Methodol Comput Appl Probab
DOI 10.1007/s11009-008-9105-3

Robust Wavelet-Domain Estimation of the Fractional Difference Parameter in Heavy-Tailed Time Series: An Empirical Study

Agnieszka Jach · Piotr Kokoszka

Received: 5 July 2007 / Revised: 9 August 2008 /
Accepted: 18 September 2008

Abstract We investigate the performance of several wavelet-based estimators of the fractional difference parameter. We consider situations where, in addition to long-range dependence, the time series exhibit heavy tails and are perturbed by polynomial and change-point trends. We make detailed study of a wavelet-domain pseudo Maximum Likelihood Estimator (MLE), for which we provide an asymptotic and finite-sample justification. Using numerical experiments, we show that unlike the traditional time-domain estimators, estimators based on the wavelet transform are robust to additive trends and change points in mean, and produce accurate estimates even under significant departures from normality. The Wavelet-domain MLE appears to dominate a regression-based wavelet estimator in terms of smaller root mean squared error. These findings are derived from a simulation study and application to computer traffic traces.

Keywords Fractional difference · Heavy tails · Trend · Wavelets

AMS 2000 Subject Classifications 62M10 · 42C40

1 Introduction

Advances in monitoring computer network traffic over the last two decades have led to the emergence of large data sets consisting of long time series exhibiting

Research supported by NSF grant DMS-0413653.

A. Jach (✉)
Departamento de Estadística, Universidad Carlos III de Madrid, 28903 Getafe, Madrid, Spain
e-mail: ajach@est-econ.uc3m.es

P. Kokoszka
Department of Mathematics and Statistics, Utah State University, Logan, UT 84322, USA
e-mail: Piotr.Kokoszka@usu.edu

burstiness (extreme variability of traffic intensities) and persistence (slowly decaying correlations). These are often modeled by stochastic processes with heavy-tailed marginal distributions and long-range dependence (LRD). Kogon and Manolakis (1995) used infinite-variance α-stable Fractional Auto-Regressive Integrated Moving Average (FARIMA) and Linear Fractional Stable Noise (LFSN) to model signals from infrared scenes. Ilow (2000) applied heavy-tailed FARIMA to predict bandwidth requirements for network traffic. Harmantzis and Hatzinakos (2001) and Karasaridis and Hatzinakos (2001) modeled bursty Ethernet traffic with LFSN and α-stable FARIMA with the tail index α in the range $1 < \alpha < 2$. Using a wavelet-based estimator, Lee and Fapojuwo (2005) found that the packet count processes for wireless data are LRD with the tail index $1 < \alpha < 2$.

This paper is concerned with estimating the fractional difference parameter d if the, possibly transformed, observations $X_0, X_1, \ldots, X_{N-1}$ follow the model

$$X_t = (1-B)^{-d} Y_t + f(t), \quad 0 < d < 1/2, \tag{1}$$

where B is the back-shift operator, $BY_t = Y_{t-1}$, and Y_t is a zero-mean short-memory (weakly dependent) linear process with heavy-tailed marginal distributions, i.e.,

$$Y_t = \sum_{j=1}^{\infty} \psi_j Z_{t-j}, \quad \sum_{j=1}^{\infty} |\psi_j| < \infty. \tag{2}$$

We work with the innovations Z_t which have an infinite fourth moment, or even infinite variance, and have regularly varying probability tails, i.e., there is a number $\alpha > 0$ such that for every constant $a > 1$

$$\lim_{z\to\infty} \frac{P(Z_t > az)}{P(Z_t > z)} = a^{-\alpha}, \quad 1 < \alpha < 4. \tag{3}$$

For the deterministic functions $f(t)$ that we use in this paper, the LRD process X_t also has regularly varying tails with the same index α as the innovation process Z_t. Property (3) is known to approximately hold true for many network traffic data types (Resnick 1997).

A popular model of this type (with $f(t) = 0$) is the FARIMA(p, d, q) model, defined by the equations

$$\Phi_p(B) X_t = \Theta_q(B)(1-B)^{-d} Z_t, \tag{4}$$

see Sections 7.12 and 7.13 of Samorodnitsky and Taqqu (1994) for further details. We only note here that if $1 < \alpha < 2$, equations Eq. 4 have a stationary solution only when

$$d < 1 - 1/\alpha. \tag{5}$$

Thus, in the infinite-variance case, the range of admissible values of d is smaller than in the finite-variance case, where only $d < 1/2$ is needed.

Over the last decade, it has been recognized that inference under the assumption of LRD can be very inaccurate if the observed time series is, in fact, not purely stationary LRD, but is perturbed by a trend or change point type nonstationarities: time and spectral domain estimates of the intensity of LRD become inflated, and significance tests spuriously reject. It is therefore important to develop methods of robust inference for LRD processes. Methods of testing for and estimating "pure"

LRD in the presence of trends and change points have recently been developed, see Berkes et al. (2006), Craigmile et al. (2005), Jach and Kokoszka (2008). These papers rely on the assumption that the data is normal, or at least the existence of the fourth moment.

In this work, we show that methods based on the wavelet transform provide good estimators of d in model (1–3) if the function $f(t)$ is piecewise polynomial, even with discontinuities. In particular, we demonstrate that wavelet pseudo Maximum Likelihood Estimator (MLE) works well in such situations. This is theoretically justified by the invariance property of the DWT with respect to additive polynomials of relatively low order (if the wavelet filter $\{h_l\}$ used to compute DWT is of length L, then the (nonboundary) wavelet coefficients are uninfluenced by the polynomials of order K, $K \leq L/2 - 1$). Also, if there are relatively few break points compared to the length of the series, only a few DWT coefficients will be affected by these breaks. By contrast, if a periodogram is used, as in the Whittle likelihood (Robinson 1995a), all Fourier coefficients are affected.

Our study complements the recent paper of Moulines et al. (2008) who studied this estimator under the assumption of a finite fourth moment and assuming $f(t) = 0$ (but allowing difference nonstationarity). We also expand on the work of Stoev and Taqqu (2005) who showed that a regression wavelet-domain estimator is applicable to α-stable processes if $f(t) = 0$. We reveal that this estimator produces good results if α is estimated from the wavelet coefficients.

The above methods of estimating d are based on the decorrelation property of the wavelet transform (Percival and Walden 2000, Chapter 9) and here as well we take advantage of this characteristic (expressed as Assumption A1 below). Independence combined with the Gaussianity (Assumption A2) of the wavelet coefficients makes it very convenient to calculate and justify the pseudo MLE. We also make use of a specific formula for level j scale parameter, σ_j, (Assumption A3) through which we introduce the unknown parameter into the derivations. This formula also permits to find MLE of σ_j explicitly and consequently leads to a simple expression of the pseudo-likelihood. For ease of reference we list the assumptions on the wavelet coefficients that we refer to henceforth. These assumptions are never exactly satisfied, but as we mentioned above, they are used to derive a tractable wavelet-domain pseudo-likelihood function. We use the notation of Percival and Walden (2000) and, in particular, denote the wavelet coefficients at level (octave) j and time $2^j k$ by $W_{j,k}$. They are obtained by circularly filtering the data with level j filter $\{h_{j,l}\}$ (based on the underlying wavelet filter $\{h_l\}$)

$$W_{j,k} = \sum_{l=0}^{L_j-1} h_{j,l} X_{2^j(k+1)-1-l \bmod N},$$

where $j = 1, 2, \ldots, J$, $J = \lfloor \log_2(N) \rfloor$ and $k = 0, 1, \ldots, N_j - 1$, $N_j = 2^{J-j}$. We assume that for appropriately chosen levels $j = j_0, \ldots, J_0$:

A1) The vectors $\boldsymbol{W}_j = [W_{j,0}, W_{j,1}, \ldots, W_{j,N_j-1}]^T$ are independent, and within each of these vectors the $W_{j,k}$ are independent.

A2) Each $W_{j,k}$ is a Gaussian random variable with mean zero and standard deviation σ_j. (This assumption is obviously not satisfied by the wavelet coefficients of heavy-tailed processes, but we use it to calculate the pseudo MLE.)

A3) The scale parameters σ_j satisfy the relationship

$$\sigma_j = 2^{(j-j_0)d}\sigma_{j_0}, \tag{6}$$

which follows from the approximate self-similarity of the time series models we consider and the definition of the wavelet coefficients, see Stoev and Taqqu (2005).

The remainder of the paper is organized as follows. In Section 2, we describe the estimation methods considered in this paper, focusing on the wavelet pseudo MLE whose justification is provided in Section 3. Section 4 contains a simulation study, while Section 5 covers application to computer traffic traces. The approximate independence of the wavelet coefficients is investigated in finite samples by applying a statistical test of independence (we address this issue in Appendix). Theoretically, it has been shown that measures of dependence appropriate for α-stable processes vanish in an asymptotic sense when calculated for the wavelet coefficients of α-stable FARIMA and similar processes (Pipiras et al. 2007, and references therein).

2 Wavelet-Domain Estimation of *d*

In this section we briefly describe five methods of estimating d. The first three, Wavelet-domain MLE (WMLE), LOG and LOG LOG are based on the wavelet transform. The last two, FD and FARIMA, are well-known time-domain approximate parametric likelihood methods. We describe the WMLE method in most detail because it has not previously been investigated neither theoretically nor via simulations in the context of heavy-tailed time series. Moulines et al. (2008) recently developed a detailed asymptotic theory for the WMLE of Gaussian processes, and to some extent for linear processes with a finite fourth moment.

The FD and FARIMA methods are included because they are known to have good asymptotic and finite-sample properties for heavy-tailed stationary fractionally differenced processes if the model is specified correctly (Kokoszka and Taqqu 1996; Taqqu and Teverovsky 1998). The widely-used semiparametric spectral-domain estimators of Geweke and Porter-Hudak (1983) and Robinson (1995a, b) are not compared, because there is no theoretical justification for their use in heavy-tailed series (spectral density does not even exist if $\alpha < 2$), and, in the presence of trends, they are likely to exhibit large bias, like the FD and FARIMA estimators. Data tapering, similar to that proposed in Lobato and Velasco (2000) and Velasco and Robinson (2000), could reduce the bias. However, these modifications are not investigated here because our focus is on estimators based on the wavelet transform.

Our derivations of the wavelet-based estimators are motivated by the approximate properties A1-A3, discussed in Section 1.

WMLE: According to Assumption A3, we have the following relationship for σ_j^2, $j = j_0 + 1, j_0 + 2, \ldots, J_0$

$$\sigma_j^2 = 2^{2(j-j_0)d}\sigma_{j_0}^2.$$

Introduce the following quantity

$$s_j(d) = 2^{2(j-j_0)d}.$$

To derive the estimator, we assume (incorrectly) that at a given level j, the wavelet coefficients $W_{j,k}$ are zero-mean normal random variables with variance $2\sigma_j^2$, i.e., with the density

$$\left(2\pi 2 s_j(d)\sigma_{j_0}^2\right)^{-1/2} \exp\left\{-\frac{1}{2}\frac{w_{j,k}^2}{2s_j(d)\sigma_{j_0}^2}\right\}.$$

Thus, the likelihood function is

$$\begin{aligned} l(d,\sigma_{j_0}^2) &= \prod_{j=j_0}^{J_0}\prod_{k=0}^{M_j-1}\left(2\pi 2 s_j(d)\sigma_{j_0}^2\right)^{-1/2} \exp\left\{-\frac{1}{2}\frac{w_{j,k}^2}{2s_j(d)\sigma_{j_0}^2}\right\} \\ &= (2\pi 2\sigma_{j_0}^2)^{-M/2}\prod_{j=j_0}^{J_0}\left(s_j(d)\right)^{-M_j/2}\exp\left\{-\frac{1}{2}\frac{\sum_{k=0}^{M_j-1} w_{j,k}^2}{2s_j(d)\sigma_{j_0}^2}\right\}, \end{aligned}$$

where $M=\sum_{j=j_0}^{J_0} M_j$ and M_j is the number of nonboundary wavelet coefficients at level j, and twice the negative log-likelihood

$$-2\ln l(d,\sigma_{j_0}^2) = M\ln(2\pi 2\sigma_{j_0}^2)+\sum_{j=j_0}^{J_0}\left[M_j\ln(s_j(d))+\frac{R_j}{2s_j(d)\sigma_{j_0}^2}\right],$$

where

$$R_j=\sum_{k=0}^{M_j-1} w_{j,k}^2.$$

For any d we can compute the MLE of $\sigma_{j_0}^2$ by noting that

$$\frac{\partial}{\partial\sigma_{j_0}^2}\left[-2\ln l\left(d,\sigma_{j_0}^2\right)\right]=\frac{M}{\sigma_{j_0}^2}-\frac{1}{2\left(\sigma_{j_0}^2\right)^2}\sum_{j=j_0}^{J_0}\frac{R_j}{s_j(d)}.$$

Hence the MLE of $\sigma_{j_0}^2$ is

$$\hat\sigma_{j_0}^2(d)=\frac{1}{2M}\sum_{j=j_0}^{J_0}\frac{R_j}{s_j(d)}.$$

If we plug this estimator into $-2\ln l$ then we will obtain a function of d only

$$\begin{aligned} -2\ln l(d,\hat\sigma_{j_0}^2(d)) &= M\ln\left(2\pi 2\hat\sigma_{j_0}^2(d)\right)+\sum_{j=j_0}^{J_0}M_j\ln\left(s_j(d)\right)+\overbrace{\sum_{j=j_0}^{J_0}\frac{R_j}{2s_j(d)\hat\sigma_{j_0}^2(d)}}^{M} \\ &= M\left\{\ln\left(2\pi 2\hat\sigma_{j_0}^2(d)\right)+1\right\}+\sum_{j=j_0}^{J_0}M_j\ln\left(s_j(d)\right). \end{aligned}$$

Note that the above expression agrees with equation (9) in Craigmile et al. (2005) for the case of $j=1,2,\ldots,J$ with $s_{j,0}(d)$ replaced by $s_j(d)$.

The value of d which minimizes the $-2\ln l(d, \hat{\sigma}^2_{j_0}(d))$ above is the WMLE.

LOG: This is the regression estimator studied by Stoev and Taqqu (2005), and also denoted "log" in that paper. Essentially, it is an estimator of the self-similarity parameter H, which can be used to estimate d, provided the stability index α is *known*, by using the relationship $d = H - 1/\alpha$. This estimator can thus be practically applied in connection with a method of estimating α, or in the finite-variance case by setting $\alpha = 2$. The latter case has been extensively studied for at least a decade, see references in Abry et al. (2006) and in-depth simulations in Teyssiere and Abry (2006).

LOG LOG: This is the LOG estimator combined with a specific method of estimating α. The stability index of the underlying long-memory process X_t is carried over to the wavelet domain. In particular, the sequences $W_{1,k}$ can be regarded as a set of approximately independent identically distributed (iid) α-stable random variables (no memory), allowing us to estimate α. The choice of $j = 1$ is justified by the highest number of observations available here. In the preliminary study we applied the novel approach of Stoev et al. (2006) based on max self-similarity to estimate α from $W_{1,k}$, but we found that the following method yields a smaller root mean squared error (RMSE) of the LOG LOG estimator. Compute LOG estimator of H based on $W_{1,k}$ and use the relationship $0 = H - 1/\alpha$ to derive an estimator of α denoted by $\hat{\alpha}$. Replace unknown α by $\hat{\alpha}$ and compute the LOG estimator of d same way as in the previous method.

FD and FARIMA: These two parametric estimators assume FD(d) and FARIMA(1,d,1) models, respectively. They use approximate time-domain Gaussian likelihood.

3 Heuristic Justification of the WMLE Procedure

We now provide an informal asymptotic justification for the WMLE procedure. A rigorous justification, under the assumption of a finite fourth moment and some other technical assumptions, has recently been developed by Moulines et al. (2008). We explain below that this estimator can, in fact, be expected to be consistent if only the second moment is assumed, or if the marginal distributions have the tail index $1 < \alpha < 2$. Since the focus of this paper is on the numerical assessment of the performance of the various estimators, we do not attempt to develop a detailed asymptotic theory under these weak assumptions, as it is likely to be complex. Such a theory exists for the LOG estimator (Stoev and Taqqu 2005), where unlike the WMLE, it is assumed that α is known.

For simplicity assume that

$$j_0 = 1, \quad J_0 = J, \quad M_j = 2^{J-j}, \quad M = \sum_{j=1}^{J} M_j = 2^J - 1 \sim 2^J,$$

where $\sim$ indicates that the ratio of the left- and right-hand sides tends to 1 as M (equivalently J) tends to infinity, so that

$$s_j(d) = 2^{2(j-1)d}, \quad \hat{\sigma}^2_{j_0}(d) = \hat{\sigma}^2(d) = \frac{1}{2M}\sum_{j=1}^{J}\frac{1}{s_j(d)}\sum_{k=0}^{M_j-1} W^2_{j,k}.$$

We consider two cases: 1) the observations X_t have a finite second moment; 2) the X_t are heavy-tailed with the index $1 < \alpha < 2$. In both cases we assume that conditions A1–A3 of Section 1 hold, in particular, that the $W_{j,k}$ are approximately independent. We denote by d_0 the true value of the difference parameter.

In case 1), by the strong law of large numbers

$$\frac{1}{M_j}\sum_{k=0}^{M_j-1} W^2_{j,k} \overset{a.s.}{\to} EW^2_{j,0} \tag{7}$$

and

$$EW^2_{j,0} = 2^{2(j-1)d_0}\sigma_1^2. \tag{8}$$

In case 2), we have

$$\frac{1}{M_j^{2/\alpha}}\sum_{k=0}^{M_j-1} W^2_{j,k} \overset{d}{\to} \sigma_j^2 S_j, \tag{9}$$

where the S_j are iid, a.s. positive, $\alpha/2$-stable random variables with a scale parameter which does not depend on d, and

$$\sigma_j^2 = 2^{2(j-1)d_0}\sigma_1^2. \tag{10}$$

Relationship (9) is rigorously proved if the $W_{j,k}$ are independent (Theorem 3.3 of Davis and Resnick 1986).

Focus first on case 1). Denote

$$L_M(d) = M\left\{\ln(4\pi\hat{\sigma}^2(d)) + 1\right\} + \sum_{j=1}^{J} M_j \ln(s_j(d)). \tag{11}$$

We will show that as M diverges to infinity, $M^{-1}L_M(d) \overset{a.s.}{\to} L(d)$, and that $L(d)$ has minimum at d_0. By Eq. 7 and Eq. 8,

$$\hat{\sigma}^2(d) = \frac{1}{2M}\sum_{j=1}^{J} 2^{-2(j-1)d}2^{J-j}\frac{1}{M_j}\sum_{k=0}^{M_j-1} W^2_{j,k} \overset{a.s.}{\to} \frac{\sigma_1^2}{2}\sum_{j=1}^{\infty} 2^{2(j-1)(d_0-d)}2^{-j} =: \sigma^2(d).$$

By setting

$$u = 2(d - d_0) + 1,$$

we obtain

$$\sigma^2(d) = \frac{\sigma_1^2}{4}\sum_{i=0}^{\infty} 2^{-iu} = \frac{\sigma_1^2}{4}\frac{1}{1-2^{-u}}.$$

Observe that $u > 0$ iff $d - d_0 > -1/2$. The second term in $L_M(d)$ is

$$\sum_{j=1}^{J} M_j \ln(s_j(d)) = (M+1)\sum_{j=1}^{J} 2^{-j} 2(j-1) d \ln 2 = (M+1) d \ln 2 \sum_{i=0}^{J-1} i 2^{-i}$$
$$\sim 2Md \ln 2 = M\{u \ln 2 + C(d_0)\}. \tag{12}$$

Thus, we obtain

$$M^{-1} L_M(d) \xrightarrow{a.s.} \left\{ \ln\left(4\pi \frac{\sigma_1^2}{4} \frac{1}{1-2^{-u}}\right) + 1 \right\} + u \ln 2 + C(d_0)$$
$$= \ln(\pi\sigma_1^2) - \ln(1-2^{-u}) + 1 + u \ln 2 + C(d_0).$$

By substituting $x = 2^{-u}$, the estimator converges to the minimum of

$$f(x) = -\ln(1-x) + \ln\frac{1}{x} = -[\ln(1-x) + \ln x].$$

Since

$$f'(x) = -\left[\frac{1}{x} - \frac{1}{1-x}\right],$$

$f'(x) = 0$ iff $x = 1/2,\ u = 1,\ d = d_0$. Since for $0 < x < 1$,

$$f''(x) = \frac{1}{x^2} + \frac{1}{(1-x)^2} > 0,$$

d_0 is the minimum.

Using the notation introduced above, we now turn to case 2) (heavy tails with $1 < \alpha < 2$). By Eq. 9 and Eq. 10 we have

$$\hat{\sigma}^2(d) = \frac{1}{2M} \sum_{j=1}^{J} 2^{-2(j-1)d} M_j^{-2/\alpha} M_j^{2/\alpha} \sum_{k=0}^{M_j-1} W_{j,k}^2$$
$$\xrightarrow{d} \frac{1}{2M} \sum_{j=1}^{J} 2^{-2(j-1)d} \left(2^{J-j}\right)^{2/\alpha} \sigma_j^2 S_j$$
$$= \frac{\sigma_1^2}{2M} (M+1)^{2/\alpha} \sum_{j=1}^{J} 2^{-(j-1)[2(d-d_0)+2/\alpha]} 2^{-2/\alpha} S_j.$$

By replacing

$$u = 2(d - d_0) + 2/\alpha,$$

we obtain

$$\hat{\sigma}^2(d) \xrightarrow{d} \frac{\sigma_1^2}{2^{1+2/\alpha}} (M+1)^{2/\alpha} M^{-1} \sum_{i=0}^{\infty} 2^{-iu} S_i.$$

Since the S_i are iid $\alpha/2$-stable,

$$\sum_{i=0}^{\infty} 2^{-iu} S_i \stackrel{d}{=} \left(\sum_{i=0}^{\infty} 2^{-iu\alpha/2}\right)^{2/\alpha} S_0 = \left(\frac{1}{1-2^{-u\alpha/2}}\right)^{2/\alpha} S_0.$$

Consequently,

$$\hat{\sigma}^2(d) \xrightarrow{d} \frac{\sigma_1^2}{2^{1+2/\alpha}} (M+1)^{2/\alpha} M^{-1} \left(1-2^{-u\alpha/2}\right)^{-2/\alpha} S_0$$

and so

$$\ln(4\pi\hat{\sigma}^2(d)) - \ln\left((M+1)^{2/\alpha} M^{-1}\right) \xrightarrow{d} C(\sigma_1^2, \alpha) - \frac{2}{\alpha}\ln(1-2^{-u\alpha/2}) + \ln S_0. \quad (13)$$

By combining Eq. 12 with Eq. 13, we see that the estimator converges to the minimum of

$$f_\alpha(x) = -\frac{2}{\alpha}\ln(1-x^{\alpha/2}) - \ln x,$$

where $x = 2^{-u}$. The derivative

$$f'_\alpha(x) = \frac{x^{\alpha/2-1}}{1-x^{\alpha/2}} - \frac{1}{x}$$

vanishes iff $x = 2^{-2/\alpha},\ u = 2/\alpha,\ d = d_0$. Direct computation shows that the second derivative is positive.

4 Finite Sample Performance

In this section we provide a numerical assessment of the estimation procedures presented in Section 2 in infinite- and finite-variance scenarios.

First, we focus on α-stable time series with $1 < \alpha < 2$. As data generating processes, we use 12 FARIMA models listed in Table 1 of Stoev and Taqqu (2005), which we restate in Table 1. These models also include the short-memory case $d = 0$.

We add a deterministic trend function, either a second-order polynomial or a change-point function, to the realizations of every model . The polynomial trend is of the form

$$f_1(t) = a(t-b)^2, \quad a = 1/(N\sqrt{N}),\ b = N/2,$$

and its modification $f_2(t) = f_1(t), t \geq b$ and 0 otherwise.

We also study a change-point trend, in which we partition the set $I = \{1, 2, \ldots, N\}$ into p intervals I_i setting

$$f_3(t) = \sum_{i=1}^{p} a_i \mathbf{1}\{t \in I_i\},$$

and focus on $p \in \{2, 4\}$ and (half) jumps of the form $a_i = (-1)^{i+1}a$, where $a \in \{0.5, 1, \ldots, 3\}$.

Realizations of selected α-stable FARIMA models with the above trend functions are presented in Fig. 1.

Table 1 Models of α-stable FARIMA(p, d, q) used in the simulations. AR and MA parameters are $\phi = 0.7$, $\theta = 0.3$, respectively

Model	α	FARIMA
1	1.2	(0,0,0)
2	1.2	(1,0,1)
3	1.2	(0,0.1,0)
4	1.2	(1,0.1,1)
5	1.5	(0,0,0)
6	1.5	(1,0,1)
7	1.5	(0,0.1,0)
8	1.5	(1,0.1,1)
9	1.5	(0,0.2,0)
10	1.5	(1,0.2,1)
11	1.5	(0,0.3,0)
12	1.5	(1,0.3,1)

We generate $R = 50$ replications of length $N = 10000$ for each of the 12 models, using Fast Fourier Transform procedure described in Stoev and Taqqu (2004) with parameters $N = 2^{14}$ and $M = 2^{14}$. For the wavelet estimators, we use only nonboundary coefficients ($j = j_0, j_0 + 1, \ldots, J_0$) obtained with Daubechies filter of length 6. The upper cut-off J_0 corresponds to the largest available scale without boundary coefficients ($J_0 = 10$). The lower cut-off $j_0 = 5$ balances the effect of the bias resulting from the short-range dependence structure in the time series, against the influence of the variance that becomes large if j_0 is big. In this simulation study, the choice of R, j_0, J_0 and the wavelet filter is the same as in Stoev and Taqqu (2005). In the LOG method, weights are derived from diagonal regression weight matrix. In the LOG LOG procedure we estimate α based on $W_{1,k}$ using a LOG scheme over

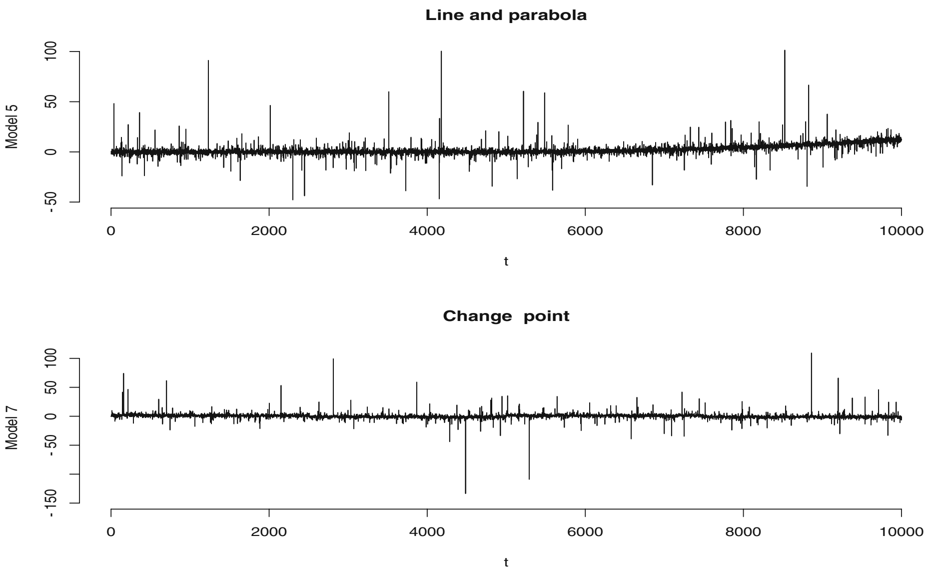

Fig. 1 Realizations of two selected models of α-stable FARIMA (Table 2) with trend function $f_2(t)$ (*top panel*) and $f_3(t)$, $a = 1$, $p = 4$ (*bottom panel*)

Table 2 Models of finite-variance FARIMA(p, d, q) models with $t(3)$ innovations used in the simulations

Model	FARIMA
1	(0,0,0)
2	(1,0,1)
3	(0,0.1,0)
4	(1,0.1,1)
5	(0,0.2,0)
6	(1,0.2,1)
7	(0,0.3,0)
8	(1,0.3,1)

AR and MA parameters are $\phi = 0.7, \theta = 0.3$, respectively

scales from $j = 3$ to $j = 8$ with nonboundary coefficients and a diagonal regression weight matrix. WMLE estimates are obtained via numerical optimization (`R` function 'optimize', $d \in (-0.01, 0.5)$). `R` code used to perform these computations is available upon request. FD and FARIMA estimates come from `R` routine `fracdiff`. RMSEs of the five estimators are presented in Figs. 2 and 3. Tables of biases and SDs are available upon request. All estimators are applied to the same realizations of the underlying model.

Polynomial-type trend, infinite variance. RMSEs in Fig. 2 for both trend functions show similar patterns. Wavelet estimators are more robust with respect to model-specification than to time-domain estimators, whose behavior is very erratic. RMSE of the FD estimator exhibits swings from about 0.05 to about 0.4, RMSE of the FARIMA estimator from 0.05 to 0.5, whilst RMSEs of the wavelet estimators consistently stay around 0.12. RMSE of the FARIMA estimator seems to explode in models 5–8 ($\alpha = 1.5$ and no or little memory). As expected, in most cases, LOG LOG performs worse than LOG estimator, and WMLE beats LOG LOG. Biases of the wavelet estimators are 0.05 or less (LOG and LOG LOG slightly outperform WMLE), thus their contribution to the RMSEs is very small. Biases of the time

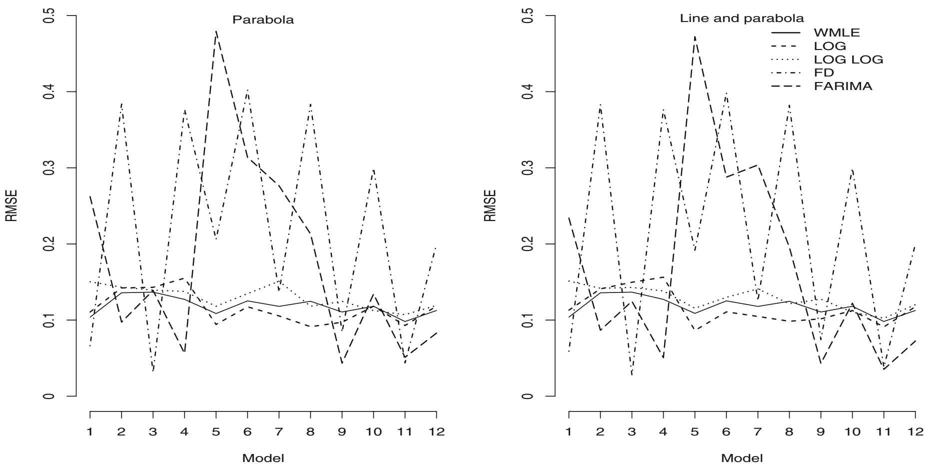

Fig. 2 RMSE of estimators of memory parameter in α-stable FARIMA models (Table 1) with trend functions $f_1(t)$ (*left panel*) and $f_2(t)$ (*right panel*)

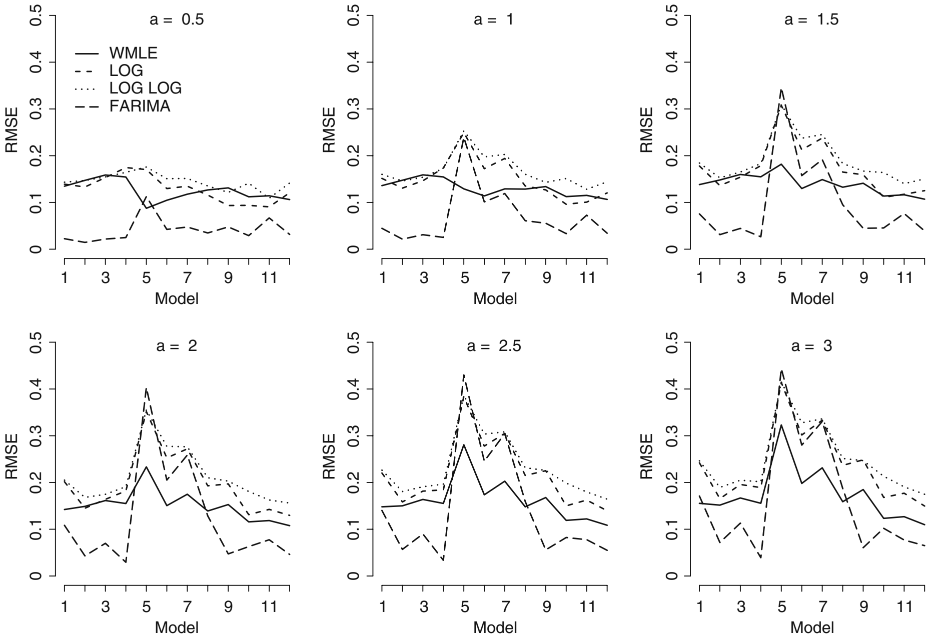

Fig. 3 RMSE of estimators of memory parameter in α-stable FARIMA models (Table 1) with change-point trend $f_3(t)$, a as indicated, $p = 4$

estimators behave similarly to the corresponding RMSEs in Fig. 2, but with slightly smaller values. The time estimators suffer from huge bias. Standard Deviations (SDs) of the wavelet estimators are about 0.12 (WMLE outperforms LOG and LOG LOG) and are larger than SDs of the FD and FARIMA estimators, except for models 5–7, where SD of FARIMA rapidly increases to match or exceed wavelet SDs.

Change-point trend, infinite variance. We show the results for a change-point trend with $p = 4$ (three jumps), because RMSEs for $p = 2$, although smaller by about 0.01, behave similarly. In our discussion we omit the results for the FD estimator, as its erratic behavior obscures the plots. There is a clear effect of a jump size, $\Delta = 2a$, on RMSEs in models 5–7, especially for the FARIMA estimator, but also for the LOG and LOG LOG estimators. WMLE seems to be less influenced by the trend than other estimators. In general, RMSEs of the wavelet estimators stay robustly around 0.12 for relatively small jumps (top panel of Fig. 3) and 0.15 for remaining, except for models 5–7, with WMLE outperforming LOG and LOG LOG. In most cases, RMSE of the FARIMA estimator is smaller than RMSEs of the wavelet estimators, however, the former becomes very large in models 5–7 for almost all jump-sizes. As with polynomial-type trends, the RMSEs of all the estimators evaluated are mostly influenced by their respective biases and to a lesser extent by their SDs. The SD of WMLE is about 0.1, and falls between the SDs of the regression estimators and that of FARIMA. In general, the latter is smaller than the SDs of all the former estimators.

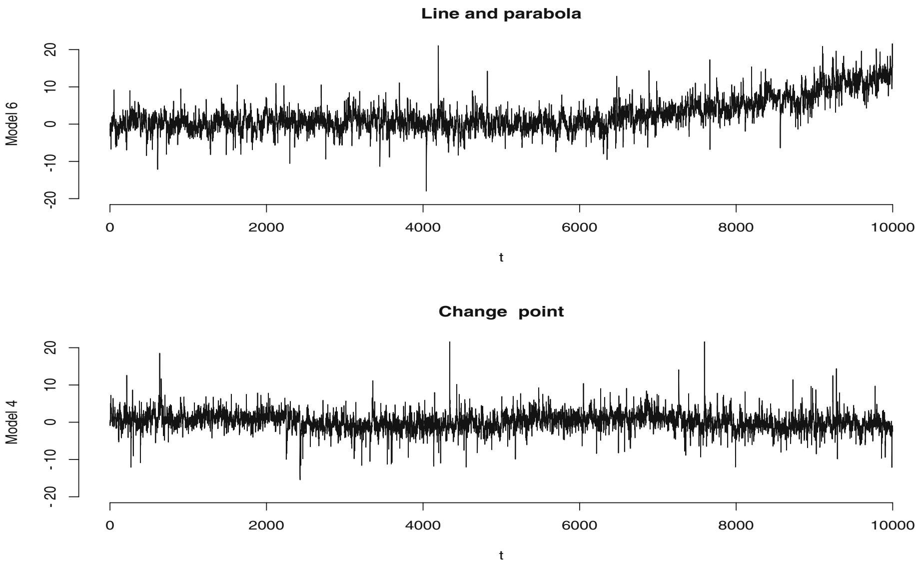

Fig. 4 Realizations of two selected models of finite-variance FARIMA with $t(3)$ innovations (Table 2) with trend function $f_2(t)$ (*top panel*) and $f_3(t)$, $a = 1$, $p = 4$ (*bottom panel*)

We now turn to the finite-variance case. We repeat our experiments by replacing α-stable innovations with Student's $t(3)$ (3 is the number of degrees of freedom), and thus we reduce the number of models from 12 to the 8 models described in Table 2.

We have chosen the $t(3)$ distribution because it has an infinite fourth moment, and so is not covered by the theory of Moulines et al. (2008). For the LOG estimator, $\alpha = 2.0$ was used. Realizations of two finite-variance FARIMA model with polynomial and change-point trends are presented in Fig. 4. Bias, SD and RMSE of the five estimators are presented.

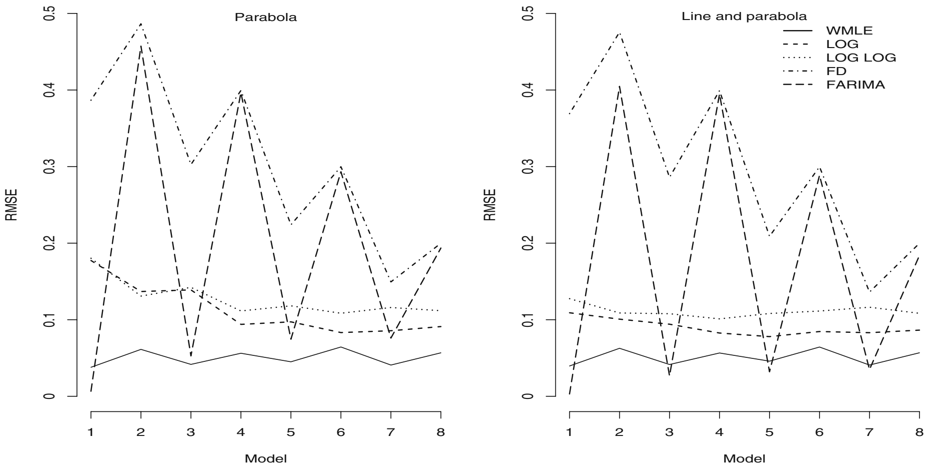

Fig. 5 RMSE of estimators of memory parameter in finite-variance FARIMA models with $t(3)$ innovations (Table 2) with trend functions $f_1(t)$ (*left panel*) and $f_2(t)$ (*right panel*)

Polynomial-type trend, finite variance. As with the infinite-variance case, RMSEs in Fig. 5 for both trends mimic each other. The WMLE estimator is optimal in almost all models with a RMSE of about 0.05, which is consistently lower, by about 0.05, than RMSEs of LOG and LOG LOG. The time estimators are hugely biased and model-sensitive, and consequently their RMSEs show undesirable behavior. The hierarchy dictated by the SD is: FD (the smallest), FARIMA, WMLE (of about 0.04), LOG and LOG LOG.

Change-point trend, finite variance. The effect of the number of jumps (three versus one) is somewhat stronger in the finite-variance case than in the α-stable case, and the performance of all estimators declines rapidly when the jump size is relatively big ($\Delta > 2$). WMLE seems to be better than the other estimators, it has a low RMSE compared to LOG and LOG LOG, and is more robust than FARIMA for various models (Fig. 6). It provides good results for jumps smaller than 2, especially when there is only one jump, but for bigger jumps it is not appropriate. All RMSEs are hugely influenced by the biases. The SD of the WMLE is very close to that of FARIMA (≈ 0.02), and is smaller than that of LOG, which in turn is smaller than the SD of LOG LOG.

We summarize the above discussion in the following points:

- Wavelet estimators clearly outperform time-domain estimators when a polynomial-type trend is present, they have smaller RMSEs and are more robust to model-specification;

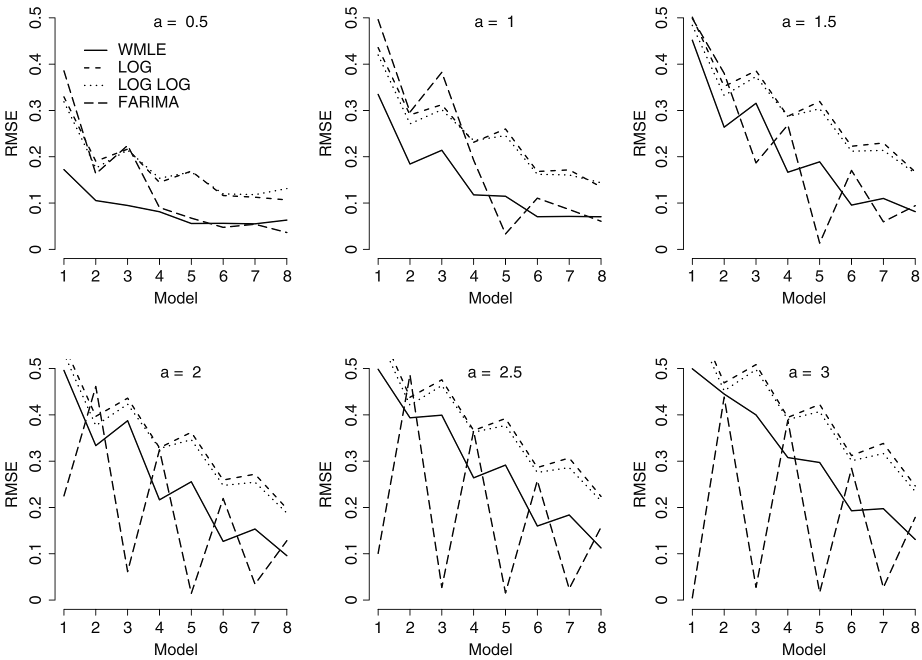

Fig. 6 RMSE of estimators of memory parameter in finite-variance FARIMA models with $t(3)$ innovations (Table 2) with change-point trend $f_3(t)$, a as indicated, $p = 4$

- WMLE outperforms LOG LOG and LOG in the finite-variance case; in the α-stable case the WMLE is slightly better than LOG LOG and slightly worse than (infeasible) LOG;
- In the α-stable case, the LOG LOG method performs well. This is surprising because the estimation of α is known to be very difficult, particularly for dependent data;
- In the infinite-variance setting, RMSEs of the wavelet estimators, although slightly larger in the presence of the change points, show more resistance to the model-type than time estimators; the RMSEs of those may become very large in some cases;
- In the finite-variance case it makes sense to estimate d if jump-sizes are comparable with the variance of the noise process; in such cases WMLE outperforms all the other estimators;
- Jump-size has a bigger influence on the performance of the estimators than the number of jumps;
- Wavelet estimators are much faster to obtain, e.g., for pAug time series of Section 5 of length $N = 2^{17}$, computations take: less than 1 s in WMLE; 1 s in LOG LOG; 2 s in FD; 4 s in FARIMA (on PC Pentium(R) 4 CPU 2.8 GHz, 1.5 GB of RAM with Windows XP).

5 Application to Network Traffic Data

We consider two groups of Ethernet traces. The first group consists of relatively old (1994, 1998) time series of aggregated packet counts that were used in the context of modeling with α-stable FARIMA. The second group includes relatively new (2003) time series of aggregated packet bytes that were utilized in a detailed study of LRD, time-of-day and day-of-week variation in Internet traffic.

First group. The total of three time series. pAug and pOct time series were obtained from Bellcore Ethernet LAN traces[1] (Leland et al. 1994) with 10ms aggregation level, and were modeled using α-stable FARIMA by Ilow (2000). LBL-TCP-3 time series was provided by the Lawrence Berkeley Laboratory[2] (Paxson and Floyd 1995). Following Harmantzis and Hatzinakos (2001), who modeled this data with α-stable FARIMA, we used a 50ms aggregation. The first time series is of length 2^{18}, the other two of length 2^{17}.

Second group. The total of 28 time series based on four daily measurements starting at 5:00, 10:00, 15:00 and 21:30, for the 7-day period, Sun Apr 06 to Sat Apr 12, 2003. Each time series consists of 2^{19} observations and was obtained with a 10ms aggregation of 2 h-long traffic[3] (Park et al. 2005) from the University of North Carolina at Chapel Hill.

[1] http://ita.ee.lbl.gov/html/contrib/BC.html.

[2] http://ita.ee.lbl.gov/html/contrib/LBL-TCP-3.html.

[3] http://www-dirt.cs.unc.edu/ts.

Table 3 Estimates of the difference parameter d and of the stability index α for the traffic traces

	Memory estimator				
TS	WMLE	LOG LOG	FD	FARIMA	$\hat{\alpha}$
pAug	0.2973	0.1641	0.4129	0.3565	1.8912
pOct	0.3379	0.3430	0.2428	0.2108	2.0754
LBL-TCP-3	0.3828	0.2647	0.2963	0.1718	1.6564
Sun 05:00	0.3077	0.2732	0.1461	0.4999	1.8536
Sun 10:00	0.3569	0.2995	0.1545	0.4997	1.8283
Sun 15:00	0.3674	0.3598	0.1912	0.4985	1.9438
Sun 21:30	0.3700	0.3508	0.1918	0.4999	1.9623
Mon 05:00	0.3134	0.2768	0.1043	0.4744	1.9061
Mon 10:00	0.4099	0.4062	0.2542	0.4999	1.9615
Mon 15:00	0.4097	0.4093	0.1928	0.4995	1.9755
Mon 21:30	0.3631	0.3397	0.1688	0.4998	1.9108
Tue 05:00	0.3145	0.2761	0.1185	0.4719	1.8831
Tue 10:00	0.3936	0.3759	0.1969	0.4999	1.9814
Tue 15:00	0.4207	0.3805	0.2242	0.4999	1.9140
Tue 21:30	0.3800	0.3682	0.1725	0.4998	1.9482
Wed 05:00	0.3524	0.2778	0.1363	0.4994	1.8774
Wed 10:00	0.4265	0.4252	0.2324	0.4999	2.0129
Wed 15:00	0.4188	0.3990	0.2210	0.4998	1.9442
Wed 21:30	0.3912	0.3923	0.1896	0.4998	1.9954
Thu 05:00	0.3626	0.4018	0.1787	0.4997	2.0566
Thu 10:00	0.3900	0.4106	0.2346	0.4997	2.0074
Thu 15:00	0.4313	0.4130	0.2265	0.4999	1.8920
Thu 21:30	0.4052	0.4129	0.2030	0.4998	1.9842
Fri 05:00	0.3610	0.3012	0.1536	0.4998	1.8758
Fri 10:00	0.4266	0.3833	0.2366	0.4999	1.9054
Fri 15:00	0.4417	0.3877	0.2323	0.4999	1.9012
Fri 21:30	0.3949	0.3510	0.1925	0.4999	1.9269
Sat 05:00	0.3515	0.2444	0.1541	0.4919	1.7988
Sat 10:00	0.3811	0.3025	0.2073	0.4999	1.8400
Sat 15:00	0.4053	0.3537	0.2056	0.4999	1.8908
Sat 21:30	0.3894	0.3395	0.1898	0.4999	1.8876

We apply four estimation procedures to the above time series with exactly the same specifications given in Section 4, and report the results in Table 3. The stable index estimates reported in the last column of Table 3 show that in many cases the marginal distributions of the traffic traces have infinite variance.

First group. Memory estimates of Ilow (2000) for pAug and pOct are 0.3014 and 0.3395, respectively, thus they fall very close to the corresponding WMLE. LOG LOG for pOct with value of 0.3430 is also very close, however this is no longer true for pAug, which has the value 0.1641. Harmantzis and Hatzinakos (2001) shortened LBL-TCP-3 time series to about 40000 observations and obtained an estimate of 0.31, which falls between WMLE and LOG LOG estimated on the entire data set.

Second group. The FARIMA estimator performs very poorly, because it hits the upper bound in the estimation procedure. This might be due to the presence of nonstationarities in the data (e.g., trends) or misspecification of the underlying

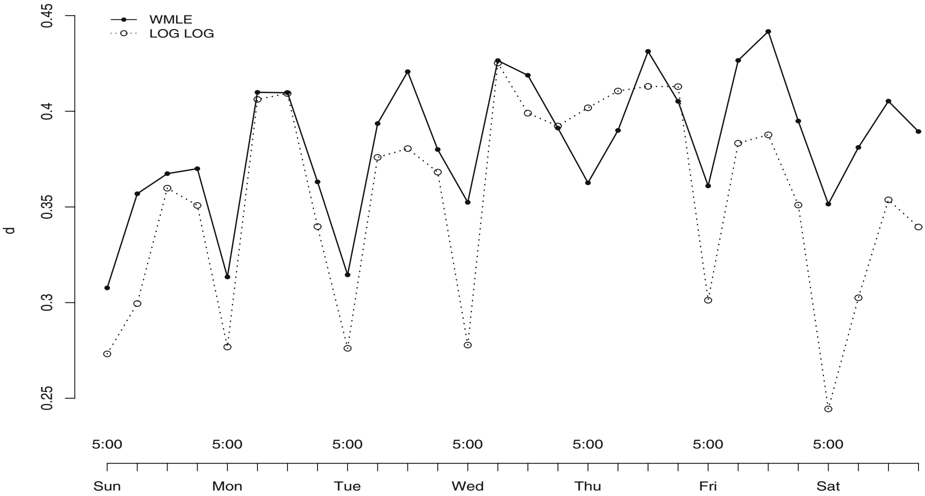

Fig. 7 Estimates of memory parameter for 28 time series (measurements starting at 5:00, 10:00, 15:00 and 21:30) for the 7-day period, Sun Apr 06 to Sat Apr 12, 2003 from the University of North Carolina at Chapel Hill. Each time series consists of 2^{19} observations

model. FD estimates seem to be much lower than the wavelet estimates. WMLE estimates follow a somewhat more regular daily pattern compared to LOG LOG (Fig. 7). There is less memory at 5:00 compared to 10:00 and 15:00, and d at 21:30 slightly decreases. There seems to be an increasing tendency in d from Sun to Fri. The LOG LOG estimates behave similarly to those of the WMLE on Sun and Mon, with slightly smaller values for the former. On Tue, the two tendencies agree, but the difference between the estimated values is bigger than on previous days. The Wed and Thu estimates of the LOG LOG seem to have a different pattern compared to other days of the week, unlike those of the WMLE.

The data examples presented in this section confirm the findings of the simulation study: both WMLE and LOG LOG estimators are more "stable" than the time-domain estimators, which is most likely due to their smaller SD. Figure 7 strongly suggests that the WMLE has a smaller SD than the LOG LOG estimator.

The simulation study of Section 4 was based on popular linear long-memory models. This is justified by our emphasis on the robustness to departures from stationarity rather than detailed traffic data modeling. It would however be of interest to investigate how the estimation methods discussed in this paper perform when applied to structural models for the traffic data, like those recently developed by D'Auria and Resnick (2007a, b). These models have the same large-scale scaling behavior as the linear models studied here and exhibit heavy tails. To model the traffic data, the authors impose a probabilistic structure that is consistent with the properties of the observations. However, simulation of these models is more complex and is the subject of ongoing research.

An alternative extension of the research presented here would be to focus on the estimation of d in certain nonlinear models of asset returns, like the LARCH model of Giraitis et al. (2000). The calculations in Section 3 show that the WMLE is consistent whenever the wavelet coefficients are approximately independent and have either a finite second moment or regularly varying tails, when the second moment is infinite, provided relationship (6) holds. This allows us to conjecture that this method is much more broadly applicable. These issues are currently under investigation.

Appendix: Independence of the Wavelet Coefficients of Heavy-Tailed LRD Processes

In this section, by applying a statistical test of significance, we provide numerical verification of the approximate independence of the DWT coefficients of α-stable FARIMA and the traffic data. Additionally, we consider models of FARIMA with $t(3)$ innovations.

We use the nonparametric difference-sign test from Chapter 1 of Brockwell and Davis (2002), and apply it to the DWT coefficients of the trend-contaminated ($f_1(t)$) models of Section 4 (Tables 1 and 2). We generate $R = 500$ replications of length $N = 10000$ for each model. In Figs. 8–9 we present the empirical sizes of the test.

In general, the empirical sizes of the difference-sign test in Figs. 8–9 are close to the nominal level of 0.05. The empirical sizes in the finite-variance case are closer to the target value than to the α-stable case (Fig. 9 versus Fig. 8). In the latter setting, the DWT seems to decorrelate models with large d better than those with small d (right-bottom panel of Fig. 8 compared to the remaining panels). In the former scenario, the no-memory case ($d = 0$) seems to deviate from the nominal level.

Now we focus on the DWT of the traffic data. In Table 4 we show p-values of the difference-sign test for differenced wavelet coefficients of the three time series of packet counts and of byte counts from Mon Apr 7th, 2003.

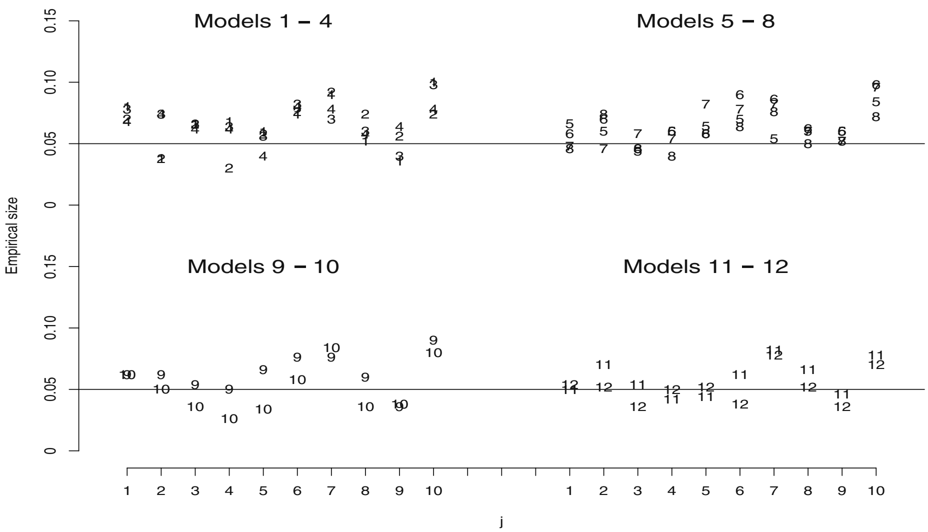

Fig. 8 Empirical sizes of the difference-sign test based on the DWT coefficients of $R = 500$ replications of 12 FARIMA models with α-stable innovations (Table 1). Nominal level: 0.05

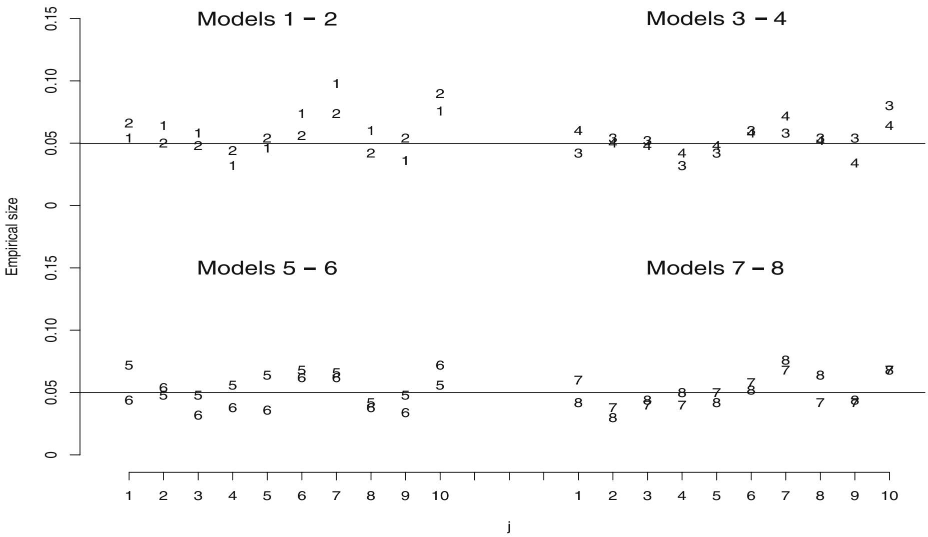

Fig. 9 Empirical sizes of the difference-sign test based on the DWT coefficients of $R = 500$ replications of 8 FARIMA models with $t(3)$ innovations (Table 2). Nominal level: 0.05

P-values for the first three levels are small in most time series and thus their within-level approximate independence is somewhat questionable. Starting from level $j = 5$, we obtain large p-values, that allow us to accept the null hypothesis of independence. Level $j = 4$ is a somewhat border line case, where in five out of seven time series we accept the null hypothesis.

Table 4 P-values for the difference-sign test based on the differenced DWT coefficients of selected time series of packet/byte counts

Level j	TS						
	pAug	pOct	LBL-TCP-3	Mon 05:00	Mon 10:00	Mon 15:00	Mon 21:30
1	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
2	< 0.001	0.2392	0.0013	< 0.001	< 0.001	< 0.001	< 0.001
3	0.0010	0.5698	0.0101	< 0.001	0.0294	< 0.001	< 0.001
4	0.0029	0.9086	0.7594	0.0006	0.3998	0.3787	0.2753
5	0.5658	0.3035	0.3297	0.0987	0.1229	0.8923	0.7047
6	0.0831	0.7592	0.4435	0.4439	0.9390	0.0661	0.3013
7	0.7592	0.1585	0.1037	0.1937	0.9138	0.6649	0.9568
8	0.2781	0.4422	0.5387	0.9389	0.9389	0.1254	0.5917
9	0.8779	0.1904	0.6625	0.5876	0.0008	0.6644	0.5876
10	0.6625	0.1198	0.7557	0.3564	0.0909	0.8779	0.4422
11	0.5338	0.6547	0.6547	0.5127	0.0809	0.1266	0.2752
12	0.3711	0.5127	0.5127	1.0000	0.5338	1.0000	0.2134
13	1.0000	0.3173	0.0455	0.0736	1.0000	0.6547	0.3711
14	0.3173	1.0000	1.0000	0.5127	0.5127	0.1904	1.0000
15	0.0833	NA	NA	0.3173	1.0000	0.3173	1.0000
16	NA	NA	NA	1	1	1	1

Nominal level: 0.05

References

Abry P, Flandrin P, Taqqu MS, Veitch D (2006) Self-similarity and long-range dependence through the wavelet lens. In: Doukhan P, Oppenheim G, Taqqu MS (eds) Theory and applications of long range dependence. Birkhäuser, Boston

Berkes I, Horváth L, Kokoszka P, Shao Q-M (2006) On discriminating between long-range dependence and changes in mean. Ann Stat 34:1140–1165

Brockwell PJ, Davis RA (2002) Introduction to time series and forecasting, 2nd edn. Springer, New York

Craigmile PF, Guttorp P, Percival DB (2005) Wavelet-based parameter estimation for polynomial contaminated fractionally differenced process. IEEE Trans Signal Process 53:3151–3161

D'Auria B, Resnick SI (2007a) Data network models of burstiness. Technical Report ORIE 1428, Cornell University

D'Auria B, Resnick SI (2007b) The influence of dependence on data network models of burstiness. Technical Report ORIE 1449, Cornell University

Davis RA, Resnick SI (1986) Limit theory for the sample covariance and correlation functions of moving averages. Ann Stat 14(2):533–558

Geweke J, Porter-Hudak S (1983) The estimation and application of long memory time series models. J Time Ser Anal 4:221–238

Giraitis L, Robinson P, Surgailis D (2000) A model for long memory conditional heteroskedasticity. Ann Appl Probab 10:1002–1024

Harmantzis F, Hatzinakos D (2001) Network traffic modeling and simulation using stable FARIMA processes. Technical Report, Presented at INFORMS Annual Meeting, Miami

Ilow J (2000) Forecasting network traffic using FARIMA models with heavy-tailed innovations. In: IEEE international conference on acoustics, speech, and signal processing, 2000, vol 6. IEEE, Piscataway, pp 3814–3817

Jach A, Kokoszka P (2008) Wavelet-domain test for long-range dependence in the presence of a trend. Statistics 42:101–113

Karasaridis A, Hatzinakos D (2001) Network heavy traffic modeling using α-stable self-similar processes. IEEE Trans Commun 49(7):1203–1214

Kogon SM, Manolakis DG (1995) Linear parametric models for signals with long-range dependence and infinite variance. In: IEEE international conference on acoustics, speech, and signal processing, 2000 vol 3. IEEE, Piscataway, pp 1597–1600

Kokoszka PS, Taqqu MS (1996) Parameter estimation for infinite variance fractional ARIMA. Ann Stat 24:1880–1913

Lee IWC, Fapojuwo AO (2005) Estimating heavy tails in long-range dependent wireless traffic. IEEE Veh Technol Conf 4:2132–2136

Leland WE, Taqqu MS, Willinger W, Wilson DV (1994) On the self-similar nature of Ethernet traffic (extended version). IEEE/ACM Trans Netw 2:1–15

Lobato I, Velasco C (2000) Long memory in stock market trading volume. J Bus Econ Stat 18:410–427

Moulines E, Roueff E, Taqqu M (2008) A wavelet Whittle estimator of the memory parameter of a nonstationary Gaussian time series. Ann Stat 36:1925–1956

Park C, Hernández-Campos F, Marron J, Smith F (2005) Long-range dependence in a changing traffic mix. Comput Networks 48:401–422

Paxson V, Floyd S (1995) Wide area traffic: the failure of Poisson modelling. IEEE/ACM Trans Netw 3:226–224

Percival DB, Walden AT (2000) Wavelet methods for time series analysis. Cambridge University Press, Cambridge

Pipiras V, Taqqu MS, Abry P (2007) Bounds for the covariance of functions of infinite variance stable random variables with applications to central limit theorems and wavelet-based estimation. Bernoulli 13:1091–1123

Resnick SI (1997) Heavy tail modeling and teletraffic data. Ann Stat 25:1805–1849

Robinson PM (1995a) Log-periodogram regression of time series with long range dependence. Ann Stat 23:1048–1072

Robinson PM (1995b) Gaussian semiparametric estimation of long range dependence. Ann Stat 23:1630–1661

Samorodnitsky G, Taqqu MS (1994) Stable non-gaussian random processes: stochastic models with infinite variance. Chapman & Hall, New York

Stoev S, Michailidis G, Taqqu M (2006) Estimating heavy-tail exponents through max self-similarity. Technical Report TR445, University of Michigan, 2006. http://www.stat.lsa.umich.edu/~sstoev/max-spectrum-1-corrected.pdf

Stoev S, Taqqu M (2004) Simulation methods for linear fractional stable motion and FARIMA using the fast fourier transform. Fractals 12:95–121

Stoev S, Taqqu MS (2005) Asymptotic self-similarity and wavelet estimation for long-range dependent FARIMA time series with stable innovations. J Time Ser Anal 26:211–249

Taqqu MS, Teverovsky V (1998) On estimating the intensity of long-range dependence in finite and infinite variance series. In: Adler R, Feldman R, Taqqu MS (eds) A practical guide to heavy tails: statistical techniques for analyzing heavy tailed distributions. Birkhäuser, Boston, pp 177–217

Teyssiere G, Abry P (2006) Wavelet analysis of nonlinear long-range dependent processes. Applications to financial time series. In: Teyssiere G, Kirman A (eds) Long memory in economics. Springer, New York, pp 173–238

Velasco C, Robinson P (2000) Whittle pseudo-maximum likelihood estimation for nonstationary time series. J Am Stat Assoc 95:1229–1243