Subsampling the mean of heavy-tailed dependent observations

Piotr Kokoszka*
Mathematics and Statistics
Utah State University
Logan, Utah 84322-3900, U.S.A.

Michael Wolf†
Dept. of Economics and Business
Universitat Pompeu Fabra
08005 Barcelona, Spain

February 2002

*Partially supported by USU New Faculty grant and NATO grant PST.CLG.977607.
†Supported by DGES grant BEC2001-1270.

Abstract

We establish the validity of subsampling confidence intervals for the mean of a dependent series with heavy-tailed marginal distributions. Using point process theory, we study both linear and nonlinear GARCH-like time series models. We propose a data-dependent method for the optimal block size selection and investigate its performance by means of a simulation study.

JEL CLASSIFICATION NOS: C10, C14, C32.

KEYWORDS: GARCH, Heavy tails, Linear time series, Subsampling.

1 Introduction

Estimation of the mean is often the first step in an analysis of a stationary time series. If the observations can be assumed to be generated by a stationary model with finite variance, there is a well-known asymptotic theory for the sample mean, see e.g. Section 7.1 of Brockwell and Davis (1991) and a large body of research devoted to the estimation of the asymptotic variance.

In this paper we assume that the observations follow the model $X_t = \mu + Y_t$, where $\{Y_t\}$ is a zero mean stationary time series with heavy tailed univariate marginal distributions. We assume that these distributions are regularly varying with index κ satisfying $1 < \kappa < 2$, so that the mean exists but the variance is infinite. Linear processes with infinite variance heavy tailed distributions have been studied by Cline and Brockwell (1985), Mikosch *et al.* (1995), Anderson and Meerschaert (1997) and Kokoszka and Taqqu (1994, 1996, 2001), among others. It has recently been established that the popular GARCH processes have regularly varying marginal distibution which may exhibit infinite variance for some choices of parameters, see Basrak, Davis and Mikosch (2002a, 2002b) and the asymptotic theory for sample autocovariances and extrema for such processes has been developed, see Davis and Mikosch (1998), Mikosch and Stărică (2000).

We investigate the validity of the subsampling confidence intervals for μ based on the statistic

$$T_n = n^{1/2}\frac{\bar{X}_n - \mu}{\hat{\sigma}_n}, \tag{1.1}$$

where

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{t=1}^{n}(X_t - \bar{X}_n)^2. \tag{1.2}$$

Thus we approximate the sampling distribution of T_n by

$$L_{n,b}(x) = \frac{1}{n-b+1}\sum_{t=1}^{n-b+1}\mathbf{1}\left\{\frac{b^{1/2}(\bar{X}_{n,b,t} - \bar{X}_n)}{\hat{\sigma}_{n,b,t}} \le x\right\}. \tag{1.3}$$

We refer to Politis *et al.* (1999) for a systematic account of the subsampling methodology. A theoretical justification for the subsampling method considered in this paper is based on Theorem 1.1 which is stated below. It is almost identical to Theorem 11.3.1 of Politis *et al.* (1999), the only difference being that we do not assume independent observations. For the sake of completeness we state here this result and outline its proof.

Suppose we have observed a sample $X_1, \ldots, X_n$ and $\hat{\theta}_n$ is an estimator of θ and J_n is the sampling distribution of $\tau_n(\hat{\theta}_n - \theta)/\hat{\sigma}_n$, where $\hat{\sigma}_n > 0$. Set also

$$J_n(x) = P\left\{\tau_n(\hat{\theta}_n - \theta)/\hat{\sigma}_n \le x\right\}. \tag{1.4}$$

Assumption 1.1 There are nondegenerate distributions J, V, W, such that W has no mass at the origin, an positive sequences $\{t_n\}$ and $\{u_n\}$ such that $\tau_n = t_n/u_n$ and

$$J_n \xrightarrow{d} J; \tag{1.5}$$

$$t_n(\hat{\theta}_n - \theta) \xrightarrow{d} V; \tag{1.6}$$

$$u_n\hat{\sigma}_n \xrightarrow{d} W. \tag{1.7}$$

Consider the subsampling approximation to $J_n(x)$ given by

$$L_{n,b}(x) = \frac{1}{n-b+1} \sum_{t=1}^{n-b+1} \mathbf{1}\left\{ \frac{\tau_b(\hat{\theta}_{n,b,t} - \hat{\theta}_n)}{\hat{\sigma}_{n,b,t}} \le x \right\}, \tag{1.8}$$

where $\hat{\theta}_{n,b,t}, \hat{\sigma}_{n,b,t}$ are computed from the observations $X_t, X_{t+1}, \ldots, X_{t+b-1}$.

The definition of strong mixing is recalled in Section 2.

THEOREM 1.1 *Suppose the process $\{X_t\}$ is strong mixing, Assumption 1.1 holds, and*

$$b \to \infty, \quad \frac{b}{n} \to 0, \quad \frac{\tau_b}{\tau_n} \to 0, \quad \frac{t_b}{t_n} \to 0..$$

Then, the following conclusions hold:

(i) If x is a continuity point of $J(\cdot)$, then $L_{n,b}(x) \xrightarrow{P} J(x)$.

(ii) If $J(\cdot)$ is continuous, then $\sup_x |L_{n,b}(x) - J(x)| \xrightarrow{P} 0$.

(iii) Denote

$$c_{n,b}(1-\alpha) = \inf\{x : L_{n,b}(x) \ge 1-\alpha\};$$

$$c(1-\alpha) = \inf\{x : J(x) \ge 1-\alpha\}.$$

If $J(\cdot)$ is continuous at $c(1-\alpha)$, then

$$P\left\{\tau_n(\hat{\theta}_n - \theta)/\hat{\sigma}_n \le c_{n,b}(1-\alpha)\right\} \to 1-\alpha,$$

that is, the subsampling confidence intervals yield asymptotically correct coverage probability.

The proof of Theorem 1.1 is the same as of Theorem 11.3.1 in Politis *et al.* (1999), except that to show the convergence

$$\frac{1}{n-b+1} \sum_{t=1}^{n-b+1} \mathbf{1}\left\{ \frac{\tau_b(\hat{\theta}_{n,b,t} - \theta)}{\hat{\sigma}_{n,b,t}} \le x \right\} \xrightarrow{P} J(x)$$

one must follow the argument in the proof of Theorem 3.2.1. of Politis *et al.* (1999), rather then use an argument for independent observations.

The difficulty of applying Theorem 1.1 lies in verifying Assumption 1.1 for a specific class of time series of interest. In Section 2 we study two popular classes of dependent processes which exhibit both dependence and heavy tails, the case of independent observations was studied in Chapter 11 of Politis *et al.* (1999). Both classes are defined in a broad nonparametric setting and are shown to contain popular parametric models. The first class consists of moving average models with heavy-tailed innovations and was independently investigated by McElroy and Politis (2002). Their method of proof relies on representing the partial sum of observations as a multiple of the partial sum of the noise plus a small remainder term. We use point process techniques which yield somewhat shorter arguments. These techniques are also useful for the second class of models which includes nonlinear time series like GARCH. Thus the point process approach may be more widely applicable in other contexts as well. Section 3 focuses on the practical implementation of the subsampling method and illustrates its applicability by means of a simulation study. The critical issue is the choice of the block size b. McElroy and Politis (2002), who report only results for several fixed choices of b, stress the need for finding a good procedure for determining b. We propose a fairly general data-dependent approach. Mathematical proofs are collected in the Appendix.

Remark 1.1 The approximation (1.8) allows for the construction of one-sided or equal-tailed two-sided confidence intervals for μ. As an alternative, two-sided symmetric confidence intervals could be constructed by estimating the two-sided distribution function

$$J_{n,|\cdot|}(x) = P\left\{\tau_n|\hat{\theta}_n - \theta|/\hat{\sigma}_n \leq x\right\}. \tag{1.9}$$

The according subsampling approximation is given by

$$L_{n,b,|\cdot|}(x) = \frac{1}{n-b+1}\sum_{t=1}^{n-b+1} \mathbf{1}\left\{\frac{b^{1/2}|\bar{X}_{n,b,t} - \bar{X}_n|}{\hat{\sigma}_{n,b,t}} \leq x\right\}. \tag{1.10}$$

The asymptotic validity of this approach follows immediately from the validity of (1.8) and the continuous mapping theorem.

2 Main results

In this section we verify that the assumptions of Theorem 1.1 are satisfied by two commonly used classes of time series. The first class consists of moving averages with innovations which are in the domain of attraction of a stable law with index $1 < \kappa < 2$. Such time series arise for example in modelling teletraffic and server workload data, see e.g. Resnick (1997). The second class includes GARCH-like processes which, unlike the processes from the previous class, do not exhibit "linear" dependence, but possess "dependence in absolute values". Such time series are commonly used in modelling financial and economic data; for example, see Gouriéroux (1997).

In Theorem 1.1 we assume that the time series under consideration is strongly mixing. We recall here the appropriate definitions and some related facts. Suppose $\{X_t,\ t \in \mathbf{Z}\}$ is a stationary random sequence. The *mixing rate function* m_k of $\{X_t\}$ is defined as

$$m_k = \sup\left\{|P(A \cap B) - P(A)P(B)|,\ A \in \sigma(X_s, s \leq 0),\ B \in \sigma(X_s, s > k)\right\}, \tag{2.1}$$

with the σ-algebras in (2.1) defined in the usual way. (The m_k in (2.1) are usually denoted α_k but we want to avoid confusion with the coefficients in the GARCH specification (2.14).) If $m_k \to 0$ as $k \to \infty$, the sequence $\{X_t\}$ is said to be *strong mixing* or α-mixing, and if there are constants $K > 0$ and $0 < a < 1$ such that $m_k < Ka^k$, it is said to be *strongly mixing with geometric rate*. We refer to Doukhan (1994) or Bradley (1986) for systematic accounts of mixing conditions. .

2.1 Heavy-tailed moving averages

We consider moving averages of the form

$$Y_t = \sum_{j=0}^{\infty} c_j Z_{t-j} \tag{2.2}$$

with the weights c_j satisfying

$$\sum_{j=0}^{\infty} |c_j| < \infty. \tag{2.3}$$

This model nests causal ARMA(p,q) and AR(∞) specifications. The iid innovations Z_t are assumed to be mean zero and in the domain of attraction of a stable law. Thus the Z_t satisfy the following assumption:

Assumption 2.1 There is $1 < \kappa < 2$, a slowly varying function L and nonnegative constants a and b satisfying $a + b = 1$ such that

$$P(|Z_t| > x) = x^{-\kappa} L(x), \tag{2.4}$$

$$\frac{P(Z_t > x)}{P(|Z_t| > x)} \to a, \quad \frac{P(Z_t < -x)}{P(|Z_t| > x)} \to b. \tag{2.5}$$

In addition, $EZ_t = 0$.

The moving averages of the form (2.2) have been studied, among others, by Davis and Resnick(1985, 1986) and Mikosch *et al.* (1995).

The moving average (2.2) has the same tail behavior as the innovations Z_t. More precisely, if (2.4), (2.5) hold, then

$$\lim_{x\to\infty} \frac{P(|\sum_{j=0}^{\infty} c_j Z_{t-j}| > x)}{P(|Z_t| > x)} = \sum_{j=0}^{\infty} |c_j|^{\kappa}. \tag{2.6}$$

Relation (2.6) was established by Cline (1983).

In order to ensure that the moving average (2.2) is strong mixing, we must assume that the innovations Z_t have a density. It follows from Gorodetskii (1977) that (2.2) is strong mixing under the following assumption:

Assumption 2.2 The density, f, of the Z_t and the weights c_j satisfy: (i) There is a constant C such that $\int_{-\infty}^{\infty} |f(x+y) - f(x)| dx \le C|y|$; (ii) $C(z) = \sum_{j=0}^{\infty} c_j z^j \neq 0$ for $|z| \le 1$; (iii) $\sum_{i=0}^{\infty} \left(\sum_{j=i}^{\infty} |c_j|\right)^{1/2} < \infty$.

Proposition 2.1 *Suppose Assumptions 2.1 and 2.2 hold. Then there is a constant M such that for the moving average* (2.2),

$$m_k \le M \sum_{i=k}^{\infty} \left(\sum_{j=i}^{\infty} |c_j|\right)^{1/2}.$$

Remark 2.1 If the Z_t have a stable distribution, then they have a density which satisfies condition (i) of Assumption 2.2. To see this, recall that in this case the characteristic function of Z_t is $\phi(\theta) = \exp\{-\sigma^{\kappa}|\theta|^{\kappa}(1 - i\beta(\text{sign}\theta)\tan(\pi\kappa/2)\}$, see Definition 1.1.6 in Samorodnitsky and Taqqu (1994) for the details. Thus, differentiating the inversion formula,

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ix\theta} \phi(\theta) d\theta$$

we obtain, for some constant C, $|f'(x)| \le C \int_0^{\infty} u \exp(-u^{\kappa}) du < \infty$.

Condition (ii) in Assumption 2.2 is satisfied, for example, by causal and invertible ARMA processes, see e.g. Section 2 in Kokoszka (1996). Condition (iii) is implied, for instance, by $\sum_{j=1}^{\infty} j|c_j|^{1/2} < \infty$. Notice also that condition (iii) implies (2.3).

In order to verify the assumptions of Theorem 1.1, we also need the following result:

Theorem 2.1 *Suppose the Y_t are defined by* (2.2). *If* (2.3) *and Assumption 2.1 hold, then*

$$\left(\frac{1}{n^{1/\kappa} L_0(n)} \sum_{t=1}^{n} Y_t, \quad \frac{1}{n^{2/\kappa} L_0^2(n)} \sum_{t=1}^{n} Y_t^2\right) \xrightarrow{d} \left((\sum_{j=0}^{\infty} c_j)V, \quad (\sum_{j=0}^{\infty} c_j^2)W^2\right),$$

where V is mean zero κ-stable with the skewness parameter $a-b$ and W^2 is positive $\kappa/2$-stable, and where L_0 is a slowly varying function.

The proof of Theorem 2.1, which is presented in the Appendix, relies on an analogous result for the innovation sequence established by Logan *et al.* (1973) and the ideas used in the proof of Theorem 4.1 in Davis and Resnick (1985) who showed componentwise convergence using very different methods for each component.

With Theorem 2.1 in hand, it is easy to check that the assumptions of Theorem 1.1 hold with

$$\theta = \mu,\ \hat{\theta}_n = \bar{X}_n,\ t_n = \frac{n^{1-1/\kappa}}{L_0(n)},\ u_n = \frac{n^{1/2-1/\kappa}}{L_0(n)}. \tag{2.7}$$

Indeed, (1.6) holds because

$$t_n(\bar{X}_n - \mu) = \frac{t_n}{n}\sum_{t=1}^{n} Y_t = \frac{1}{n^{1/\kappa}L_0(n)}\sum_{t=1}^{n} Y_t \xrightarrow{d} (\sum_{j=0}^{\infty} c_j)V.$$

Since $u_n\bar{Y}_n = n^{-1/2}(n^{1/\kappa}L_0(n))^{-1}\sum_{t=1}^{n} Y_t \xrightarrow{P} 0$, we have

$$u_n^2\hat{\sigma}_n^2 = \frac{u_n^2}{n}\sum_{t=1}^{n} Y_t^2 - (u_n\bar{Y}_n)^2 \sim \frac{1}{n^{2/\kappa}L_0^2(n)}\sum_{t=1}^{n} Y_t^2 \xrightarrow{d} (\sum_{j=0}^{\infty} c_j^2)W^2,$$

so (1.7) also holds. Relation (1.5) follows now from the joint convergence in Theorem 2.1.

We have thus established the following result:

THEOREM 2.2 *If Assumptions 2.1 and 2.2 are satisfied, then for the moving average* (2.2) *the conclusions of Theorem 1.1 hold with* $\tau_n = n^{1/2}, \theta = \mu,\ \hat{\theta}_n = \bar{X}_n, \hat{\sigma}_n$ *defined in* (1.2) *and* J *being the distribution of* $[(\sum_{j=0}^{\infty} c_j)V]/[(\sum_{j=0}^{\infty} c_j^2)^{1/2}W]$ *with* V *and* W *defined in Theorem 2.1.*

2.2 GARCH-type processes

In this section we consider a nonparametric specification intended to model a time series which exhibits no "correlation" but has a significant "correlation in absolute values". As mentioned above, series with such characteristics arise in finance and economics. Condition (2.8), in which $\xrightarrow{v}$ denotes vague convergence, together with (2.9) is equivalent to the requirement that the one-dimensional marginal distributions are in the domain of attraction of an κ-stable law, see e.g. Meerschaert and Scheffler (2001), Proposition 6.1.37. If we assume, as we do in this paper, that a stochastic process has infinite variance, we cannot assume that the observations are uncorrelated because the covariances do not exist. Instead we assume condition (2.10) below which means that truncated variables are uncorrelated. Other assumptions are collected in Assumption 2.3. We have found it convenient to use the theory of point processes, as it has been successfully applied in the context of GARCH processes in Davis and Mikosch (1998) and Mikosch and Stărică (2000). Our approach draws heavily on Davis and Hsing (1995) and we refer the reader to this paper for further details. In particular, condition (2.11) is implied by a very weak form of mixing, which in turn is implied by strong mixing which is necessary for the validity of the subsampling method. Our proofs rely, however, only on condition (2.11) and the other conditions in Assumption 2.3.

ASSUMPTION 2.3 The sequence $\{Y_t\}$ is strictly stationary with symmetric univariate marginal distributions which satisfy

$$nP(Y_1/a_n \in \cdot) \xrightarrow{v} \mu(\cdot), \tag{2.8}$$

with the a_n defined by $nP(|Y_1| > a_n) \to 1$ and the measure μ given by

$$2\mu(dx) = \kappa|x|^{-\kappa-1}\mathbf{1}\{x < 0\}dx + \kappa x^{-\kappa-1}\mathbf{1}\{x > 0\}dx. \tag{2.9}$$

Moreover we assume that for every $y >$ and $t \neq s$

$$E[Y_t \mathbf{1}\{|Y_t| \leq y\} Y_s \mathbf{1}\{|Y_s| \leq y\}] = 0 \tag{2.10}$$

and

$$\sum_{t=1}^{n} \delta_{Y_t/a_n} \xrightarrow{d} \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \delta_{P_i Q_{ij}}, \tag{2.11}$$

with the limiting point process as in Theorem 2.3 and Corollary 2.4 of Davis and Hsing (1995).

Remark 2.2 We assume a symmetric distribution to avoid lengthy mathematical arguments and notation. The case of a nonsymmetric distribution could be handled similarly as in Davis and Hsing (1995) by introducing appropriate centering constants.

Theorem 2.3 *If Assumption 2.3 holds, then*

$$\left(\frac{1}{a_n} \sum_{t=1}^{n} Y_t, \frac{1}{a_n^2} \sum_{t=1}^{n} Y_t^2 \right) \xrightarrow{d} (S_1, S_2), \tag{2.12}$$

where S_1 is the distributional limit, as $\varepsilon \to 0$, of $\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} P_i Q_{ij} \mathbf{1}\{P_i |Q_{ij}| > \varepsilon\}$ (the existence of this limit was established in Theorem 3.1 of Davis and Hsing (1995)) and W^2 is equal in distribution to $\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} P_i^2 Q_{ij}^2$. The random variable S_1 is symmetric κ-stable and S_2 is positive $\kappa/2$-stable.

As verified in Section 2.1, the asymptotic validity of the subsampling confidence intervals for the mean follows from Theorem 2.3 under the additional assumption that the process is strong mixing.

In the remainder of this section we focus on the popular class of GARCH processes. The observations $Y_1, \ldots, Y_n$ are said to follow a GARCH(p, q) model if they satisfy the equations:

$$Y_t = \sigma_t \varepsilon_t, \tag{2.13}$$

$$\sigma_t^2 = \omega + \sum_{j=1}^{p} \alpha_j Y_{t-j}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2. \tag{2.14}$$

The *innovations* ε_k in (2.13) are iid and $\omega, \alpha_j, \beta_j$ are nonnegative parameters.

Several authors formulated conditions under which a GARCH process is strong mixing with geometric rate, see Boussama (2000), Maercker and Moser (1999) Basrak *et al.* (2002b), Carasco and Chen (2002). These conditions are not rescritive but are difficult to verify as they are often formulated in terms of abstract quantities which are very difficult to estimate from the available observations. Basrak, Davis and Mikosch (2002a, 2002b) showed that under similar conditions the finite dimensional distributions of of GARCH processes are multivariate regularly varying, a property which implies pareto-like tails considered in this paper. The special cases of ARCH(1) and GARCH(1,1) are considered, respectively, in Davis and Mikosch (1998) and Mikosch and Stărică (2000). Finally, notice that if the innovations ε_t in (2.13) are symmetric, then (2.10) holds.

We conclude this section by noting that for GARCH(1,1) the tail index κ can be found as the solution of the equation $E(\alpha_1 \varepsilon_1^2 + \beta_1)^{\kappa/2} = 1$, see Theorem 2.1 in Mikosch and Stărică (2000). This equation can be solved analytically only in a few special cases; in general simulations must be used.

3 Choice of the block size and a simulation study

3.1 Choice of the block size

The application of the subsampling method requires a choice of the block size b; the problem is very similar to the choice of the bandwidth in applying smoothing or kernel methods. Unfortunately, the asymptotic requirements $b \to \infty$ and $b/n \to \infty$ as $n \to \infty$ give little guidance when faced with a finite sample. Instead, we propose to exploit the semi-parametric nature of models treated in this paper to estimate a 'good' block size in practice.

Our goal is to construct a $1-\alpha$ confidence interval for the mean μ, but the methodology described below can be adapted to other parameters of interest as well. In finite samples, a subsampling interval will typically not exhibit coverage probability exactly equal to $1-\alpha$; moreover, the actual coverage probability generally depends on the block size b. Indeed, one can think of the actual coverage level $1-\lambda$ of a subsampling confidence interval as a function of the block size b, conditional on the underlying probability mechanism P—that is, the fully specified moving average or GARCH-type model in our application—and the nominal confidence level $1-\alpha$. The idea is now to adjust the 'input' b in order to obtain the actual coverage level close to the nominal one. Hence, one can consider the block size calibration function $g : b \to 1-\lambda$. If $g(\cdot)$ were known, one could construct an 'optimal' confidence interval by finding $\tilde{b}$ that minimizes $|g(b) - (1-\alpha)|$ and use $\tilde{b}$ as the block size; note that $|g(b) - (1-\alpha)| = 0$ may not always have a solution.

Of course, the function $g(\cdot)$ depends on the underlying probability mechanism P and is therefore unknown. We now propose a semi-parametric bootstrap method to estimate it. The idea is that in principle we could simulate $g(\cdot)$ if P were known by generating data of size n according to P and computing subsampling confidence intervals for θ for a number of different block sizes b. This process is then repeated many times and for a given b one estimates $g(b)$ as the fraction of the corresponding intervals that contain the true parameter. The method we propose is identical except that P is replaced by an estimate $\hat{P}_n$ whose mean is equal to $\bar{X}_n$, the sample mean of the original data.

We suggest to make use of the assumed model class in the estimation of $\hat{P}_n$. For example, if a general moving average process is assumed, one would start by determining the order of the process by a model selection criterion that is robust against infinite variance; for example, see Bhansali (1988). (Note that even if the true process has order infinity, for a fixed sample size n, a finite-order model should serve as a good approximation.) Say the so-estimated order is $\hat{q}$. Fitting an MA($\hat{q}$) model to the zero mean data $\hat{Y}_t = X_t - \bar{X}_n$, say by the Whittle estimator technique of Mikosch *et al.* (1995), then yields estimated coefficients $\hat{c}_0, \ldots, \hat{c}_{\hat{q}}$ and centered residuals $\hat{Z}_{\hat{q}+1}, \ldots, \hat{Z}_n$. We can now define $\hat{P}_n$ as the law of the following sequence $X_1^*, \ldots, X_n^*$ (and the definition makes it obvious how to generate such a sequence in practice):

- Draw $Z^*_{-\hat{q}+1}, \ldots, Z^*_n$ iid from the empirical distribution of the centered $\hat{Z}_{\hat{q}+1}, \ldots, \hat{Z}_n$.
- Let $Y_t^* = \sum_{j=0}^{\hat{q}} \hat{c}_j Z^*_{t-j}$, for $t = 1, \ldots, n$.
- Let $X_t^* = \bar{X}_n + Y_t^*$, for $t = 1, \ldots, n$.

Of course if a finite ARMA(p,q) model of known order is assumed, this model should be used instead; the modifications are obvious.

To give another example, if a GARCH(1,1) model is assumed, one would start again by computing the $\hat{Y}_t = X_t - \bar{X}_n$. Then, the model parameters ω, α_1, and β_1 are estimated from the $\hat{Y}_t$ by quasi maximum likelihood, assuming conditional normality. Using the estimated parameters, and resampling from the centered and normalized residuals, one then builds up the Y_t^* sequence. And in the last step, the sample mean $\bar{X}_n$ of the original data is added to

them in order to arrive at the X_t^* sequence. Again, the probability mechanism that gives rise to this sequence is $\hat{P}_n$.

The following algorithm describes how to pick the block size b in practice.

Algorithm 3.1 (Choice of the Block Size)

1. Fix a selection of reasonable block sizes b between limits b_{low} and b_{up}.
2. Generate K pseudo sequences $X_{k1}^*, \ldots, X_{kn}^*$, $k = 1, \ldots, K$, from an estimated model $\hat{P}_n$. For each sequence, $k = 1, \ldots, K$, and for each b, compute a subsampling confidence interval $\mathrm{CI}_{k,b}$ for μ.
3. Compute $\hat{g}(b) = \#\{\bar{X}_n \in \mathrm{CI}_{k,b}\}/K$.
4. Find the value $\tilde{b}$ that minimizes $|\hat{g}(b) - (1-\alpha)|$.

Remark 3.1 There is no universal good block size. For each combination of confidence level and confidence interval type (one-sided, equal-tailed, or symmetric) a separate block size should be computed.

Remark 3.2 Algorithm 3.1 is by an order of magnitude more expensive than the computation of the final subsampling interval once the block size has been determined. While it is advisable to choose the selection of candidate block sizes in Step 2 as fine as possible (ideally, include every integer between b_{low} and b_{up}), this may computationally not be feasible, especially in simulation studies. In those instances, a coarse grid should be employed.

3.2 Simulation Study

We now present a small simulation study. Two data generating processes (DGP) are considered. The first DGP is an AR(1) model with stable innovations with index κ[1] By choosing positive values for ω, α_1 and β_1 such that the equation $E(\alpha_1\varepsilon_1^2+\beta_1)^{\kappa/2} = 1$ has a solution $1 < \kappa < 2$, we can generate GARCH(1,1) time series with finite mean but infinite variance. We also consider the IGARCH model defined by the requirement that $\alpha_1 + \beta_1 = 1$ because it is often used in practice; for example, see Engle and Bollerslev (1986). This is a model with infinite variance but $\kappa = 2$, so it is not covered by the theory developed in the present paper. We must therefore rely solely on simulations to assess the performance of the subsampling method. A theoretical investigation of this case would be difficult.

Without loss of generality, the true mean μ is always set equal to zero. Of interest is the coverage probability of two-sided subsampling confidence intervals with nominal coverage levels 95% and 90%. We include two types of intervals in the study, the two-sided equal-tailed interval and the two-sided symmetric interval. The sample sizes considered are $n = 200$ and $n = 500$. To keep the computational cost at a reasonable level in this simulation study, we choose $K = 300$ in Algorithm 3.1 and select a very coarse grid of 3 input block sizes. (Note that when applying the method to a real life data set one should choose $K = 1000$ and a finer grid.) As outlined above, we resample from the (standardized and) centered residuals and do not use the knowledge of their distributional form.

[1]The stable innovations were generated using software of John Nolan; see the web page `http://academic2.american.edu/~jpnolan/`. as in McElroy and Politis (2002), who only present results for fixed block sizes:

$$Y_t = \phi Y_{t-1} + Z_t.$$

The second DGP is a GARCH(1,1) model with normal innovations.

The results, based on 2,000 replications, are presented in Tables 1–6. It can be seen that the 'optimal' fixed block size depends on the sample size, the DGP, the parameter values, the interval type, and, to a lesser extent, the confidence level. Hence, it is very important to have a reliable data-dependent method to select a good block size in practice. It is seen that our method works very well when κ is close to 2 but that the intervals overcover when κ is close to 1. In general, the coverage properties are better for the symmetric interval in agreement with previous simulation studies for other parameters of interest and/or probability mechanisms; for example, see Chapters 9, 11, and 12 of Politis *et al.* (1999).

4 Appendix

Proof of Theorem 2.1. In this proof we denote for brevity

$$a_n = n^{1/\kappa} L_0(n).$$

According to the discussion on p. 789 of Logan *et al.* (1973)

$$\left(a_n^{-1} \sum_{t=1}^{n} Z_t,\ a_n^{-2} \sum_{t=1}^{n} Z_t^2 \right) \xrightarrow{d} \left(V,\ W^2 \right). \tag{4.1}$$

We first prove the result for finite moving averages

$$Y_t^{(m)} = \sum_{j=0}^{m} c_j Z_{t-j}.$$

Consider the random vector

$$\left(a_n^{-1} \sum_{t=1}^{n} Z_t, a_n^{-1} \sum_{t=1}^{n} Z_{t-1}, \ldots, a_n^{-1} \sum_{t=1}^{n} Z_{t-m}, a_n^{-2} \sum_{t=1}^{n} Z_t^2, a_n^{-2} \sum_{t=1}^{n} Z_{t-1}^2, \ldots, a_n^{-2} \sum_{t=1}^{n} Z_{t-m}^2 \right). \tag{4.2}$$

By an elementary argument, see p. 190 of Davis and Resnick (1985), it follows from (4.1) that vector (4.2) converges is distribution to the random vector

$$(V, V, \ldots, V, W^2, W^2, \ldots, W^2). \tag{4.3}$$

Let λ and μ be arbitrary real numbers. Multiplying (4.2) and (4.3) on the right by the vector

$$(\lambda c_0, \lambda c_1, \ldots, \lambda c_m, \mu c_0^2, \mu c_1^2, \ldots, \mu c_m^2)^T$$

and using the continuous mapping theorem, we obtain

$$\lambda a_n^{-1} \sum_{t=1}^{n} Y_t^{(m)} + \mu a_n^{-2} \sum_{t=1}^{n} \sum_{j=0}^{m} c_j^2 Z_{t-j}^2 \xrightarrow{d} \lambda \left(\sum_{j=0}^{m} c_j \right) V + \mu \left(\sum_{j=0}^{m} c_j^2 \right) W^2.$$

Thus in order to claim that

$$\left(a_n^{-1} \sum_{t=1}^{n} Y_t^{(m)},\quad a_n^{-2} \sum_{t=1}^{n} [Y_t^{(m)}]^2 \right) \xrightarrow{d} \left((\sum_{j=0}^{m} c_j) V,\quad (\sum_{j=0}^{m} c_j^2) W^2 \right) \tag{4.4}$$

we must show that

$$a_n^{-2} \sum_{t=1}^{n} \sum_{1 \le j < k \le m} c_j c_k Z_{t-j} Z_{t-k} \xrightarrow{P} 0. \tag{4.5}$$

Using $E|Z| < \infty$ and (2.3), we obtain

$$a_n^{-2} E \left| \sum_{t=1}^{n} \sum_{1 \le j < k \le m} c_j c_k Z_{t-j} Z_{t-k} \right|$$

$$\le a_n^{-2} \sum_{1 \le j < k \le m} |c_j||c_k| E|Z_{t-j}| E|Z_{t-k}| = O(n a_n^{-2}) = o(1).$$

Consequently (4.5) holds, and the verification of (4.4) is complete.

In light of Theorem 3.2 of Billingsley (1999), to complete the proof of Theorem 2.1, it suffices to verify that for any $r > 0$

$$\lim_{m \to \infty} \limsup_{n \to \infty} P \left(a_n^{-1} \left| \sum_{t=1}^{n} [Y_t - Y_t^{(m)}] \right| > r \right) = 0 \tag{4.6}$$

and

$$\lim_{m \to \infty} \limsup_{n \to \infty} P \left(a_n^{-2} \left| \sum_{t=1}^{n} [Y_t^2 - (Y_t^{(m)})^2] \right| > r \right) = 0. \tag{4.7}$$

Relation (4.6) is verified on pp. 190-191 of Davis and Resnick (1985). To verify (4.7) observe that by defining $c_j^{(m)} = 2c_j$ if $j \le m$, $c_j^{(m)} = c_j$ if $j > m$, we have

$$Y_t^2 - (Y_t^{(m)})^2 = \left[\sum_{k=0}^{\infty} c_k^{(m)} Z_{t-k} \right] \left[\sum_{j=m+1}^{\infty} c_j Z_{t-j} \right] = \sum_{j=m+1}^{\infty} c_j^2 Z_{t-j}^2 + \sum_{k=0}^{\infty} \sum_{m < j \ne k} c_k^{(m)} c_j Z_{t-k} Z_{t-j}.$$

Note that

$$E \left| a_n^{-2} \sum_{t=1}^{n} \sum_{k=0}^{\infty} \sum_{m < j \ne k} c_k^{(m)} c_j Z_{t-k} Z_{t-j} \right| = O \left(a_n^{-2} \sum_{t=1}^{n} \sum_{k=0}^{\infty} \sum_{m < j \ne k} |c_k^{(m)}||c_j| \right) = O(n a_n^{-2}) = o(1).$$

Consequently, relation (4.7) follows from Lemma 4.1 below.

Lemma 4.1 *Under the assumptions of Theorem 2.1, for any* $r > 0$

$$\lim_{m \to \infty} \limsup_{n \to \infty} P \left(a_n^{-2} \sum_{t=1}^{n} \sum_{j > m} c_j^2 Z_{t-j}^2 > r \right) = 0.$$

Proof. We will verify that

$$\lim_{m \to \infty} \limsup_{n \to \infty} P \left(a_n^{-2} \sum_{t=1}^{n} \sum_{j > m} c_j^2 Z_{t-j}^2 \mathbf{1} \{ |Z_{t-j}| > a_n \} > r \right) = 0 \tag{4.8}$$

and

$$\lim_{m \to \infty} \limsup_{n \to \infty} P \left(a_n^{-2} \sum_{t=1}^{n} \sum_{j > m} c_j^2 Z_{t-j}^2 \mathbf{1} \{ |Z_{t-j}| \le a_n \} > r \right) = 0. \tag{4.9}$$

Verification of (4.8)*:* Fix p so that $1 < 2p < \kappa$ and observe that

$$E \left[a_n^{-2} \sum_{t=1}^{n} \sum_{j > m} c_j^2 Z_{t-j}^2 \mathbf{1} \{ |Z_{t-j}| > a_n \} \right]^p \tag{4.10}$$

$$\leq a_n^{-2p}\left(\sum_{j>m}|c_j|^{2p}\right) nE\left[|Z|^{2p}\mathbf{1}\left\{|Z_{t-j}|>a_n\right\}\right].$$

Note that

$$E\left[|Z|^{2p}\mathbf{1}\left\{|Z_{t-j}|>a_n\right\}\right] = a_n^{2p}P(|Z|>a_n)+\int_{a_n^{2p}}^{\infty}P(|Z|>t^{1/2p})dt. \tag{4.11}$$

Thus the expectation in (4.10) is bounded above by

$$\left(\sum_{j>m}|c_j|^{2p}\right) nP(|Z|>a_n)+\left(\sum_{j>m}|c_j|^{2p}\right) a_n^{-2p}n\int_{a_n^{2p}}^{\infty}P(|Z|>t^{1/2p})dt. \tag{4.12}$$

Since $nP(|Z|>a_n)\to 1$, the $\lim_{m\to\infty}\limsup_{n\to\infty}$ of the first term in (4.12) is zero. To deal with the second term, note that $U(t)=P(|Z|>t^{1/2p})\in RV_{-\kappa/2p}$, so by Theorem 0.6 of Resnick (1987)

$$\int_{a_n^{2p}}^{\infty}P(|Z|>t^{1/2p})dt\sim\frac{2p}{\kappa-2p}a_n^{2p}P(|Z|>a_n).$$

Thus the second term in (4.12) is asymptotically equal to the first term up to a multiplicative constant.

Verification of (4.9)*:* Note that

$$E\left[a_n^{-2}\sum_{t=1}^{n}\sum_{j>m}c_j^2Z_{t-j}^2\mathbf{1}\left\{|Z_{t-j}|\leq a_n\right\}\right]=\left(\sum_{j>m}|c_j|^{2p}\right)a_n^{-2}nE\left[Z^2\mathbf{1}\left\{|Z|\leq a_n\right\}\right] \tag{4.13}$$

and

$$E\left[Z^2\mathbf{1}\left\{|Z|\leq a_n\right\}\right]=\int_0^{a_n^2}P(|Z|>t^{1/2})dt. \tag{4.14}$$

Since $U^*(t)=P(|Z|>t^{1/2})\in RV_{-\kappa/2}$, by Theorem 0.6 of Resnick (1987),

$$\int_0^{a_n^2}P(|Z|>t^{1/2})dt\sim\frac{2}{2-\kappa}a_n^2P(|Z|>a_n). \tag{4.15}$$

Combining (4.13). (4.14) and (4.15), we see that the expectation in (4.13) is bounded above by $C\left(\sum_{j>m}|c_j|^{2p}\right)nP(|Z|>a_n)$, and so (4.9) follows.

Proof of Theorem 2.3. As in the proof of Theorem 3.1 in Davis and Hsing (1995), note that for any $\varepsilon>0$ and real t_1,t_2 the map

$$T_\varepsilon:\sum_i\delta_{x_i}\mapsto\sum_i(t_1x_i+t_2x_i^2)\mathbf{1}\{|x_i|>\varepsilon\} \tag{4.16}$$

is continuous with respect to the point process $\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\delta_{P_iQ_{ij}}$. Therefore by (2.11) and the Continuous Mapping Theorem, we obtain

$$t_1S_{1n}(\varepsilon)+t_2S_{2n}(\varepsilon)\xrightarrow{d}t_1S_1(\varepsilon)+t_2S_2(\varepsilon), \tag{4.17}$$

where

$$S_{1n}(\varepsilon)=a_n^{-1}\sum_{t=1}^{n}Y_t\mathbf{1}\{|Y_t|>\varepsilon a_n\},\qquad S_{2n}(\varepsilon)=a_n^{-2}\sum_{t=1}^{n}Y_t^2\mathbf{1}\{|Y_t|>\varepsilon a_n\} \tag{4.18}$$

and

$$S_1(\varepsilon) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} P_i Q_{ij} \mathbf{1}\{P_i|Q_{ij}| > \varepsilon\}, \qquad S_2(\varepsilon) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} P_i^2 Q_{ij}^2 \mathbf{1}\{P_i|Q_{ij}| > \varepsilon\}. \tag{4.19}$$

The remainder of the proof relies on Theorem 3.2 of Billingsley (1999). We will show that there are random variables S_1 and S_2 such that $(S_1(\varepsilon), S_2(\varepsilon))$ converges in distribution to (S_1, S_2), as $\varepsilon \to 0$, and that for any $r > 0$

$$\lim_{\varepsilon\to 0} \limsup_{n\to\infty} P[|S_{1n} - S_{1n}(\varepsilon)| > r] = 0 \tag{4.20}$$

and

$$\lim_{\varepsilon\to 0} \limsup_{n\to\infty} P[|S_{2n} - S_{2n}(\varepsilon)| > r] = 0, \tag{4.21}$$

where

$$S_{1n} = \frac{1}{a_n}\sum_{t=1}^{n} Y_t, \qquad S_{2n} = \frac{1}{a_n^2}\sum_{t=1}^{n} Y_t^2. \tag{4.22}$$

Finally, we will identify the distributions of S_1 and S_2.

Denote by

$$\phi_\varepsilon(t_1, t_2) = E\exp[it_1 S_1(\varepsilon) + it_2 S_2(\varepsilon)]$$

the joint characteristic function of $S_1(\varepsilon)$ and $S_2(\varepsilon)$. We will show that $\phi_\varepsilon(t_1, t_2)$ is uniformly Cauchy on the set $\{(t_1, t_2) : \max(|t_1|, |t_2|) \le 1\}$. This implies that $\phi_\varepsilon(t_1, t_2)$ converges pointwise to a function which is continuous at the origin, so by the multivariate Continuity Theorem, see e.g. Remark on p. 147 of Durrett (1991), there exist random variables S_1 and S_2 such that $(S_1(\varepsilon), S_2(\varepsilon))$ converges in distribution to (S_1, S_2), as $\varepsilon \to 0$.

Similarly as in Davis and Hsing (1995) we write

$$\phi_v(t_1, t_2) - \phi_u(t_1, t_2) =: E_1(t_1, t_2; u, v; \delta) + E_2(t_1, t_2; u, v; \delta) =: E_1 + E_2, \tag{4.23}$$

where

$$\begin{aligned} E_1 = E\{&\exp(it_1 S_1(v) + it_2 S_2(v))\,[1 - \exp(it_1(S_1(u) - S_1(v)) + it_2(S_2(u) - S_2(v)))] \\ &\times \mathbf{1}\{\max(|S_1(u) - S_1(v)|, |S_2(u) - S_2(v)|) \le \delta\}\}; \\ E_2 = E\{&\exp(it_1 S_1(v) + it_2 S_2(v))\,[1 - \exp(it_1(S_1(u) - S_1(v)) + it_2(S_2(u) - S_2(v)))] \\ &\times \mathbf{1}\{\max(|S_1(u) - S_1(v)|, |S_2(u) - S_2(v)|) > \delta\}\}. \end{aligned}$$

Fix $\eta > 0$ and choose δ so that $E_1 < \eta/2$ provided $\max(|t_1|, |t_2|) \le 1$. Observe that

$$|E_2| \le 2P[|S_1(u) - S_1(v)| > \delta] + 2P[|S_2(u) - S_2(v)| > \delta].$$

On p. 897 of Davis and Hsing (1995) it is verified that for sufficiently small $\varepsilon > 0$

$$\sup_{0<u<v<\varepsilon} 2P[|S_1(u) - S_1(v)| > \delta] < \eta/4, \tag{4.24}$$

provided that for each $r > 0$

$$\lim_{\varepsilon\to 0} \limsup_{n\to\infty} P\left[a_n^{-1}|\sum_{t=1}^{n} Y_t \mathbf{1}\{|Y_j| \le \varepsilon a_n\}| > r\right] = 0. \tag{4.25}$$

Note that condition (4.25) follows from assumptions (2.8) and (2.10), c.f. (4.14) and (4.15). Indeed,

$$\mathrm{Var}\left[a_n^{-1}\sum_{t=1}^{n} Y_t \mathbf{1}\{|Y_j| \le \varepsilon a_n\}\right] = a_n^{-2} n E[Y_1^2 \mathbf{1}\{|Y_1| \le \varepsilon a_n\} \tag{4.26}$$

$$\sim 2(2-\kappa)^{-1}\varepsilon^2 nP[|Y_1| > a_n y] \sim 2(2-\kappa)^{-1}\varepsilon^{2-\kappa}, \quad \text{as } n \to \infty.$$

In addition to (4.24) we must show that for sufficiently small ε

$$\sup_{0<u<v<\varepsilon} 2P[|S_2(u) - S_2(v)| > \delta] < \eta/4. \tag{4.27}$$

Relation (4.27) follows from Lemma 4.2 below.

Relation (4.20) is the same as (4.25) and has already been verified, whereas relation (4.21) follows from (4.26).

We have established that (2.12) holds for some random variables S_1 and S_2. Applying the projection onto the first coordinate we obtain the marginal distribution of S_1 from Theorem 3.1 of Davis and Hsing (1995). Similarly, setting $W_i = \sum_{j=1}^{\infty} Q_{ij}^2$, and using the notation introduced in Lemma 4.2, we get the representation $\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} P_i^2 Q_{ij}^2 = \gamma^{2/\kappa}\sum_{i=1}^{\infty}\Gamma_i^{-\kappa/2}W_i$ with $E|W_i|^{\kappa/2} < \infty$. The series $\sum_{i=1}^{\infty}\Gamma_i^{-\kappa/2}W_i$ converges absolutely a.s., see e.g. Remark 4 on p. 29 of Samorodnitsky and Taqqu (1994), so $S_2(\varepsilon) \overset{a.s.}{\to} \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} P_i^2 Q_{ij}^2$. ∎

Lemma 4.2 *Under Assumption 2.3, for each $\delta > 0$,*

$$\lim_{u,v\to 0} P\left[\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\Gamma_i^{-2/\kappa}Q_{ij}^2\mathbf{1}\{u < \Gamma_i^{-1/\kappa}|Q_{ij}| \le v\} > \delta\right] = 0,$$

where $\Gamma_i = \sum_{k=1}^{i}\xi_k$ and the ξ_k are iid exponential with mean 1. (Recall that we can take $P_i = \gamma^{1/\kappa}\Gamma_i^{-1/\kappa}$ with the constant γ defined in Theorem 2.3 of Davis and Hsing (1995).)

Proof. It is well-known that in the series representations of the type considered in the present lemma, the term involving Γ_1 dominates the remaining terms, see e.g. the discussion on pp. 26-28 of Samorodnitsky and Taqqu (1994). We will therefore first show that

$$\lim_{u,v\to 0} P\left[\Gamma_1^{-2/\kappa}\sum_{j=1}^{\infty}Q_{1j}^2\mathbf{1}\{u < \Gamma_1^{-1/\kappa}|Q_{ij}| \le v\} > \delta\right] = 0 \tag{4.28}$$

and then verify that

$$\lim_{u,v\to 0} P\left[\sum_{i=2}^{\infty}\sum_{j=1}^{\infty}\Gamma_i^{-2/\kappa}Q_{ij}^2\mathbf{1}\{u < \Gamma_i^{-1/\kappa}|Q_{ij}| \le v\} > \delta\right] = 0. \tag{4.29}$$

To prove relation (4.28), it suffices to show that

$$\lim_{u,v\to 0} P\left[\sum_{j=1}^{\infty}|Q_{1j}|^{\kappa}\mathbf{1}\{\Gamma_1 u^{\kappa} < |Q_{1j}|^{\kappa} \le \Gamma_1 v^{\kappa}\} > \Gamma_1\delta\right] = 0. \tag{4.30}$$

The probability in (4.30) is equal to

$$\int_0^{\infty} P\left[\sum_{j=1}^{\infty}|Q_{1j}|^{\kappa}\mathbf{1}\{xu^{\kappa} < |Q_{1j}|^{\kappa} \le xv^{\kappa}\} > x\delta\right]e^{-x}dx,$$

so by the Dominated Convergence Theorem it is enough to check that that for any fixed $x > 0$

$$\lim_{u,v\to 0} P\left[\sum_{j=1}^{\infty}|Q_{1j}|^{\kappa}\mathbf{1}\{xu^{\kappa} < |Q_{1j}|^{\kappa} \le xv^{\kappa}\}\right] = 0$$

which in turn will follows from

$$\lim_{u,v\to 0}\sum_{j=1}^{\infty} E\left[|Q_{1j}|^{\kappa}\mathbf{1}\{xu^{\kappa} < |Q_{1j}|^{\kappa} \le xv^{\kappa}\}\right] = 0. \tag{4.31}$$

By Theorem 2.6 of Davis and Hsing (1995), $\sum_{j=1}^{\infty} E|Q_{1j}|^{\kappa} < \infty$, so relation (4.31) follows from the Dominated Convergence Theorem.

To verify (4.29), observe that if for $i \ge 2$, $E\Gamma_i^{-2/\kappa} < \infty$ and that in this case $E\Gamma_i^{-2/\kappa} = \Gamma(i-2/\kappa)/\Gamma(i) \sim i^{-2/\kappa}$. Therefore, since $Q_{ij}^2 \le |Q_{ij}|^{\kappa}$, we have

$$E\left|\sum_{i=2}^{\infty}\sum_{j=1}^{\infty}\Gamma_i^{-2/\kappa}Q_{ij}^2\right| \le \left(\sum_{i=2}^{\infty}E\Gamma_i^{-2/\kappa}\right)\left(\sum_{j=1}^{\infty}EQ_{1j}^2\right) = O\left(\sum_{i=2}^{\infty}i^{-2/\kappa}\right) = O(1).$$

Thus relation (4.29) follows from Markov's inequality and the Dominated Convergence Theorem.

References

Anderson, P. L. and Meerschaert, M. M. (1997). Periodic moving averages of random variables with regularly varying tails. *Ann. Statist.*, **24,** 771–785.

Basrak, B., Davis, R. A. and Mikosch, T. (2002a). A characterization of regular variation. *Annals of Applied Probablility*; Forthcoming. Preprint available at `http://www.math.ku.dk/~mikosch/preprint.html`.

Basrak, B., Davis, R. A. and Mikosch, T. (2002b). Regular variation of GARCH processes. *Stochastic Processes and Their Applications*; Forthcoming. Preprint available at `http://www.math.ku.dk/~mikosch/preprint.html`.

Bhansali, R. J. (1988). Consistent order determination for processes with infinite variance. *J. Roy. Statist. Soc.*, **50,** 46–60.

Billingsley, P. (1999). *Convergence of Probability Measures; Second Edition.* Wiley, New York.

Boussama, F. (2000). Ergodicity, mixing and estimation in GARCH models. Ph.D. Thesis. University of Paris 6. In French; translated by F. Comte as "Probabilistic properties of multivariate GARCH processes".

Bradley, R. C. (1986). Basic properties of strong mixing conditions. In *Dependence in Probability and Statistics* (eds E. Eberlein and M. S. Taqqu), pp. 165–192. Birkhäuser.

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods.* Springer-Verlag, New York.

Carasco, M. and Chen, X. (2002). Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory*, **18,** 000–000; preprint available at `http://www.econ.rochester.edu/Faculty/Carrasco.html`.

Cline, D. B. H. (1983). Estimation and linear prediction for regression, autoregression and ARMA with infinite variance data. Ph.D. Thesis. Colorado State University.

Cline, D. B. H. and Brockwell, P. J. (1985). Linear prediction of ARMA processes with infinite variance. *Stochastic Processes and their Applications*, **19,** 281–296.

Davis, R. A. and Hsing, T. (1995). Point process and partial sum convergence for weakly dependent random variables'. *Ann. Probab.*, **23,** 879–917.

Davis, R. A. and Mikosch, T. (1998). Limit theory for the sample acf of stationary process with heavy tails with applications to ARCH. *The Annals of Statistics*, **26,** 2049–2080.

Davis, R. A. and Resnick, S. I. (1985). Limit theory for moving averages of random variables with regularly varying tail probabilities. *The Annals of Probability*, **13,** number 1, 179–195.

Davis, R. A. and Resnick, S. I. (1986). Limit theory for the sample covariance and correlation functions of moving averages. *Ann. Statistics*, **14,** number 2, 533–558.

Doukhan, P. (1994). *Mixing: Properties and Examples.* Lecture Notes in Statistics. Springer Verlag.

Durrett, R. (1991). *Probability: Theory and Examples.* Wadsworth & Brooks/Cole, Pacific Grove, California.

Engle, R. F. and Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometric Reviews*, **5,** 1–50.

Gorodetskii, V. V. (1977). On the strong mixing property for linear sequences. *Theory Probab. Appl.*, **22,** 411–413.

Gouriéroux, C. (1997). *ARCH Models and Financial Applications.* Springer-Verlag, New York.

Kokoszka, P. S. (1996). Prediction of infinite variance fractional ARIMA. *Probability and Mathematical Statistics*, **16/1,** 65–83.

Kokoszka, P. S. and Taqqu, M. S. (1994). Infinite variance stable ARMA processes. *Journal of Time Series Analysis*, **115,** 203–220.

Kokoszka, P. S. and Taqqu, M. S. (1996). Parameter estimation for infinite variance fractional ARIMA. *The Annals of Statistics*, **24,** 1880–1913.

Kokoszka, P. S. and Taqqu, M. S. (2001). Can one use the Durbin–Levinson algorithm to generate infinite variance fractional ARIMA time series? *Journal of Time Series Analysis*, **22,** 317–337.

Logan, B. F., Mallows, C. L., Rice, S. O. and Shepp, L. A. (1973). Limit distributions of self-normallized sums. *Ann. Probability*, **1,** 788–809.

Maercker, G. and Moser, M. (1999). Yule-Walker estimators in GARCH(1,1) models: asymptotic normality and bootstrap. Beiträge zur Statistik 58. Universität Heidelberg. Available at `http://www.statlab.uni-heidelberg.de/reports/www.html` .

McElroy, T. and Politis, D. N. (2002). Robust inference for the mean in the presence of serial correlation and heavy-tailed distributions. *Econometric Theory*; Forthcoming.

Meerschaert, M. M. and Scheffler, H. P. (2001). *Limit Theorems for Sums of Independent Random Vectors*. Wiley, New York.

Mikosch, T., Gadrich, T., Klüppelberg, C. and Adler, R. J. (1995). Parameter estimation for ARMA models with infinite variance innovations. *The Annals of Statistics*, **23,** 305–326.

Mikosch, T. and Stărică, C. (2000). Limit theory for the sample autocorrelation and extremes of a GARCH(1, 1) process. *The Annals of Statistics*, **28,** 1427–1451.

Politis, D. N., Romano, J. P. and Wolf, M. (1999). *Subsampling*. Springer.

Resnick, S. I. (1987). *Extreme values, regular variation and point processes*. Springer-Verlag, New York.

Resnick, S. I. (1997). Heavy tail modeling and teletraffic data. *The Annals of Statistics*, **25,** 1805–1849; Special invited paper with discussion.

Samorodnitsky, G. and Taqqu, M. S. (1994). *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Chapman & Hall.

Piotr Kokoszka
Mathematics and Statistics
Utah State University
3900 Old Main Hill
Logan, Utah 84322-3900
United States
Email: piotr@stat.usu.edu
Phone: +1 (435) 797 0746
Fax: +1 (435) 797 1822
http://stat.usu.edu/∼piotr/

Michael Wolf
Economics and Business
Universitat Pompeu Fabra
Ramon Trias Fargas 25–27
08005 Barcelona
Spain
Email: michael.wolf@econ.upf.es
Phone: +34 (93) 542 2552
+ 34 (93) 542 1746
http://www.econ.upf.es/∼wolf/

5 Tables

Table 1: Estimated coverage probabilities of nominal 90% and 95% subsampling confidence intervals based on 2,000 replications. The DGP is an AR(1) model with stable innovations and the sample size is $n = 200$. ET stands for equal-tailed and SYM stands for symmetric. The data-dependent choice of block size is denoted by $\tilde{b}$.

		$\phi = 0.5;$	$\kappa = 1.2$		
Type	Target	$b = 10$	$b = 30$	$b = 50$	$\tilde{b}$
ET	0.90	0.82	0.78	0.72	0.81
SYM	0.90	0.98	0.96	0.91	0.97
ET	0.95	0.88	0.80	0.74	0.88
SYM	0.95	0.99	0.97	0.94	0.99
		$\phi = 0.5;$	$\kappa = 1.5$		
Type	Target	$b = 10$	$b = 25$	$b = 40$	$\tilde{b}$
ET	0.90	0.87	0.82	0.77	0.86
SYM	0.90	0.96	0.93	0.90	0.94
ET	0.95	0.92	0.85	0.80	0.92
SYM	0.95	0.99	0.96	0.93	0.98
		$\phi = 0.5;$	$\kappa = 1.8$		
Type	Target	$b = 10$	$b = 20$	$b = 30$	$\tilde{b}$
ET	0.90	0.90	0.86	0.82	0.89
SYM	0.90	0.92	0.88	0.85	0.90
ET	0.95	0.95	0.90	0.85	0.95
SYM	0.95	0.97	0.94	0.91	0.95

Table 2: Estimated coverage probabilities of nominal 90% and 95% subsampling confidence intervals based on 2,000 replications. The DGP is an AR(1) model with stable innovations and the sample size is $n = 500$. ET stands for equal-tailed and SYM stands for symmetric. The data-dependent choice of block size is denoted by $\tilde{b}$.

		$\phi = 0.5;$	$\kappa = 1.2$		
Type	Target	$b = 20$	$b = 80$	$b = 140$	$\tilde{b}$
ET	0.90	0.80	0.75	0.70	0.80
SYM	0.90	0.98	0.95	0.90	0.96
ET	0.95	0.85	0.77	0.72	0.85
SYM	0.95	0.99	0.97	0.93	0.98
		$\phi = 0.5;$	$\kappa = 1.5$		
Type	Target	$b = 20$	$b = 60$	$b = 100$	$\tilde{b}$
ET	0.90	0.85	0.81	0.77	0.85
SYM	0.90	0.95	0.92	0.88	0.93
ET	0.95	0.89	0.83	0.79	0.89
SYM	0.95	0.98	0.95	0.92	0.97
		$\phi = 0.5;$	$\kappa = 1.8$		
Type	Target	$b = 20$	$b = 50$	$b = 80$	$\tilde{b}$
ET	0.90	0.90	0.85	0.80	0.89
SYM	0.90	0.92	0.88	0.85	0.90
ET	0.95	0.93	0.88	0.84	0.93
SYM	0.95	0.96	0.93	0.90	0.95

Table 3: Estimated coverage probabilities of nominal 90% and 95% subsampling confidence intervals based on 2,000 replications. The DGP is a GARCH(1,1) model with normal innovations and the sample size is $n = 200$. ET stands for equal-tailed and SYM stands for symmetric. The (approximate) index κ was determined by numerical simulation. The data-dependent choice of block size is denoted by $\tilde{b}$.

$\omega = 1, \alpha_1 = 1.3, \beta_1 = 0.05;$				$\kappa \approx 1.19$	
Type	Target	$b = 10$	$b = 35$	$b = 60$	$\tilde{b}$
ET	0.90	0.89	0.82	0.75	0.88
SYM	0.90	0.98	0.95	0.90	0.93
ET	0.95	0.94	0.86	0.79	0.94
SYM	0.95	0.99	0.97	0.93	0.97
$\omega = 1, \alpha_1 = 1.1, \beta_1 = 0.1;$				$\kappa \approx 1.43$	
Type	Target	$b = 10$	$b = 35$	$b = 60$	$\tilde{b}$
ET	0.90	0.90	0.84	0.76	0.90
SYM	0.90	0.97	0.95	0.90	0.93
ET	0.95	0.95	0.87	0.79	0.95
SYM	0.95	0.99	0.97	0.93	0.97
$\omega = 1, \alpha_1 = 0.9, \beta_1 = 0.15;$				$\kappa \approx 1.83$	
Type	Target	$b = 10$	$b = 30$	$b = 50$	$\tilde{b}$
ET	0.90	0.90	0.84	0.78	0.90
SYM	0.90	0.95	0.91	0.85	0.90
ET	0.95	0.95	0.86	0.80	0.95
SYM	0.95	0.99	0.95	0.90	0.95

Table 4: Estimated coverage probabilities of nominal 90% and 95% subsampling confidence intervals based on 2,000 replications. The DGP is a GARCH(1,1) model with normal innovations and the sample size is $n = 500$. ET stands for equal-tailed and SYM stands for symmetric. The (approximate) index κ was determined by numerical simulation. The data-dependent choice of block size is denoted by $\tilde{b}$.

$\omega = 1, \alpha_1 = 1.3, \beta_1 = 0.05;$				$\kappa \approx 1.19$	
Type	Target	$b = 20$	$b = 85$	$b = 150$	$\tilde{b}$
ET	0.90	0.86	0.82	0.73	0.86
SYM	0.90	0.97	0.95	0.90	0.92
ET	0.95	0.93	0.85	0.76	0.93
SYM	0.95	0.99	0.97	0.93	0.96
$\omega = 1, \alpha_1 = 1.1, \beta_1 = 0.1;$				$\kappa \approx 1.43$	
Type	Target	$b = 20$	$b = 60$	$b = 100$	$\tilde{b}$
ET	0.90	0.88	0.85	0.80	0.87
SYM	0.90	0.97	0.95	0.90	0.91
ET	0.95	0.93	0.87	0.83	0.93
SYM	0.95	0.99	0.97	0.95	0.96
$\omega = 1, \alpha_1 = 0.9, \beta_1 = 0.15;$				$\kappa \approx 1.83$	
Type	Target	$b = 20$	$b = 60$	$b = 100$	$\tilde{b}$
ET	0.90	0.89	0.85	0.80	0.88
SYM	0.90	0.95	0.91	0.88	0.90
ET	0.95	0.93	0.88	0.83	0.93
SYM	0.95	0.98	0.95	0.91	0.95

Table 5: Estimated coverage probabilities of nominal 90% and 95% subsampling confidence intervals based on 2,000 replications. The DGP is a GARCH(1,1) model with normal innovations and the sample size is $n = 200$. ET stands for equal-tailed and SYM stands for symmetric. The data-dependent choice of block size is denoted by $\tilde{b}$.

	$\omega = 1, \alpha_1 = 0.1, \beta_1 = 0.9;$			$\kappa = 2$	
Type	Target	$b = 10$	$b = 20$	$b = 30$	$\tilde{b}$
ET	0.90	0.91	0.88	0.85	0.89
SYM	0.90	0.91	0.89	0.86	0.90
ET	0.95	0.96	0.92	0.89	0.95
SYM	0.95	0.97	0.94	0.92	0.95
	$\omega = 1, \alpha_1 = 0.5, \beta_1 = 0.5;$			$\kappa = 2$	
Type	Target	$b = 10$	$b = 30$	$b = 50$	$\tilde{b}$
ET	0.90	0.91	0.85	0.78	0.90
SYM	0.90	0.95	0.91	0.85	0.90
ET	0.95	0.96	0.89	0.83	0.95
SYM	0.95	0.99	0.95	0.90	0.95
	$\omega = 1, \alpha_1 = 0.9, \beta_1 = 0.1;$			$\kappa = 2$	
Type	Target	$b = 10$	$b = 25$	$b = 40$	$\tilde{b}$
ET	0.90	0.91	0.88	0.83	0.90
SYM	0.90	0.96	0.94	0.90	0.91
ET	0.95	0.95	0.90	0.86	0.95
SYM	0.95	0.98	0.96	0.94	0.95

Table 6: Estimated coverage probabilities of nominal 90% and 95% subsampling confidence intervals based on 2,000 replications. The DGP is a GARCH(1,1) model with normal innovations and the sample size is $n = 500$. ET stands for equal-tailed and SYM stands for symmetric. The data-dependent choice of block size is denoted by $\tilde{b}$.

	$\omega = 1, \alpha_1 = 0.1, \beta_1 = 0.9;$			$\kappa = 2$	
Type	Target	$b = 20$	$b = 50$	$b = 80$	$\tilde{b}$
ET	0.90	0.90	0.88	0.84	0.89
SYM	0.90	0.92	0.90	0.87	0.90
ET	0.95	0.94	0.90	0.87	0.94
SYM	0.95	0.96	0.94	0.91	0.95
	$\omega = 1, \alpha_1 = 0.5, \beta_1 = 0.5;$			$\kappa = 2$	
Type	Target	$b = 20$	$b = 70$	$b = 200$	$\tilde{b}$
ET	0.90	0.90	0.85	0.80	0.90
SYM	0.90	0.94	0.90	0.86	0.91
ET	0.95	0.94	0.88	0.83	0.94
SYM	0.95	0.98	0.95	0.91	0.96
	$\omega = 1, \alpha_1 = 0.9, \beta_1 = 0.1;$			$\kappa = 2$	
Type	Target	$b = 20$	$b = 60$	$b = 100$	$\tilde{b}$
ET	0.90	0.89	0.85	0.80	0.89
SYM	0.90	0.93	0.90	0.86	0.90
ET	0.95	0.93	0.88	0.83	0.93
SYM	0.95	0.96	0.94	0.90	0.95