SUBSAMPLING THE MEAN OF HEAVY-TAILED DEPENDENT OBSERVATIONS

By Piotr Kokoszka and Michael Wolf

Utah State University, Universitat Pompeu Fabra

First Version received February 2002

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 focus on 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.

Keywords. Estimation of the mean; heavy tails; GARCH; 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. Brockwell and Davis, 1991 Section 7.1), 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 regularly vary 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. McElroy and Politis (2002) considered constructing subsampling confidence intervals for μ assuming $\{Y_t\}$ is such a linear process. It has recently been established that the popular GARCH processes have regularly varying marginal distribution which may exhibit infinite variance for some choices of parameters (see Basrak *et al.*, 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). In this paper, we extend the theory of McElroy and Politis (2002) to such processes and develop a data-driven procedure for choosing the optimal subsampling block size.

2000 MSC Classification: 62M10.

0143-9782/04/02 217–234 JOURNAL OF TIME SERIES ANALYSIS Vol. 25, No. 2

The paper is organized as follows. In Section 2, we provide the necessary background on subsampling confidence intervals. Section 3 contains the relevant statistical theory, while Section 4 focuses on its practical implementation. The mathematical proofs and tables are collected in the Appendix.

2. SUBSAMPLING CONFIDENCE INTERVALS

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

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

where

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

Thus we approximate the sampling distribution of T_n by

$$L_{n,b}(x) = \frac{1}{n-b+1}\sum_{t=1}^{n-b+1}\mathbf{1}\left\{\frac{b^{1/2}(\bar{X}_{n,b,t} - \bar{X}_n)}{\hat{\sigma}_{n,b,t}} \le x\right\}. \tag{3}$$

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 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

$$J_n(x) = P\left\{\frac{\tau_n(\hat{\theta}_n - \theta)}{\hat{\sigma}_n \le x}\right\}. \tag{4}$$

ASSUMPTION 1. *There are nondegenerate distributions J, V, W, such that W has no mass at the origin, and positive sequences $\{t_n\}$ and $\{u_n\}$ such that, $\tau_n = t_n/u_n$ and*

$$J_n \xrightarrow{d} J, \tag{5}$$

$$t_n(\hat{\theta}_n - \theta) \xrightarrow{d} V, \tag{6}$$

$$u_n\hat{\sigma}_n \xrightarrow{d} W. \tag{7}$$

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

$$L_{n,b}(x) = \frac{1}{n-b+1}\sum_{t=1}^{n-b+1} \mathbf{1}\left\{\frac{\tau_b(\hat{\theta}_{n,b,t} - \hat{\theta}_n)}{\hat{\sigma}_{n,b,t}} \le x\right\}, \tag{8}$$

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

In Theorem 1 we assume that the time series under consideration is strong mixing. We recall the definition and some related facts here which will be referred to in the sequel. Suppose $\{X_t, t \in Z\}$ is a stationary random sequence. The *mixing rate function* m_k of $\{X_t\}$ is defined as

$$m_k = \sup\{|P(A \cap B) - P(A)P(B)|,\ A \in \sigma(X_s, s \le 0),\ B \in \sigma(X_s, s > k)\}, \tag{9}$$

with the σ-algebras in (9) defined in the usual way. (The m_k in (9) are usually denoted α_k but confusion with the coefficients in the GARCH specification has to be avoided (19).) 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.

THEOREM 1. *Suppose the process* $\{X_t\}$ *is strong mixing, Assumption* 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*

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

i.e., the subsampling confidence intervals yield asymptotically correct coverage probability.

The proof of Theorem 1 is the same as that 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 than use an argument for independent observations.

The difficulty in applying Theorem 1 lies in verifying Assumption 1 for a specific class of time series of interest. The case of independent observations was studied in Chapter 11 of Politis *et al.* (1999). Moving average models with heavy-tailed innovations were 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. In the present paper we focus on GARCH-type sequences introduced in Section 3 and use point-process techniques which can also be used to establish the results of McElroy and Politis (2002).

Remark 1. The approximation (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

$$J_{n,|\cdot|}(x) = P\Big\{\tau_n|\hat{\theta}_n - \theta|/\hat{\sigma}_n \le x\Big\}. \tag{10}$$

The according subsampling approximation is given by

$$L_{n,b,|\cdot|}(x) = \frac{1}{n-b+1}\sum_{t=1}^{n-b+1} \mathbf{1}\left\{\frac{b^{1/2}|\bar{X}_{n,b,t} - \bar{X}_n|}{\hat{\sigma}_{n,b,t}} \le x\right\}. \tag{11}$$

The asymptotic validity of this approach follows immediately from the validity of (8) and the continuous mapping theorem.

3. ASSUMPTIONS AND RESULTS

In this section we describe a nonparametric specification intended to model a time series which exhibits no 'correlation' but has a significant 'correlation in absolute values'. Series with such characteristics arise in finance and economics. Condition (12), in which $\overset{v}{\rightarrow}$ denotes vague convergence, together with (13) is equivalent to the requirement that the one-dimensional marginal distributions are in the domain of attraction of a κ-stable law (see e.g. Meerschaert and Scheffler, 2001, Propn 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 (14) which means that truncated variables are uncorrelated. Other assumptions are present in Assumption 2. We have found it convenient to use the theory of point processes, as it has been successfully applied in the context of GARCH processes by 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 (15) is implied by a very weak form of mixing assumed by Davis and Hsing (1995), 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 (15) and the other conditions in Assumption 2.

ASSUMPTION 2. *The sequence* $\{Y_t\}$ *is strictly stationary with symmetric univariate marginal distributions which satisfy*

$$nP(Y_1/a_n \in \cdot) \xrightarrow{v} \mu(\cdot), \tag{12}$$

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

$$2\mu(dx) = \kappa|x|^{-\kappa-1}\mathbf{1}\{x < 0\}dx + \kappa x^{-\kappa-1}\mathbf{1}\{x > 0\}dx. \tag{13}$$

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

$$E[Y_t\mathbf{1}\{|Y_t| \le y\}Y_s\mathbf{1}\{|Y_s| \le y\}] = 0 \tag{14}$$

and

$$\sum_{t=1}^{n} \delta_{Y_t/a_n} \xrightarrow{d} \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \delta_{P_iQ_{ij}}, \tag{15}$$

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

REMARK 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. *If Assumption* 2 *holds, then*

$$\left(\frac{1}{a_n}\sum_{t=1}^{n} Y_t, \frac{1}{a_n^2}\sum_{t=1}^{n} Y_t^2\right) \xrightarrow{d} (S_1, S_2), \tag{16}$$

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

Under the additional assumption that the process is strong mixing, Theorem 2 implies the validity of the subsampling confidence intervals for μ. Indeed, the assumptions of Theorem 1 hold with

$$\theta = \mu, \quad \hat{\theta}_n = \bar{X}_n, \quad t_n = na_n^{-1}, \quad u_n = n^{1/2}a_n^{-1}. \tag{17}$$

More specifically, (6) holds because

$$t_n(\bar{X}_n - \mu) = \frac{t_n}{n}\sum_{t=1}^{n} Y_t = \frac{1}{a_n}\sum_{t=1}^{n} Y_t \xrightarrow{d} S_1$$

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

$$u_n^2\hat{\sigma}_n^2 = \frac{u_n^2}{n}\sum_{t=1}^{n} Y_t^2 - (u_n\bar{Y}_n)^2 \sim \frac{1}{a_n^2}\sum_{t=1}^{n} Y_t^2 \xrightarrow{d} S_2$$

so (7) also holds. Relation (5) follows now from the joint convergence in Theorem 2. To summarize, we have thus established the following result:

THEOREM 3. *If Assumption* 2 *is satisfied and the process* $\{Y_t\}$ *is strong mixing, then the conclusions of Theorem* 1 *hold, with* $\tau_n = n^{1/2}$, $\theta = \mu$, $\hat{\theta}_n = \bar{X}_n$, $\hat{\sigma}_n$ *defined in* (2) *and J being the distribution of* $S1S_2^{-1/2}$ *with* S_1 *and* S_2 *as in Theorem* 2.

We now 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:

$$Y_t = \sigma_t\varepsilon_t, \tag{18}$$

$$\sigma_t^2 = \omega + \sum_{j=1}^{p}\alpha_j Y_{t-j}^2 + \sum_{j=1}^{q}\beta_j\sigma_{t-j}^2. \tag{19}$$

The *innovations* ε_k in (18) are i.i.d. and ω, α_j, β_j are non-negative parameters.

Several authors formulated conditions under which a GARCH process is strong mixing with geometric rate (see Maercker and Moser, 1999; Boussama, 2000; Basrak *et al.*, 2002b; Carasco and Chen, 2002). These conditions are not restrictive 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 *et al.* (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 in Davis and Mikosch (1998) and Mikosch and Stărică (2000), respectively. Finally, notice that if the innovations ε_t in (18) are symmetric, then (14) holds.

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. The estimation of κ from the observations $Y_1, \ldots, Y_n$ is discussed in Berkes *et al.* (2003).

We conclude this section by recalling that McElroy and Politis (2002) established analogs of Theorems 2 and 3 for linear sequences of the form

$$Y_t = \sum_{j=0}^{\infty} c_j Z_{t-j} \tag{20}$$

with the weights c_j satisfying

$$\sum_{j=0}^{\infty} |c_j| < \infty. \tag{21}$$

This model nests causal ARMA(p, q) and AR(∞) specifications. The i.i.d. innovations Z_t are in the domain of attraction of a κ-stable law with $1 < \kappa < 2$.

4. CHOICE OF THE BLOCK SIZE AND A SIMULATION STUDY

4.1. *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 aim 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 – i.e. 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 the 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 (e.g. 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.) Suppose the so-estimated order is $\hat{q}$. Fitting an MA($\hat{q}$) model to the zero-mean data $\hat{Y}_t = X_t - \bar{X}_n$, 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 $Z^*_{-\hat{q}+1}, \ldots, Z^*_n$ i.i.d. 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 $\hat{P}_n$.

Algorithm 1 describes how to pick the block size b, in practice.

Algorithm 1 (Choice of the Block Size)

Step 1. Fix a selection of reasonable block sizes b between limits b_{low} and b_{up}.
Step 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 $\text{CI}_{k,b}$ for μ.
Step 3. Compute $\hat{g}(b) = \#\{\bar{X}_n \in \text{CI}_{k,b}\}/K$.
Step 4. Find the value $\tilde{b}$ that minimizes $|\hat{g}(b) - (1 - \alpha)|$.

Remark 3. 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 4. Algorithm 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.

4.2. *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 as in McElroy and Politis (2002), who present results only for fixed block sizes:

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

The second DGP is a GARCH(1,1) model with normal innovations. 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 (e.g. 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 1 and select a very coarse grid of three 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 their distributional form.

The results, based on 2000 replications, are presented in Tables I–VI. We first discuss the results for the AR(1) model with stable innovations. It is seen that equal-tailed confidence intervals tend to undercover while symmetric confidence intervals tend to overcover. Performance improves as κ gets closer to 2. For $\kappa = 1.8$ the data-dependent choice of block size yields coverages close to the nominal level. On the other hand, for $\kappa = 1.2$ they are quite far away from the nominal level even for $n = 500$. Similar findings had been obtained by Politis *et al.* (1999, Ch. 11) who made inference for the mean of i.i.d. random variables having a stable distribution. It appears that very large sample sizes are needed when the stable index is close to 1.

To understand intuitively why the results get worse as κ approaches 1, note that for such observations values very far away from the mean are very likely, a realization of a series of this type will look like very small 'oscillations' around the mean with a few long 'spikes' that dominate the picture. To focus attention, suppose the time series $\{Y_t\}$ has one large positive 'outlier'. This means that $\bar{Y}$ will be large while most of the $\bar{Y}_{n,b,t}$ will be small because these subsamples will not include the 'outlier'. Thus, as the differences $\bar{Y}_{n,b,t} - \bar{Y}$, which are equal to the differences $\bar{X}_{n,b,t} - \bar{X}$, are small, the distribution $L_{n,b}$ will be shifted too much to the left and its right tail may not reach sufficiently far to zero. A reversed picture would arise if we had one dominating negative 'outlier', and both cases would result in an increased probability of not covering the mean. By contrast, for symmetric confidence intervals we are looking at the distribution of $|\bar{Y}_{n,b,t} - \bar{Y}|$ which, for the reasons explained above, tend to be shifted too much to the right and away from zero, and hence its upper quantiles will be too large. If we add these excessive quantiles to the sample mean, we will get too wide a confidence interval and a resultant overcoverage. We note that these explanations are intuitive and somewhat speculative and we present them with the hindsight of the available simulation results.

We also note that the usual bootstrap procedure does not work even for i.i.d. observations with infinite variance as realized by Athreya (1987). Instead, resamples must have a size smaller than that of the original sample (see Arcones and Giné, 1989, 1991).

The results are somewhat superior for the GARCH(1,1) model. Also here they improve as κ approaches 2 and tend to be better for the symmetric intervals. But even for $\kappa \approx 1.19$, coverages are reasonable for the sample sizes

TABLE I

ESTIMATED COVERAGE PROBABILITIES OF NOMINAL 90% AND 95% SUBSAMPLING CONFIDENCE INTERVALS BASED ON 2000 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 $\tilde{b}$

Type	Target	$b = 10$	$b = 30$	$b = 50$	$\tilde{b}$
$\phi = 0.5$; $\kappa = 1.2$					
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
		$b = 10$	$b = 25$	$b = 40$	$\tilde{b}$
$\phi = 0.5$; $\kappa = 1.5$					
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
		$b = 10$	$b = 20$	$b = 30$	$\tilde{b}$
$\phi = 0.5$; $\kappa = 1.8$					
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 II

ESTIMATED COVERAGE PROBABILITIES OF NOMINAL 90% AND 95% SUBSAMPLING CONFIDENCE INTERVALS BASED ON 2000 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 $\tilde{b}$

Type	Target	$b = 20$	$b = 80$	$b = 140$	$\tilde{b}$
$\phi = 0.5$; $\kappa = 1.2$					
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
		$b = 20$	$b = 60$	$b = 100$	$\tilde{b}$
$\phi = 0.5$; $\kappa = 1.5$					
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
		$b = 20$	$b = 50$	$b = 80$	$\tilde{b}$
$\phi = 0.5$; $\kappa = 1.8$					
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 III

ESTIMATED COVERAGE PROBABILITIES OF NOMINAL 90% AND 95% SUBSAMPLING CONFIDENCE INTERVALS BASED ON 2000 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 $\tilde{b}$

Type	Target	$b = 10$	$b = 35$	$b = 60$	$\tilde{b}$
$\omega = 1$, $\alpha_1 = 1.3$, $\beta_1 = 0.05$; $\kappa \approx 1.19$					
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
		$b = 10$	$b = 35$	$b = 60$	$\tilde{b}$
$\omega = 1$, $\alpha_1 = 1.1$, $\beta_1 = 0.1$; $\kappa \approx 1.43$					
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
		$b = 10$	$b = 30$	$b = 50$	$\tilde{b}$
$\omega = 1$, $\alpha_1 = 0.9$, $\beta_1 = 0.15$; $\kappa \approx 1.83$					
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 IV

ESTIMATED COVERAGE PROBABILITIES OF NOMINAL 90% AND 95% SUBSAMPLING CONFIDENCE INTERVALS BASED ON 2000 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 $\tilde{b}$

Type	Target	$b = 20$	$b = 85$	$b = 150$	$\tilde{b}$
$\omega = 1$, $\alpha_1 = 1.3$, $\beta_1 = 0.05$; $\kappa \approx 1.19$					
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
		$b = 20$	$b = 60$	$b = 100$	$\tilde{b}$
$\omega = 1$, $\alpha_1 = 1.1$, $\beta_1 = 0.1$; $\kappa \approx 1.43$					
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
		$b = 20$	$b = 60$	$b = 100$	$\tilde{b}$
$\omega = 1$, $\alpha_1 = 0.9$, $\beta_1 = 0.15$; $\kappa \approx 1.83$					
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 V

ESTIMATED COVERAGE PROBABILITIES OF NOMINAL 90% AND 95% SUBSAMPLING CONFIDENCE INTERVALS BASED ON 2000 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 $\tilde{b}$

Type	Target	$b = 10$	$b = 20$	$b = 30$	$\tilde{b}$
$\omega = 1$, $\alpha_1 = 0.1$, $\beta_1 = 0.9$; $\kappa = 2$					
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
		$b = 10$	$b = 30$	$b = 50$	$\tilde{b}$
$\omega = 1$, $\alpha_1 = 0.5$, $\beta_1 = 0.5$; $\kappa = 2$					
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
		$b = 10$	$b = 25$	$b = 40$	$\tilde{b}$
$\omega = 1$, $\alpha_1 = 0.9$, $\beta_1 = 0.1$; $\kappa = 2$					
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 VI

ESTIMATED COVERAGE PROBABILITIES OF NOMINAL 90% AND 95% SUBSAMPLING CONFIDENCE INTERVALS BASED ON 2000 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 $\tilde{b}$

Type	Target	$b = 20$	$b = 50$	$b = 80$	$\tilde{b}$
$\omega = 1,\ \alpha_1 = 0.1,\ \beta_1 = 0.9;\quad \kappa = 2$					
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
		$b = 20$	$b = 70$	$b = 200$	$\tilde{b}$
$\omega = 1,\ \alpha_1 = 0.5,\ \beta_1 = 0.5;\quad \kappa = 2$					
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
		$b = 20$	$b = 60$	$b = 100$	$\tilde{b}$
$\omega = 1,\ \alpha_1 = 0.9,\ \beta_1 = 0.1;\quad \kappa = 2$					
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

considered, in contrast to the linear case. Note the very good coverages for $\kappa = 2$ also, the case of an IGARCH model, not covered by our theory. The data-dependent method to pick the block size is seen to perform very well and is comparable with the 'optimal' fixed block size, which would be unknown in practice.

Our intuition as to why the results are superior for the GARCH(1,1) model is that realizations of such models do not exhibit isolated spikes but rather 'clusters of high volatility', so there will be more subsamples with large sample means in the presence of a large overall sample mean, and consequently the differences $\bar{X}_{n,b,t} - \bar{X}$ will not be 'too positive large' or 'too negative large ' 'too often'.

APPENDIX: MATHEMATICAL PROOFS

PROOF OF THEOREM 2. 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

$$T_\varepsilon : \sum_i \delta_{x_i} \mapsto \sum_i (t_1 x_i + t_2 x_i^2) \mathbf{1}\{|x_i| > \varepsilon\} \tag{A.1}$$

is continuous with respect to the point process $\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \delta_{P_i Q_{ij}}$. Therefore by (15) and the continuous mapping theorem, we obtain

$$t_1 S_{1n}(\varepsilon) + t_2 S_{2n}(\varepsilon) \xrightarrow{d} t_1 S_1(\varepsilon) + t_2 S_2(\varepsilon), \tag{A.2}$$

where

$$S_{1n}(\varepsilon) = a_n^{-1} \sum_{t=1}^{n} Y_t \mathbf{1}\{|Y_t| > \varepsilon a_n\}, \quad S_{2n}(\varepsilon) = a_n^{-2} \sum_{t=1}^{n} Y_t^2 \mathbf{1}\{|Y_t| > \varepsilon a_n\} \tag{A.3}$$

and

$$S_1(\varepsilon) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} P_i Q_{ij} \mathbf{1}\{P_i|Q_{ij}| > \varepsilon\}, \quad S_2(\varepsilon) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} P_i^2 Q_{ij}^2 \mathbf{1}\{P_i|Q_{ij}| > \varepsilon\}. \tag{A.4}$$

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 $r > 0$

$$\lim_{\varepsilon \to 0} \limsup_{n \to \infty} P[|S_{1n} - S_{1n}(\varepsilon)| > r] = 0 \tag{A.5}$$

and

$$\lim_{\varepsilon \to 0} \limsup_{n \to \infty} P[|S_{2n} - S_{2n}(\varepsilon)| > r] = 0, \tag{A.6}$$

where

$$S_{1n} = \frac{1}{a_n} \sum_{t=1}^{n} Y_t, \quad S_{2n} = \frac{1}{a_n^2} \sum_{t=1}^{n} Y_t^2. \tag{A.7}$$

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 $\{(t_1, t_2) : \max(|t_1|, |t_2|) \leq 1\}$. 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

$$\phi_v(t_1, t_2) - \phi_u(t_1, t_2) =: E_1(t_1, t_2; u, v; \delta) + E_2(t_1, t_2; u, v; \delta) =: E_1 + E_2, \tag{A.8}$$

where

$$E_1 = E\{\exp(it_1 S_1(v) + it_2 S_2(v))[1 - \exp(it_1(S_1(u) - S_1(v)) + it_2(S_2(u) - S_2(v)))] \times \mathbf{1}\{\max(|S_1(u) - S_1(v)|, |S_2(u) - S_2(v)|) \leq \delta\}\};$$

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

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

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

Davis and Hsing (1995, p. 897) verified that for sufficiently small $\varepsilon > 0$

$$\sup_{0<u<v<\varepsilon} 2P[|S_1(u) - S_1(v)| > \delta] < \eta/4, \tag{A.9}$$

provided that for each $r > 0$

$$\lim_{\varepsilon\to 0}\limsup_{n\to\infty} P\left[a_n^{-1}|\sum_{t=1}^{n} Y_t \mathbf{1}\{|Y_j| \le \varepsilon a_n\}| > r\right] = 0. \tag{A.10}$$

Note that condition (A.10) follows from (12) and (14) of Assumption 2. Indeed,

$$\begin{aligned}\mathrm{var}\left[a_n^{-1}\sum_{t=1}^{n} Y_t\mathbf{1}\{|Y_j| \le \varepsilon a_n\}\right] &= a_n^{-2} nE[Y_1^2\mathbf{1}\{|Y_1| \le \varepsilon a_n\}] \sim 2(2-\kappa)^{-1}\varepsilon^2 nP[|Y_1| > a_n y] \\ &\sim 2(2-\kappa)^{-1}\varepsilon^{2-\kappa}, \text{ as } n\to\infty. \end{aligned}\tag{A.11}$$

In addition to (A.9) we must show that for sufficiently small ε

$$\sup_{0<u<v<\varepsilon} 2P[|S_2(u) - S_2(v)| > \delta] < \frac{\eta}{4}. \tag{A.12}$$

Relation (A.12) follows from Lemma 1 below.

Relation (A.5) is the same as (A.10) and has already been verified, whereas relation (A.6) follows from (A.11).

We have established that (16) 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 1, we get the representation

$$\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} P_i^2 Q_{ij}^2 = \gamma^{2/\kappa}\sum_{i=1}^{\infty}\Gamma_i^{-\kappa/2} W_i \quad \text{with } E|W_i|^{\kappa/2} < \infty.$$

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 1. *Under Assumption 2, 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 i.i.d. 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).] 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. 26–8 of Samorodnitsky and Taqqu, 1994). We will therefore first show that

$$\lim_{u,v\to 0} P\left[\Gamma_1^{-2/\kappa}\sum_{j=1}^{\infty} Q_{1j}^2 \mathbf{1}\{u < \Gamma_1^{-1/\kappa}|Q_{ij}| \le v\} > \delta\right] = 0 \tag{A.13}$$

and then verify that

$$\lim_{u,v\to 0} P\left[\sum_{i=2}^{\infty}\sum_{j=1}^{\infty} \Gamma_i^{-2/\kappa} Q_{ij}^2 \mathbf{1}\{u < \Gamma_i^{-1/\kappa}|Q_{ij}| \le v\} > \delta\right] = 0. \tag{A.14}$$

To prove relation (A.13), it suffices to show that

$$\lim_{u,v\to 0} P\left[\sum_{j=1}^{\infty} |Q_{1j}|^{\kappa} \mathbf{1}\{\Gamma_1 u^{\kappa} < |Q_{1j}|^{\kappa} \le \Gamma_1 v^{\kappa}\} > \Gamma_1\delta\right] = 0. \tag{A.15}$$

The probability in (A.15) is equal to

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

so by the dominated convergence theorem it is enough to check that that for any fixed $x > 0$

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

which in turn follows from

$$\lim_{u,v\to 0} \sum_{j=1}^{\infty} E\left[|Q_{1j}|^{\kappa} \mathbf{1}\{x u^{\kappa} < |Q_{1j}|^{\kappa} \le x v^{\kappa}\}\right] = 0. \tag{A.16}$$

By Theorem 2.6 of Davis and Hsing (1995), $\sum_{j=1}^{\infty} E|Q_{1j}|^{\kappa} < \infty$, so relation (A.16) follows from the dominated convergence theorem.

To verify (A.14), 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 (A.14) follows from Markov's inequality and the dominated convergence theorem. QED

ACKNOWLEDGEMENTS

Piotr Kokoszka was partially supported by USU New Faculty grant and NATO grant PST.CLG.977607. Michael Wolf was supported by DGES grant BEC2001-1270.

NOTES

1. The stable innovations were generated using software of John Nolan; see the webpage http://academic2.american.edu/~jpnolan/

Corresponding author: Piotr Kokoszka, Mathematics and Statistics, Utah State University, 3900 Old Main Hill, Logan, Utah 84322-3900, USA. Tel.: +1 (435) 797 0746; Fax: +1 (435) 797 1822; E-mail: piotr@stat.usu.edu; http://stat.usu.edu/~piotr/
Michael Wolf, Economics and Business, Universitat Pompeu Fabra, Ramon Trias Fargas 25–27, 08005 Barcelona, Spain. Tel.: +34 (93) 542 2552; Fax: + 34 (93) 542 1746; E-mail: michael.wolf@econ.upf.es; http://www.econ.upf.es/~wolf/

REFERENCES

Anderson, P. L. and Meerschaert, M. M. (1997) Periodic moving averages of random variables with regularly varying tails. *Annals of Statistics* 24, 771–85.

Arcones, M. and Giné, E. (1989) The bootstrap of the mean with arbitrary bootstrap sample size. *Annals of the Institute Henri Poincaré* 22, 457–81.

Arcones, M. and Giné, E. (1991) Additions and corrections to 'the bootstrap of the mean with arbitrary bootstrap sample size'. *Annals of the Institute Henri Poincaré* 27, 583–95.

Athreya, K. B. (1987) Bootstrap of the mean in the infinite variance case. *The Annals of Statistics* 15(2), 724–31.

Basrak, B., Davis, R. A. and Mikosch, T. (2002a) A characterization of regular variation. *Annals of Applied Probability* 12, 908–20.

Basrak, B., Davis, R. A. and Mikosch, T. (2002b) Regular variation of GARCH processes. *Stochastic Processes and Their Applications* 99, 95–115.

Berkes, I., Horváth, L. and Kokoszka, P. S. (2003) Estimation of the maximal moment exponent of a GARCH(1,1) sequence. *Econometric Theory* 19, 565–86.

Bhansali, R. J. (1988) Consistent order determination for processes with infinite variance. *Journal of the Royal Statistical Society* 50, 46–60.

Billingsley, P. (1999) *Convergence of Probability Measures* (2nd edn). New York: Wiley.

Boussama, F. (1998) Ergodicity, mixing and estimation in GARCH models. PhD 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). Boston: Birkhäuser, pp. 165–92.

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

Carasco, M. and Chen, X. (2002) Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory* 18, 17–39.

Cline, D. B. H. and Brockwell, P. J. (1985) Linear prediction of ARMA processes with infinite variance. *Stochastic Processes and their Applications* 19, 281–96.

Davis, R. A. and Hsing, T. (1995) Point process and partial sum convergence for weakly dependent random variables'. *Annals of Probability* 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, 2049–80.

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

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

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

Kokoszka, P. S. and Taqqu, M. S. (1994) Infinite variance stable ARMA processes. *Journal of Time Series Analysis* 115, 203–20.

Kokoszka, P. S. and Taqqu, M. S. (1996) Parameter estimation for infinite variance fractional ARIMA. *The Annals of Statistics* 24, 1880–913.

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, 317–37.

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* 18, 1019–39.

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

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, 305–26.

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, 1427–51.

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

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