochastic volatility models. *Journal of Econometrics*, 101(2):195–218, 2001.
- [RR00] Sidney Resnick and Holger Rootzén. Self-similar communication models and very heavy tails. *The Annals of Applied Probability*, 10(3):753–778, 2000.
- [RvdB00] Sidney Resnick and Eric van den Berg. Weak convergence of high-speed network traffic models. *Journal of Applied Probability*, 37(2):575–597, 2000.
- [SP03] Yixiao Sun and Peter C. B. Phillips. Nonlinear log-periodogram regression for perturbed fractional processes. *Journal of Econometrics*, 115(2):355–389, 2003.
- [SV02] Donatas Surgailis and Marie-Claude Viano. Long memory properties and covariance structure of the EGARCH model. *ESAIM. Probability and Statistics*, 6:311–329, 2002.
- [Tak54] Lajos Takács. On secondary processes derived from a Poisson process and their physical applications. With an appendix by Alfréd Rényi. *Magyar Tud. Akad. Mat. Fiz. Osz. Közl.*, 4, 1954.
- [Taq03] Murad Taqqu. Fractional brownian motion and long-range dependence. In Paul Doukhan, Georges Oppenheim and Murad S. Taqqu (eds), *Theory and applications of Long-Range Dependence*. Boston, Birkhäuser, 2003.
- [TH94] Norma Terrin and Clifford M. Hurvich. An asymptotic Wiener-Itô representation for the low frequency ordinates of the periodogram of a

- long memory time series. *Stochastic Processes and their Applications*, 54(2):297–307, 1994.
- [TL86] Murad S. Taquq and Joshua M. Levy. Using renewal processes to generate long range dependence and high variability. In E. Eberlein and M.S. Taquq (eds), *Dependence in Probability and Statistics*. Boston, Birkhäuser, 1986.
- [TWS97] Murad Taquq, Walter Willinger, and Robert Sherman. Proof of a fundamental result in self-similar traffic modeling. *Computer Communication Review.*, 27, 1997.
- [Vel00] Carlos Velasco. Non-Gaussian log-periodogram regression. *Econometric Theory*, 16(1):44–79, 2000.
- [Whi02] Ward Whitt. *Stochastic-process limits*. Springer Series in Operations Research. Springer-Verlag, New York, 2002.

Appendix

Proof (of Theorem 3). For simplicity, we set the clock-time spacing $\Delta t = 1$. Define

$$S_{\tau,n}(\theta) = \sum_{k=1}^{\lfloor n\theta \rfloor} \tau_k \quad 0 \leq \theta \leq 1 ,$$

$$S_{\Delta N,n}(\theta) = \sum_{t'=1}^{\lfloor n\theta \rfloor} \Delta N_{t'} \quad 0 \leq \theta \leq 1 .$$

Since $\alpha < 2$ and $\{\tau_k\}$ is an i.i.d. sequence, by the functional central limit theorem (FCLT) for random variables in the domain of attraction of a stable law (see [EKM97, Theorem 2.4.10]), $l(n)n^{-1/\alpha}\{S_{\tau,n}(\theta) - \lfloor n\theta \rfloor\mu_\tau\}$ converges weakly in $\mathcal{D}(0, 1)$ to an α -stable motion, for some slowly varying function l . Now define

$$U_n(\theta) = (2\pi)^{-1/2}l(n)n^{-1/\alpha}\{S_{\Delta N,n}(\theta) - \lfloor n\theta \rfloor/\mu_\tau\} .$$

By the equivalence of FCLTs for the counting process and its associated partial sums of duration process (see [IW71]), U_n also converges weakly in $\mathcal{D}([0, 1])$ to an α -stable motion, say S . Summation by parts yields, for any nonzero Fourier frequency ω_j (with fixed $j > 0$)

$$l(n)n^{1/2-1/\alpha}J_{n,j}^{\Delta N} = (2\pi)^{-1/2}l(n)n^{-1/\alpha} \sum_{t'=1}^n \{\Delta N_{t'} - 1/\mu_\tau\} e^{it'\omega_j}$$

$$= \sum_{t'=1}^n \{U_n(t'/n) - U_n((t' - 1)/n)\} e^{it'\omega_j} = \int_0^1 e^{2ij\pi x} dU_n(x) .$$

Hence by the continuous mapping theorem

$$\sqrt{2\pi} l(n)n^{1/2-1/\alpha} J_{n,j}^{\Delta N} \xrightarrow{d} \int_0^1 e^{2i\pi jx} dS(x)$$

which is a stochastic integral with respect to a stable motion, hence has a stable law.

To prove the second statement of the theorem, note that for fixed j and as $n \rightarrow \infty$, $f(\omega_j) \sim l_1(n)\omega_j^{-2d}$ for some slowly varying function l_1 , so

$$\begin{aligned} f^{-1/2}(\omega_j) J_{n,j}^{\Delta N} &= \frac{l(n)n^{1/\alpha-1/2}}{f^{1/2}(\omega_j)} \frac{J_{n,j}^{\Delta N}}{l(n)n^{1/\alpha-1/2}} \\ &\sim C_1 l(n)n^{1/\alpha+\alpha/2-3/2} \frac{J_{n,j}^{\Delta N}}{\mu_\tau^{-1-1/\alpha} l(n)n^{1/\alpha-1/2}}. \end{aligned} \quad (21)$$

Since $1/\alpha + \alpha/2 - 3/2 < 0$, we have $l(n)n^{1/\alpha+\alpha/2-3/2} \rightarrow 0$. Hence by Slutsky's Theorem, (21) converges to zero. \square

Proof (of Theorem 4). Let $S_n(t) = n^{-H} \sum_{k=1}^{\lfloor nt \rfloor} (\tau_k - \mathbb{E}[\tau_k])$, $t \in (0, 1)$. It is shown in Surgailis and Viano (2002) that $S_n(t) \xrightarrow{d} B_H(t)$ in $\mathcal{D}([0, 1])$ where $B_H(t)$ is fractional Brownian motion with Hurst parameter $H = d + 1/2$. Thus, by Iglehart and Whitt (1971), it follows that $t^{-H} N \rightarrow AB_H$ in $\mathcal{D}([0, 1])$, where A is a nonzero constant. The result follows as above by the continuous mapping theorem and summation by parts. \square