Subsampling unit root tests for heavy-tailed observations

Agnieszka Jach and Piotr Kokoszka*
Utah State University

August 11, 2003

Abstract

It is shown that the subsampling methodology can be used to develop unit root tests when the noise sequence is heavy-tailed with infinite variance. Using least-squares residuals, we construct processes which approximately satisfy the null hypothesis and then, using subsampling, we approximate the null distribution of test statistics. We establish the asymptotic validity of this method and demonstrate its applicability in finite samples by means of a simulation study and a data example.

Keywords: heavy tails, stable distribution, subsampling, unit root
2000 MSC: 62P20 60E07 62M10,

1 Introduction

In econometric theory and practice it is very important to be able to tell if the observations $y_1, \ldots, y_n$ follow a stationary time series model or if their differences $\Delta y_t = y_t - y_{t-1}$ are stationary. If the Δy_t form a stationary process, we say that the original series $\{y_t\}$ has a unit root or is integrated. In unit root tests, under the null hypothesis, the y_t have a unit root and under the alternative they form a stationary sequence. Unit root tests in various settings have been extensively studied in econometric and statistical literature in the last two decades, we refer to Chapter 17 of Hamilton (1994) and Philllips and Xiao (1998) for informative in-depth reviews.

To explain the contribution of the present paper, suppose the observations y_t follow the model

$$y_t = \rho y_{t-1} + u_t, \tag{1.1}$$

*Partially supported by NSF grant INT-0223262 and NATO grant PST.CLG.977607

where $-1 < \rho \leq 1$ is an unknown parameter and the u_t are independent. The unit root problem for this model consists in testing $H_0 : \rho = 1$ against $H_A : |\rho| < 1$; if $\rho = 1$, the process has a unit root, if $|\rho| < 1$, the y_t follow a stationary AR(1) model. We assume that under H_0, $y_0 = 0$, and under H_A, y_0 has the stationary distribution.

Essentially all work in this area assumed that the noise sequence $\{u_t\}$, usually in a more complex setting than (1.1), is stationary with finite variance. The goal is then to establish the null distribution of a test statistic. Typically the limiting distribution is a complicated functional of a Brownian motion (or a related process with stationary increments) and the parameters describing the noise process $\{u_t\}$, see e.g Wang *et al.* (2003), Ling and Li (2003) and Ling and McAleer (2003) for recent contributions in this direction.

Politis *et al.* (1999), Chapter 12, and Paparoditis and Politis (2003) used, respectively, subsampling and a residual block bootstrap to develop unit root tests for which the exact form of the limiting null distribution is not required. Earlier bootstrap procedures included Basawa *et al.* (1991) and Heimann and Kreiss (1996). Local power of bootstrap unit root tests for model (1.1) was studied in Swensen (2003). All the procedures assume that the noise process has at least finite variance, usually finite fourth moment.

Some economic and financial time series are believed not to have finite variance, see e.g. Guillaume *et al.* (1997) and Mittnik and Rachev (2000). Model (1.1) with infinite variance u_t was considered by Chan and Tran (1989) whose theory was extended in Chapter 15 of Mittnik and Rachev (2000) to more complex models and to t-tests. The approach of Chan and Tran (1989) and Mittnik and Rachev (2000) relies on finding the asymptotic distribution of a test statistic which is shown to depend on the tail index parameter κ, as explained in Section 3. The parameter κ is known to be very difficult to estimate, see e.g. Section 6.4 of Embrechts *et al.* (1997)

This paper shows that the subsampling methodology of Politis *et al.* (1999) can be used to develop unit root tests when the noise sequence is heavy-tailed with infinite variance without the need to estimate κ. Instead we have to choose a subsampling parameter b which is introduced in Section 2. Our approach is different from that advocated in Chapter 12 of Politis *et al.* (1999) in that it does not rely on constructing confidence intervals for the parameter ρ, see Remark 3.1. Instead, we construct from the data processes which approximately satisfy the null hypothesis. This method is similar to the one developed by Paparoditis and Politis (2003) who carefully explain that in order to derive a useful test, it must be ensured that the null distribution of a test statistic can be approximated no matter whether the data actually follow H_0 or H_A. As in the residual block bootstrap of Paparoditis and Politis (2003), we construct centered residuals, but instead of applying block bootstrap to them, we use shorter subsamples of centered residuals to construct many shorter pseudo-processes which satisfy H_0. As in Paparoditis and Politis (2003), these pseudo-processes allow us to approximate the null distribution of test statistics.

Subsampling is a methodology which is applicable to data assumed to be generated by a stationary or nearly stationary random process or field. It is presented in Politis *et al.* (1999). To explain the central idea, assume that $\{X_k\}$ is a strictly stationary sequence. Suppose we want to approximate the distribution of a statistic $T_n = T_n(X_1, \ldots, X_n)$ whose distribution function is $L_n(x) = P(T_n \le x)$. For this we consider $n - b + 1$ "copies" of T_n: $T_{b,k} = T_b(X_k, \ldots, X_{k+b-1})$ from which we compute the empirical distribution function

$$\hat{L}_b(x) = \frac{1}{n-b+1} \sum_{k=1}^{n-b+1} \mathbf{1}\{T_{b,k} \le x\}.$$

If the X_k are mixing in an appropriate sense and if the statistic T_n has a continuous limiting distribution, then

$$\sup_x |\hat{L}_b(x) - L_n(x)| \xrightarrow{P} 0,$$

provided $b \to \infty$, $b/n \to 0$. There a number of extensions of the above result.

In order to justify the use of subsampling, it is enough to know that a statistic of interest has a limiting distribution, it is not necessary to know the specific form of this distribution which may depend on unknown model parameters. Unlike for bootstrap methods, it is much easier to verify mathematically that subsampling is applicable to a given problem; there is no need to work with conditional distributions given the sample. As most nonparametric methods, like e.g. the block bootstrap, subsampling requires a practical choice of b for a given sample size n.

Assuming that H_0 is true, Zarepour and Knight (1999) and Horváth and Kokoszka (2003) recently established the validity of bootstrap approximations to the distribution of the unit root statistic in model (1.1) with infinite variance errors. In order to ensure the convergence of the bootstrap distribution, it is necessary to assume that the bootstrap sample size m satisfies $m \to \infty$ and $m/n \to 0$ as the sample size $n \to \infty$. The question of the practical choice of m has not been addressed yet. The need to reduce the bootstrap sample size if the observations have infinite variance was earlier observed in different contexts by Athreya (1987), Arcones and Giné (1989) and Knight (1989). The subsampling method relies on the choice of the subsample size b which must satisfy $b \to \infty$ and $b/n \to 0$. Unlike for bootstrap, this is required no matter whether the observations have finite or infinite variance. In our context, the need to choose b is thus not a unique disadvantage of the subsampling method because the same problem arises for bootstrap. In this paper we also address the question of the practical choice of b.

We establish the asymptotic validity of our method for model (1.1) and a somewhat more general model

$$y_t = \mu + \rho y_{t-1} + u_t \tag{1.2}$$

with possibly non-zero intercept μ, assuming that the u_t are independent. We investigate the applicability of our procedure to finite samples by means of a simulation study and an application to financial data.

The paper is organized as follows: In Section 2 we describe the test procedures for the model (1.1). Section 3 focuses on the mathematical formulation of the problem by stating the necessary assumptions and establishing the asymptotic validity of the method. Model (1.2) is discussed in Section 4. The results of a simulation study and an application to financial data are described in Section 5. Tables and graphs are collected in Appendix A, whereas the mathematical proofs are presented in Appendix B.

2 Test procedure

To describe the idea, we first focus on the simple model (1.1). In Section 4, we also consider the model with the intercept, but it is clear that the method can, at least in principle, be extended to more complex settings whenever the residuals can be constructed under both the null and the alternative. As will be seen from the calculations presented in Appendix B and the simulations, in order for the method to work parameter estimates must converge sufficiently fast.

Consider the least-squares estimator (LSE) of ρ in model (1.1):

$$\hat{\rho}_n = (\sum_{k=2}^{n} y_{k-1}^2)^{-1} \sum_{k=2}^{n} y_{k-1} y_k. \tag{2.1}$$

Chan and Tran (1989) showed that under H_0 the test statistic

$$T_n = n(\hat{\rho}_n - 1) \tag{2.2}$$

converges in distribution even when the u_k have infinite variance. Our goal is to develop an approximation to the null distribution of T_n which is valid for both $\rho = 1$ and $|\rho| < 1$, as we do not know which one is true. We compute the residuals

$$\hat{u}_k = y_k - \hat{\rho}_n y_{k-1}, \ k = 2, \dots, n$$

and the centered residuals

$$u_k^0 = \hat{u}_k - \frac{1}{n-1} \sum_{k=2}^{n} \hat{u}_k.$$

From the centered residuals we construct $n-b$ processes of length b which satisfy the null hypothesis. For $k = 2, \dots, n-b+1$, the kth process is defined by

$$\begin{aligned} y_0(k) = 0, \quad y_1(k) = u_k^0, \quad y_2(k) = u_k^0 + u_{k+1}^0, \dots, \\ y_b(k) = u_k^0 + u_{k+1}^0 + \dots, u_{k+b-1}^0. \end{aligned} \tag{2.3}$$

Let $\hat{\rho}_{b,k}^0$ be the LSE computed from $y_1(k), y_2(k), \dots, y_b(k)$. Denote by $q_{T,b}^0(\alpha)$ the αth quantile of the empirical distribution of the $n-b$ values $T_{b,k}^0 := b(\hat{\rho}_{b,k}^0 - 1)$. We reject H_0 if $n(\hat{\rho}_n - 1) < q_{T,b}^0(\alpha)$, because the empirical distribution of the $T_{b,k}^0$ is an approximation to the sampling distribution of T_n under H_0.

In the same way we can construct a t-test based on the statistic

$$t_n := \frac{\hat{\rho}_n - 1}{SD(\hat{\rho}_n)}, \tag{2.4}$$

where

$$SD(\hat{\rho}_n) = \left(\sum_{j=2}^{n} y_{j-1}^2 \right)^{-1/2} (s_n^2)^{1/2}; \tag{2.5}$$

$$s_n^2 = (n-1)^{-1} \sum_{j=2}^{n} (y_j - \hat{\rho}_n y_{j-1})^2. \tag{2.6}$$

If, analogously as above, $q_{t,b}^0(\alpha)$ is the αth empirical quantile of the subsampling approximation to the distribution of t_n, we reject H_0 if $t_n < q_{t,b}^0(\alpha)$.

3 Assumptions and results

In this section we state the assumptions which we need to prove the asymptotic validity of our approach. We assume that the errors u_t are independent. In Section 5 we present simulation results which show that the method also works reasonably well for weakly dependent u_t (we consider u_t which follow a GARCH model).

ASSUMPTION 3.1 *The iid errors u_t are in the domain of attraction of a stable law with index $1 < \kappa < 2$ and $Eu_t = 0$.*

To verify the consistency of the test procedure, we will need an additional assumption that

$$E|u_0|^\kappa = \infty. \tag{3.1}$$

Recall that Assumption 3.1 implies $P(|u_t| > x) = x^{-\kappa} l(x)$, where l is a slowly varying function, and consequently that $E|u_t|^p = \infty$ for $p > \kappa$. Condition (3.1) is not restrictive and is satisfied, for example, by all κ-stable random variables.

Our approach relies on the following result derived in Section 4.4 of Resnick (1986).

LEMMA 3.1 *Under Assumption 3.1*

$$\left(a_n^{-1} \sum_{i=1}^{[nt]} u_i, \quad a_n^{-2} \sum_{i=1}^{[nt]} u_i^2 \right) =: (U_n(t), V_n(t)) \to (U(t), V(t)) \text{ in } D([0,1], R^2),$$

where

$$a_n = \inf\{x : P(|u_0| > x) \le n^{-1}\} \tag{3.2}$$

and $(U(t), V(t))$ is a two-dimensional Levy process.

The exact definition of the Levy Process $(U(t), V(t))$ appearing in Lemma 3.1 is not needed in the following, but we recall that the quantiles a_n can be represented as

$$a_n = n^{1/\kappa} L(n), \tag{3.3}$$

for some slowly varying function L.

Chan and Tran (1989) observed that if (1.1) holds with $\rho = 1$, then

$$n^{-1} a_n^{-2} \sum_{j=1}^{n} y_{j-1}^2 = \int_0^1 U_n^2(t)dt \xrightarrow{d} \int_0^1 U^2(t)dt \tag{3.4}$$

and

$$a_n^{-2} \sum_{j=1}^{n} u_j y_{j-1} = \frac{1}{2}[U_n^2(1) - V_n(1)] \xrightarrow{d} \frac{1}{2}[U^2(1) - V(1)]. \tag{3.5}$$

Relations (3.4) and (3.5) follow from Lemma 3.1 and so represent the joint convergence. We therefore conclude that if $\rho = 1$ in model (1.1), then

$$T_n = n(\hat{\rho}_n - 1) = \frac{a_n^{-2}\sum_{j=1}^n u_j y_{j-1}}{n^{-1}a_n^{-2}\sum_{j=1}^n y_{j-1}^2} \xrightarrow{d} \frac{1}{2}\frac{U^2(1) - V(1)}{\int_0^1 U^2(t)dt}. \tag{3.6}$$

For a given κ, the distribution of the right-hand side of (3.6) can be obtained by simulation, but the value of κ is typically unknown and difficult to estimate.

To formalize the idea of the subsampling test, recall from Section 2 that

$$T_{b,k}^0 = b(\hat{\rho}_{b,k}^0 - 1) \tag{3.7}$$

and define

$$\hat{G}_b(x) = \frac{1}{n-b}\sum_{k=2}^{n-b+1} \mathbf{1}\{T_{b,k}^0 \le x\}. \tag{3.8}$$

Denote also (c.f. (3.6))

$$G(x) = P\left(\left[2\int_0^1 U^2(t)dt\right]^{-1}[U^2(1) - V(1)] \le x\right). \tag{3.9}$$

We need to show that the empirical distribution $\hat{G}_b$ converges pointwise in probability to G. As G is continuous, it follows that $\sup_x |\hat{G}_b(x) - G(x)|$ converges to zero a.s. Since under H_0, the distribution of T_n (2.2) also conveges uniformly a.s. to G, we otain from the triangle inequality that the uniform distance between the distribution of T_n and $\hat{G}_b$ tends to zero a.s. Thus Theorem 3.1 shows that the null distribution of the test statistic T_n can be approximated by $\hat{G}_b$ regardless whether H_0 or H_A is true.

THEOREM 3.1 *Suppose the observations $y_1, \ldots, y_n$ follow model* (1.1) *and, as $n \to \infty$, $b \to \infty$ and $b/n \to 0$.*

a) If $\rho = 1$ and Assumption 3.1 holds, then for every x, $\hat{G}_b(x) \xrightarrow{P} G(x)$.

b) If $|\rho| < 1$, Assumption 3.1 and (3.1) *hold, then then for every x, $\hat{G}_b(x) \xrightarrow{P} G(x)$.*

Part a) of Theorem 3.1 implies that the subsampling test has asymptotically correct size and part b) in conjunction with Proposition 3.1 below that the test is consistent.

PROPOSITION 3.1 *Suppose the observations $y_1, \ldots, y_n$ follow model* (1.1)*. If $|\rho| < 1$, Assumption 3.1 and* (3.1) *hold, then $n(\hat{\rho}_n - 1) \xrightarrow{P} -\infty$.*

Theorem 3.1 and Proposition 3.1 are proved in Appendix B. The idea of the proof of Theorem 3.1 is quite simple: If we could replace in (2.3) the u_k^0 by the u_k, and so the $T_{b,k}^0$ in (3.8) by the correspondingly defined $T_{b,k}$, then the standard theory developed in Politis *et al.* (1999) would apply because the $T_{b,k}$ are b-dependent and hence strong mixing. The difficulty lies in verifying that the difference between the $T_{b,k}^0$ and $T_{b,k}$ is small in an appropriate sense.

The test based on the t-statistic (2.4) can be justified in the same way, we omit the verification. The asymptotic null distribution of the t-statistic is given in Theorem 15.1 of Mittnik and Rachev (2000).

Remark 3.1 If the u_t have finite fourth moment, the statistic $R_n = (\hat{\rho}_n - \rho)/SD(\hat{\rho}_n)$ has an asymptotic distribution under both H_0 and H_A, see Chapter 12 of Politis *et al.* (1999) for the details. If the u_t have infinite variance and are in the domain of attraction of a stable law, then R_n converges under H_0, but under H_A, i.e. when the process y_t is stationary, $a_n^2 n^{-1/2} \tilde{a}_n^{-1} R_n$ converges in distribution, where the sequences a_n and $\tilde{a}_n$ are defined, respectively, by (3.2) and (B.1). In particular, if the u_t are symmetric and in the domain of normal attraction of a κ-stable law, then in order to ensure convergence under H_A, R_n must be multiplied by $n^{1/\kappa - 1/2}/\log n$. For this reason, the approach based on confidence intervals cannot be readily extended to unit root testing in the infinite variance case. An approach based on subsampling raw data cannot be applied either because under H_0 the data do not form an even approximately stationary process, c.f. sections 3.5 and 4.5 of Politis *et al.* (1999).

4 Model with the intercept

In this section we briefly consider model (1.2). The LSE's are

$$\begin{bmatrix} \hat{\mu}_n \\ \hat{\rho}_n \end{bmatrix} = \begin{bmatrix} n & \sum y_{j-1} \\ \sum y_{j-1} & \sum y_{j-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum y_j \\ \sum y_{j-1} y_j \end{bmatrix}. \tag{4.1}$$

The sums in (4.1) extend from 1 to n with the understanding that $y_0 = 0$.

In model (1.2) it is desirable to test the composite null hypothesis

$$H_0: \quad \rho = 1 \text{ and } \mu = 0. \tag{4.2}$$

Following Dickey and Fuller (1981) we consider the statistic

$$\Phi_n = \frac{(n-3)(RSS_0 - RSS_1)}{2RSS_1}. \tag{4.3}$$

In (4.3), RSS_1 is the unconstrained residual sum of squares

$$RSS_1 = \sum_{k=2}^{n} (y_k - \hat{\rho}_n y_{k-1} - \hat{\mu}_n)^2 = \sum_{k=2}^{n} \hat{u}_k^2 \tag{4.4}$$

and

$$RSS_0 = \sum_{k=2}^{n} (y_k - y_{k-1})^2 \tag{4.5}$$

is the residual sum of squares under the constraints $\rho = 1, \mu = 0$. Dickey and Fuller (1981) derived the asymptotic distribution of Φ_n under the assumption that the u_t are normal and obtained the critical values by Monte Carlo simulations. Notice that $\Phi_n \geq 0$ and large values of Φ_n lead to the rejection of H_0. The statistic Φ_n also has a limiting distribution if the u_t are heavy-tailed with infinite variance, but the limit is then different from that obtained by Dickey and Fuller (1981) and depends on κ. Therefore we propose to use subsampling to approximate the sampling distribution of Φ_n under H_0. From the residuals $\hat{u}_k = y_k - \hat{\rho}_n y_{k-1} - \hat{\mu}_n$ computed using (4.1), we construct $n-b$ processes according to (2.3) and compute $n-b$ subsample statistics $\Phi^0_{b,k}$, $k = 2, \ldots, n-b+1$. Denoting by $q^0_{\Phi,b}(1-\alpha)$ the $(1-\alpha)$th quantile of the empirical distribution of the $n-b$ values $\Phi^0_{b,k}$, we reject the null hypothesis (4.2) if $\Phi_n > q^0_{\Phi,b}(1-\alpha)$.

The asymptotic theory for the subsampling unit root test in model (1.2) is similar. Denote

$$\hat{G}_{\Phi,b}(x) = \frac{1}{n-b} \sum_{k=2}^{n-b+1} \mathbf{1}\{\Phi^0_{b,k} \le x\}, \tag{4.6}$$

where, as in Section 2, $\Phi^0_{b,k}$ is computed according to (4.3), but from the sequence (2.3) with

$$\hat{u}_k = y_k - \hat{\rho}_n y_{k-1} - \mu_k, \quad k = 2, \ldots, n, \tag{4.7}$$

using the estimators $\hat{\mu}$ and $\hat{\rho}$ defined in (4.1).

THEOREM 4.1 *Suppose the observations $y_1, \ldots, y_n$ follow model* (1.2) *with $\rho \in (-1, 1]$. If Assumption 3.1 and* (3.1) *hold and $b \to \infty$, $b/n \to 0$, as $n \to \infty$, then for every x, $\hat{G}_{\Phi,b}(x) \xrightarrow{P} G_\Phi(x)$, where $G_\Phi(x)$ is a continuous distribution function.*

An outline of the proof of Theorem 4.1 is given in Appendix B, where the distribution function G_Φ is also identified. To verify the consistency of the test, we have to show that if $|\rho| < 1$ or $\mu \neq 0$, then $\Phi_n \xrightarrow{P} \infty$. This is much more tedious than the proof of Proposition 3.1 because we have to deal with a ratio whose numerator and denominator each contains six terms whose asymptotic order must be taken into account. For this reason we do not present a mathematical argument that $\Phi_n \xrightarrow{P} \infty$ under H_A, but use simulations in Section 5 to show that the test is consistent.

5 Simulations and a data example

In this section we present the results of a simulation study intended to assess the finite sample performance of the test procedures described in Section 2 and address the issue of the choice of the subsample size b. We consider models (1.1) and (1.2) and the corresponding two tests:

$$H_0 : \rho = 1 \ \text{ against } \ H_A : |\rho| < 1 \tag{5.1}$$

and

$$H_0 : \rho = 1, \mu = 0 \ \text{ against } \ H_A : |\rho| < 1 \tag{5.2}$$

using, respectively, statistics T_n (2.2) and Φ_n (4.3). We also investigated the test based on the studentized statistic t_n (2.4), but since the results for the latter test are not better than those for the test based on T_n, we focus on the simpler statistic T_n.

5.1 Model without intercept

In this section we describe the results for the test (5.1) assuming model (1.1). Our simulations have shown that rejecting the null if $T_n < q^0_{T,b}(\alpha)$, as explained in Section 2, leads to a test that rejects too often for all choices of b for the sample sizes considered in our simulation study. We therefore propose to use a finite sample correction and reject if $T_n < q^0_{T,b}(\alpha - c)$, where $c = c(n, \alpha, b)$ is a function of n (sample size), α (significance level) and b (subsample size), but not of κ which we do not wish to estimate. After many experiments, we found that the folowing formula for c gives satisfactory results:

$$c = \frac{\sqrt{\alpha b}}{2\sqrt{1.6n}}. \tag{5.3}$$

We now proceed with a detailed description of the results for the test (5.1) which rejects H_0 if $T_n < q^0_{T,b}(\alpha - c)$ with c given by (5.3). The noise sequence u_t belongs to the domain of attraction of a stable law with index $1 < \kappa < 2$ and $Eu_t = 0$. In the following, we use the usual abbreviation, $S\kappa S$, for *symmetric κ-stable*; $S\kappa S$ random variables are archetypal examples of innovations satisfying Assumption 3.1. We also use innovations u_t with Student t–distribution with κ degrees of freedom to which we refer as $t(\kappa)$ errors. Tables A.1–A.3 contain the empirical size and power for $S\kappa S$ errors with $\kappa = 1.3, 1.6, 1.9$ and table A.4 for $t(\kappa)$ errors with $\kappa = 1.6$. The results are based on $R = 5000$ replications.

Since we use correction (5.3), it really does not matter which b we take; when using the correction (5.3) the results are very much the same for all reasonable choices of b. In the tables below we present the results for the values of b which coincide with the "optimal" values for the statistic Φ_n. As can be seen in Tables A.1–A.4 the results are good. The empirical sizes of the unit root tests are close to nominal. For sample sizes of 200 and greater, the power is also very high even for $\rho = 0.95$. We also checked the empirical power for $\rho = 0$ and the values were in range $(0.99, 1)$. The full set of results is available from the authors upon request.

According to our assumptions the ratio b/n tends to zero as $n \to \infty$. For sample sizes used in our simulations ($n = 100, 200, \ldots, 500$) a simple practical recommendation for the test (5.1) is to use the subsample size b approximately equal to anything between 15 and 10 percent of the sample size n and apply the correction (5.3). The results are almost the same as those reported in Tables A.1–A.4.

5.2 Model with intercept

We now focus on the performance of the subsampling method for model (1.2). We thus consider test (5.2) and reject if $\Phi_n > q^0_{\Phi,b}(1 - \alpha)$. In this setting, the optimal b depends on κ. To illustrate this bevavior, we present in Figures A.1 and A.2 representative graphs which show, respectively, the empirical size of the 10% and 5% tests as a function of b. Since many values of b had to be considered, computing time constraints forced us to compute the empirical size for every b based on only one thousand replications. Despite the sampling variability, Figures A.1 and A.2 clearly show that for a specified nominal size the optimal subsample size b depends on the tail index κ and a rough idea about the value of κ is needed. In most economic applications, it is reasonable to assume that $\kappa > 1.5$ because for smaller values of κ the simulated trajectories exhibit very high variability not observed in economic or financial data, see Chapter 1 of Mittnik and Rachev (2000) for a number of examples. For these reasons, in the case of model (1.2), we report detailed results for $\kappa = 1.5, 1.7, 1.9$. For $1.5 \leq \kappa < 2$, we recommend to use the values of b from Table A.5, which lead to slightly conservative 5% level tests and fairly accurate 10% level tests. These empirical sizes are based on $R = 10000$ replications and so are more accurate that those reported in Figures A.1 and A.2.

5.3 Dependent errors

The theory developed in the present paper assumes that the errors u_t are independent. We believe that at the expense of more technical proofs it can be extended to weakly dependent errors, but such an extension would go beyond the intended scope of the paper. It behooves us however to investigate to some extent the robustness of the proposed approach to the assumption of the independence of the u_t. We experimented with errors following several GARCH(1,1) models with standard normal innovations and found that the empirical sizes are fairly close to the nominal sizes if one uses the values of b from Table A.5 provided the tail index is greater than 1.5.

The errors u_t follow a GARCH(1,1) model if

$$u_t = \sigma_t \varepsilon_t \tag{5.4}$$

with iid innovations ε_t and the conditional volatility function

$$\sigma_t^2 = \omega + \alpha u_{t-1}^2 + \beta \sigma_{t-1}^2. \tag{5.5}$$

If the u_t have finite variance, they are dependent but uncorrelated. For the GARCH specification, there is however no clear-cut transition between finite and infinite variance u_t. Mikosch and Stărică (2000) have recently shown that under week regularity conditions, the tails of the u_t satisfy $P(|u_t| > x) \sim x^{-\kappa}$, where the value of κ depends in a "smooth" way on the parameters α and β in (5.5). More specifically, for $Z(t)$ defined as $Z(t) = E(\beta + \alpha\varepsilon_0^2)^t$, $\kappa = 2\mu_0$ where μ_0 satisfies $Z(\mu_0) = 1$ and $Z(t) < 1$ for all $0 < t \leq \mu_0$. Thus given the distribution of the innovations ε_t in (5.4) and the parameters in (5.5), the tail index κ cam be found numerically by Monte Carlo simulation. A modification of this method allows to estimate the tail index from the data, see Berkes *et al.* (2003).

In Table A.10 we report the results for the GARCH(1,1) model with $\omega = 1$, $\alpha = 0.9$ and $\beta = 0.15$ and standard normal residuals $\{\varepsilon_t\}$. For this set of parameters the index κ is about 1.4, so this model represents almost a limit of the applicability of the method considered in this paper. The test becomes somewhat too conservative, but the empirical sizes and powers are still reasonable. For GARCH models with κ closer to 2, there results are somewhat better.

5.4 Application to yield curves of corporate bonds

We conclude this section by applying the test investigated in Section 5.2 to yield curves on corporate bonds. These proprietary data have been made available by Professor Svetlozar (Zari) Rachev (UCSB and University of Karlsruhe). Figure A.3 shows two yield curves and corresponding returns on these yield curves. The returns are seen to exhibit a number of "outliers" which are a manifestation of their heavy-tailed distributions.

The statistic Φ_n allows us to compute the P-values. To illustrate, for a series of length $n = 500$, we choose $b = 70$, according to Table A.5, and compute the 430 values of $\Phi^0_{b,k}$. The percentage of these values which are greater that the value of Φ_n computed from the original data is the approximate P-value. For the data sets in Figure A.3 these P-values are 0.1395 for A and 0.0023 for B (the mean has been substracted before applying the test). These P-values confirm the impression one might get from a visual examination of the graphs; graph A looks like a random walk, whereas graph B looks more like a stationary process.

6 Conclusions

We proposed a method for unit root testing in models with heavy-tailed infinite variance errors. The method could be viewed as "residual subsampling" and is a variation on the subsampling methodology of Politis *et al.* (1999). We established the asymptotic validity of the test procedures and investigated their finite sample performance. The procedures have been shown to perform well for the models we considered. Our approach can readily be extended to more complex models, but its performance for such models has not been investigated. As most nonparametric methods, our procedure depends on a choice of a "bandwidth parameter", in our case the subsample size b. For sample sizes $100 \leq n \leq 500$ and the two models considered in this paper, we gave clear recommendations how b should be selected. As explained in Section 5.4, our approch allows to compute the P-values and can be implemented with low computational cost.

Unlike the asymptotic tests, the subsampling procedure does not require the knowledge of the often very complex asymptotic distribution which always depends on the unknown tail index κ. Despite several important theoretical contributions, we are not aware of any systematic empirical study aimed at a practical implementation of the asymptotic tests and assessing their performance. The same can be said about bootstrap tests. If the errors have infinite variance, both classes of tests have been studied only theoretically and it is often not entirely clear how they could be implemented in practice.

Appendices

A Tables and Graphs

		$\rho = 1$		$\rho = 0.95$		$\rho = 0.9$		$\rho = 0.5$	
n	$b \setminus \alpha$	0.10	0.05	0.10	0.05	0.10	0.05	0.10	0.05
100	25	0.1116	0.0748	0.5914	0.4510	0.8888	0.8210	0.9996	0.9996
200	35	0.1022	0.0640	0.8910	0.7890	0.9968	0.9866	1	1
300	50	0.1096	0.0602	0.9786	0.9258	0.9992	0.9980	1	1
400	60	0.0952	0.0504	0.9962	0.9764	1	0.9998	1	1
500	70	0.1064	0.0568	0.9994	0.9930	1	1	1	1

TABLE A.1 Empirical size and power of the subsampling test for model (1.1) with iid $S\kappa S$ errors, $\kappa = 1.3$. The results are based on $R = 5000$ replications.

		$\rho = 1$		$\rho = 0.95$		$\rho = 0.9$		$\rho = 0.5$	
n	$b \setminus \alpha$	0.10	0.05	0.10	0.05	0.10	0.05	0.10	0.05
100	25	0.1100	0.0770	0.5254	0.4120	0.8486	0.7626	0.9996	0.9996
200	35	0.1042	0.0654	0.8502	0.7280	0.9944	0.9812	1	1
300	50	0.1024	0.0614	0.9698	0.8916	0.9920	0.9950	1	1
400	60	0.0968	0.0500	0.9938	0.9698	1	1	1	1
500	70	0.0914	0.0478	0.9990	0.9920	0.9998	0.9998	1	1

TABLE A.2 Empirical size and power of the subsampling test for model (1.1) with iid $S\kappa S$ errors, $\kappa = 1.6$. The results are based on $R = 5000$ replications.

		$\rho = 1$		$\rho = 0.95$		$\rho = 0.9$		$\rho = 0.5$	
n	$b \setminus \alpha$	0.10	0.05	0.10	0.05	0.10	0.05	0.10	0.05
100	25	0.1066	0.0738	0.4884	0.3770	0.8244	0.7288	1	1
200	35	0.0930	0.0580	0.8166	0.6866	0.9930	0.9732	1	1
300	50	0.0860	0.0428	0.9592	0.8746	0.9998	0.9968	1	1
400	60	0.0842	0.0454	0.9924	0.9642	1	1	1	1
500	70	0.0844	0.0436	0.9990	0.9918	1	1	1	1

TABLE A.3 Empirical size and power of the subsampling test for model (1.1) with iid $S\kappa S$ errors, $\kappa = 1.9$. The results are based on $R = 5000$ replications.

n	$b \setminus \alpha$	$\rho = 1$ 0.10	$\rho = 1$ 0.05	$\rho = 0.95$ 0.10	$\rho = 0.95$ 0.05	$\rho = 0.9$ 0.10	$\rho = 0.9$ 0.05	$\rho = 0.5$ 0.10	$\rho = 0.5$ 0.05
100	25	0.1092	0.0762	0.5652	0.4420	0.8828	0.8020	1	1
200	35	0.0984	0.0620	0.8722	0.7580	0.9964	0.9848	1	1
300	50	0.0990	0.0530	0.9724	0.9072	0.9996	0.9978	1	1
400	60	0.1018	0.0570	0.9960	0.9754	0.9996	0.9994	1	1
500	70	0.0936	0.0526	0.9985	0.9934	1	0.9998	1	1

TABLE A.4 Empirical size and power of the subsampling test for model (1.1) based on processes given by (2.3) for iid $t(\kappa)$, $\kappa = 1.6$. The results are based on $R = 5000$ replications.

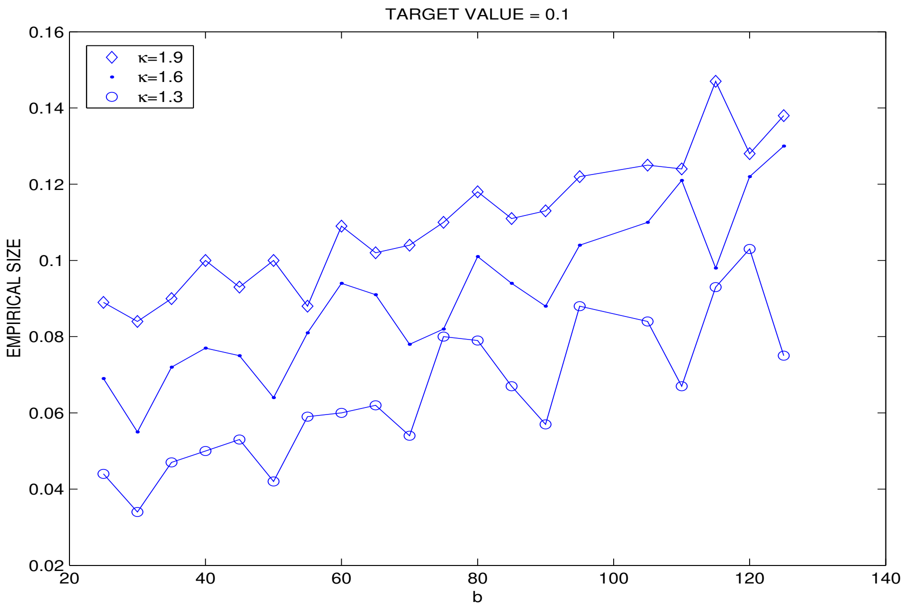

FIGURE A.1 Empirical size of the subsampling tests for model (1.2) with $S\kappa S$ errors and target value 0.1. The results are based on $R = 1000$ replications of length $n = 500$.

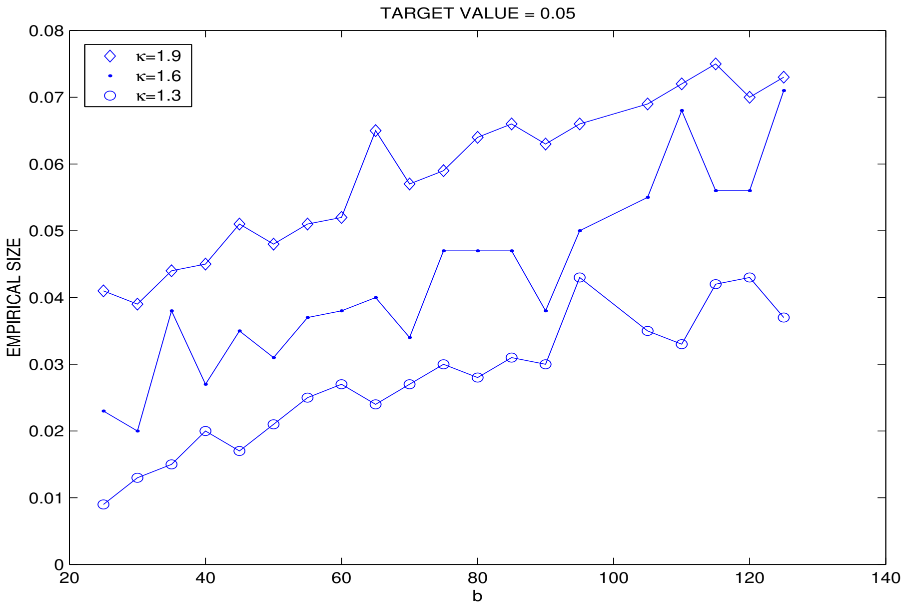

FIGURE A.2 Empirical size of the subsampling tests for model (1.2) with $S\kappa S$ errors and target value 0.05. The results are based on $R = 1000$ replications of length $n = 500$.

		$\kappa = 1.5$		$\kappa = 1.7$		$\kappa = 1.9$	
n	$b \setminus \alpha$	0.10	0.05	0.10	0.05	0.10	0.05
100	25	0.0949	0.0412	0.1059	0.0480	0.1249	0.0619
200	35	0.0908	0.0440	0.1012	0.0512	0.1023	0.0478
300	50	0.1005	0.0490	0.0968	0.0532	0.1062	0.0533
400	60	0.0912	0.0475	0.0999	0.0485	0.1016	0.0497
500	70	0.1016	0.0495	0.0951	0.0464	0.1016	0.0497

TABLE A.5 Empirical size of the subsampling test for model (1.2) with $S\kappa S$ errors based on the statistic Φ_n given by (4.3). The results are based on $R = 10000$ replications. The values of parameters are $\rho = 1$ and $\mu = 0$.

		$\kappa = 1.5$		$\kappa = 1.7$		$\kappa = 1.9$	
n	$b \setminus \alpha$	0.10	0.05	0.10	0.05	0.10	0.05
100	25	0.1231	0.0473	0.1446	0.0647	0.1823	0.0961
200	35	0.2810	0.1353	0.3463	0.1957	0.3948	0.2436
300	50	0.5675	0.3373	0.6361	0.4227	0.6748	0.4959
400	60	0.8037	0.5737	0.8611	0.6745	0.8866	0.7389
500	70	0.9997	0.9997	1	1	1	1

TABLE A.6 Empirical power of the subsampling test for model (1.2) with $S\kappa S$ errors based on the statistic Φ_n given by (4.3). The results are based on $R = 10000$ replications. The values of parameters are $\rho = 0.95$ and $\mu = 0$.

		$\kappa = 1.5$		$\kappa = 1.7$		$\kappa = 1.9$	
n	$b \setminus \alpha$	0.10	0.05	0.10	0.05	0.10	0.05
100	25	0.3149	0.1444	0.3640	0.1876	0.4114	0.2381
200	35	0.7973	0.5750	0.8514	0.6786	0.8819	0.7375
300	50	0.9402	0.8843	0.9779	0.9437	0.9900	0.9636
400	60	0.9697	0.9557	0.9930	0.9861	0.9996	0.9977
500	70	0.9797	0.9727	0.9978	0.9958	1	0.9998

TABLE A.7 Empirical power of the subsampling test for model (1.2) with $S\kappa S$ errors based on the statistic Φ_n given by (4.3). The results are based on $R = 10000$ replications. The values of parameters are $\rho = 0.9$ and $\mu = 0$.

		$\kappa = 1.5$		$\kappa = 1.7$		$\kappa = 1.9$	
n	$b \setminus \alpha$	0.10	0.05	0.10	0.05	0.10	0.05
100	25	0.9738	0.9658	0.9924	0.9899	0.9991	0.9980
200	35	0.9887	0.9870	0.9992	0.9987	0.9998	0.9998
300	50	0.9951	0.9933	0.9999	0.9999	1	1
400	60	0.9988	0.9982	0.9999	0.9999	1	1
500	70	0.9997	0.9997	1	1	1	1

TABLE A.8 Empirical power of the subsampling test for model (1.2) with $S\kappa S$ errors based on the statistic Φ_n given by (4.3). The results are based on $R = 10000$ replications. The values of parameters are $\rho = 0.5$ and $\mu = 0$.

		$\rho = 1$		$\rho = 0.95$		$\rho = 0.9$		$\rho = 0.5$	
n	$b \setminus \alpha$	0.10	0.05	0.10	0.05	0.10	0.05	0.10	0.05
100	25	0.0896	0.0368	0.1010	0.0376	0.2838	0.1120	0.9674	0.9562
200	35	0.0790	0.0334	0.2766	0.1184	0.7834	0.5486	0.9916	0.9892
300	50	0.0676	0.0314	0.5530	0.3100	0.9454	0.8872	0.9946	0.9944
400	60	0.0758	0.0320	0.7994	0.5570	0.9672	0.9526	0.9960	0.9954
500	70	0.0878	0.0400	0.9188	0.7920	0.9822	0.9756	0.9978	0.9978

TABLE A.9 Empirical size and power of the subsampling test for model (1.1) based on the statistic Φ_n given by (4.3) for iid $t(\kappa)$, $\kappa = 1.6$. The results are based on $R = 5000$ replications.

		$\rho = 1$		$\rho = 0.95$		$\rho = 0.9$		$\rho = 0.5$	
n	$b \setminus \alpha$	0.10	0.05	0.10	0.05	0.10	0.05	0.10	0.05
100	25	0.0996	0.0480	0.1520	0.0712	0.3316	0.1726	0.9672	0.9204
200	35	0.0862	0.0404	0.3072	0.1666	0.6896	0.5030	0.9950	0.9928
300	50	0.0882	0.0448	0.5176	0.3436	0.8754	0.7626	0.9976	0.9966
400	60	0.0772	0.0366	0.6920	0.4996	0.9414	0.8778	0.9974	0.9970
500	70	0.0816	0.0416	0.8006	0.6606	0.9650	0.9322	0.9984	0.9982

TABLE A.10 Empirical size and power of the subsampling test for model (1.2) based on the statistic Φ_n given by (4.3), where the errors u_t follow GARCH(1,1) model with $\omega = 1$, $\alpha = 0.9$, $\beta = 0.15$ and standard normal residuals. The index κ is about 1.4. The results are based on $R = 5000$ replications. The values of parameters are $\rho = 1$ and $\mu = 0$.

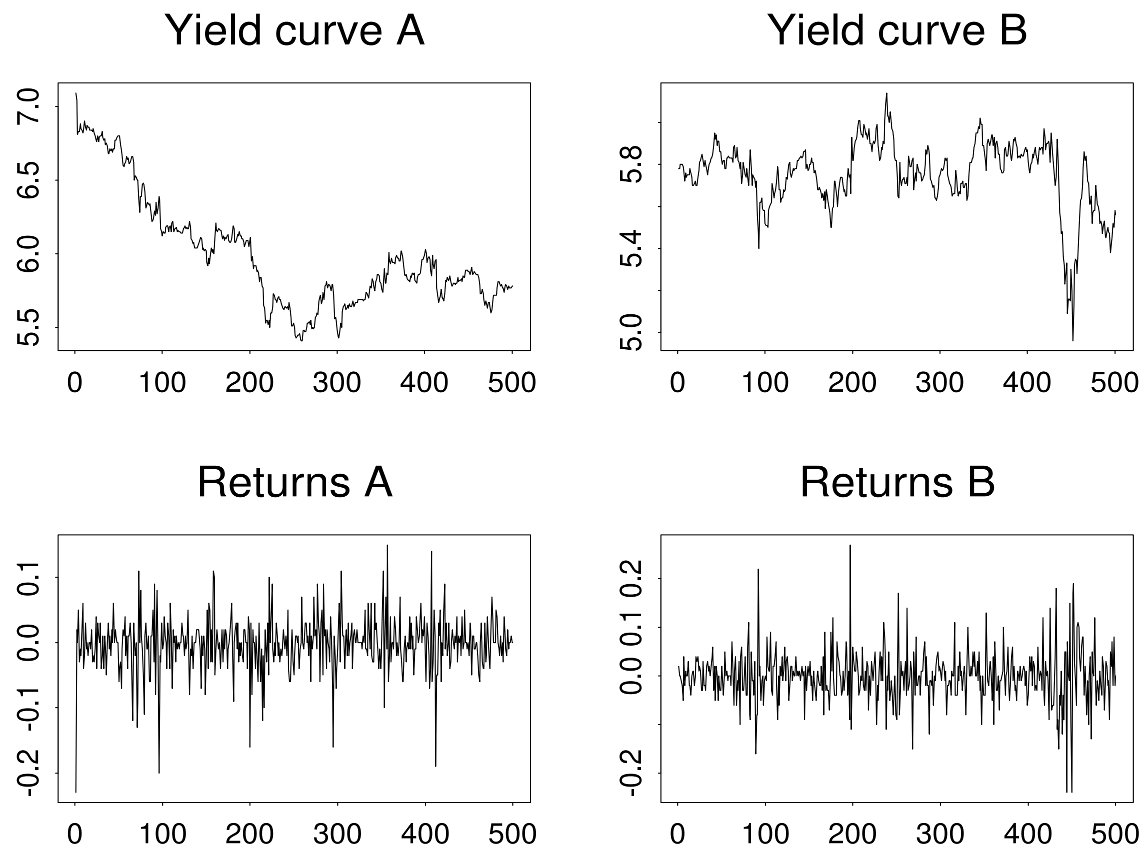

FIGURE A.3 Yield curves of corporate bonds and corresponding returns. The bottom panels represent first differences of the top panels.

B Proofs

We begin with the proof of Proposition 3.1. Throughout this section we use the notation introduced in Sections 2 and 3.

PROOF OF PROPOSITION 3.1. Define

$$\tilde{a}_n = \inf\{x : P(|u_0 u_1| > x) \leq n^{-1}\}. \tag{B.1}$$

It follows from Theorem 4.4 of Davis and Resnick (1986) that

$$D_n := \tilde{a}_n^{-1} a_n^2 (\hat{\rho}_n - \rho) \xrightarrow{d} S_1/S_0, \tag{B.2}$$

where S_1 is a κ-stable and S_0 a positive $\kappa/2$-stable random variable. Since $\tilde{a}_n = n^{1/\kappa} L_2(n)$ for some slowly varying function L_2, see Davis and Resnick (1986), p. 542, we conclude that

$$\tilde{a}_n^{-1} a_n^2 = n^{1/\kappa} \tilde{L}(n) \tag{B.3}$$

for some slowly varying function $\tilde{L}$. Therefore for any $M > 0$,

$$P(n(\hat{\rho}_n - 1) > -M) = P(1 - \rho - \tilde{a}_n a_n^{-2} D_n < M/n) \to P(1 \leq \rho) = 0.$$

We now present a number of technical lemmas which will be needed in the proof of Theorem 3.1. Lemma B.1 is elementary, but we state it here because it is often used in the arguments below. Lemmas B.2, B.3 and B.4 are needed to prove part a) of Theorem 3.1, Lemmas B.5, B.6 and B.8 are needed for part b).

LEMMA B.1 *Suppose S_n, U and $U_{k,n}$, $1 \leq k \leq n$, are random variables such that $S_n \xrightarrow{P} 0$, $U_{k,n} \stackrel{d}{=} U_{1,n}$, $U_{1,n} = O_P(1)$. Then for every $\varepsilon > 0$,*

$$\limsup_{n\to\infty} \max_k P(|S_n U_{k,n}| > \varepsilon) = 0.$$

LEMMA B.2 *Under the assumptions of Theorem 3.1 a), for $k = 2, \ldots, n$,*

$$u_k^0 = u_k + r_k(n), \tag{B.4}$$

where

$$r_k(n) = -\frac{1}{n-1}\sum_{j=2}^{n} u_j + \frac{1}{n} T_n \left(y_{k-1} - \frac{1}{n-1}\sum_{j=2}^{n} y_{j-1} \right) \tag{B.5}$$

satisfies

$$\max_{1<k\leq n-b+1} |r_k(n)| = O_P(n^{-1} a_n). \tag{B.6}$$

PROOF. Recall that $U(t)$ is the limit in $D[0,1]$ of the process $a_n^{-1}\sum_{j=1}^{[nt]} u_j$. Therefore, (B.6) follows from $T_n = O_P(1)$ and the following relations:

$$\frac{n}{a_n}\frac{1}{n-1}\sum_{j=2}^{n} u_j \xrightarrow{d} U(1);$$

$$\frac{n}{a_n}\frac{1}{n} y_{k-1} = a_n^{-1}\sum_{j=1}^{k-1} u_j, \quad \sup_k \left| a_n^{-1}\sum_{j=1}^{k-1} u_j \right| \xrightarrow{d} \sup_{0<t<1} |U(t)|;$$

$$\frac{n}{a_n}\frac{1}{n}\frac{1}{n-1}\sum_{j=2}^{n} y_{j-1} = \frac{1}{n-1}\sum_{j=2}^{n} a_n^{-1}\sum_{i=1}^{j-1} u_i \xrightarrow{d} \int_0^1 U(t)dt.$$

We now introduce some notation

$$U_{b,k}(t) = a_b^{-1}\sum_{i=k}^{[(k+b-1)t]} u_i, \quad V_{b,k} = a_b^{-2}\sum_{i=k}^{k+b-1} u_i^2, \quad 1 \leq k \leq n-b+1; \tag{B.7}$$

$$U^0_{b,k}(t) = a_b^{-1} \sum_{i=k}^{[(k+b-1)t]} u_i^0, \quad V^0_{b,k} = a_b^{-2} \sum_{i=k}^{k+b-1} u_i^{02}, \quad 1 < k \le n-b+1. \tag{B.8}$$

Define also

$$T^N_{b,k} = \frac{1}{2}[U^2_{b,k}(1) - V_{b,k}], \quad T^D_{b,k} = \int_0^1 U^2_{b,k}(t)dt \tag{B.9}$$

and analogously $T^{0N}_{b,k}$ and $T^{0D}_{b,k}$, so that $T^0_{b,k} = T^{0N}_{b,k}/T^{0D}_{b,k}$.

Lemma B.3 *Under the assumptions of Theorem 3.1 a), for every real x*

$$\lim_{n\to\infty} \frac{1}{n-b} \sum_{k=2}^{n-b+1} P(T^0_{b,k} \le x) \to G(x).$$

Proof. Using the notation introduced in (B.7) - (B.9) define

$$R^N_{b,k} = T^{0N}_{b,k} - T^N_{b,k}, \quad R^D_{b,k} = T^{0D}_{b,k} - T^D_{b,k}. \tag{B.10}$$

Observe that $T^0_{b,k} \le x$ is equivalent to

$$T^N_{b,k} + R^N_{b,k} \le xT^D_{b,k} + xR^D_{b,k}. \tag{B.11}$$

Notice that (B.11) yields for every $\varepsilon > 0$

$$P(T^0_{b,k} \le x) \le P(T^N_{b,k} \le xT^D_{b,k} + 2\varepsilon) + P(R^N_{b,k} \le -\varepsilon) + P(xR^D_{b,k} \ge \varepsilon) \tag{B.12}$$

and

$$P(T^0_{b,k} \le x) \ge P(T^N_{b,k} \le xT^D_{b,k} - 2\varepsilon) - [P(R^N_{b,k} \ge \varepsilon) + P(xR^D_{b,k} < -\varepsilon)]. \tag{B.13}$$

Suppose we can show that for every $\varepsilon > 0$

$$\limsup_{n\to\infty} \max_{1<k\le n-b+1} P(|R^N_{b,k}| > \varepsilon) = 0 \tag{B.14}$$

and that the same relation holds for $R^D_{b,k}$. To complete the proof, we can then use a standard argument involving limsup and liminf which utilizes the fact that the distribution function of

$$\frac{1}{2}[U^2(1) - V(1)] - x\int_0^1 U^2(t)dt \tag{B.15}$$

is continuous ((B.15) is a $\kappa/2$ stable random variable), see the end of the proof of Lemma B.4.

To save space we only outline the argument for $R^N_{b,k}$. Setting

$$S_{b,k}(t) = a_b^{-1} \sum_{i=k}^{[(k+b-1)t]} r_i(n) \tag{B.16}$$

we obtain by Lemma B.2 that

$$\sup_{0\le t\le 1} \max_{1<k\le n-b+1} |S_{b,k}(t)| = O_P\left(\frac{b}{a_b}\frac{a_n}{n}\right) = o_P(1), \tag{B.17}$$

because of the assumptions $b/n \to 0$ and $\kappa > 1$. Using (B.4), (B.7)–(B.9), (B.10) and (B.16), we obtain

$$R^N_{b,k} = U_{b,k}(1)S_{b,k}(1) + \frac{1}{2}S^2_{b,k}(1) - a_b^{-2}\sum_{i=k}^{k+b-1} u_i r_i(n) - \frac{1}{2}a_b^{-2}\sum_{i=k}^{k+b-1} r_i^2(n). \tag{B.18}$$

Using Lemma B.1 and (B.17), we conclude that

$$\limsup_{n\to\infty}\max_{1<k\le n-b+1} P(|U_{b,k}(1)S_{b,k}(1)| > \varepsilon) = 0. \tag{B.19}$$

The remaining three terms in (B.18) are dealt with in the same way.

Lemma B.4 *Under the assumptions of Theorem 3.1 a),* $\mathrm{Var}\left[\hat{G}_b(x)\right] \to 0$.

Proof. Set $q = n - b + 1$. Then

$$\begin{aligned}
\mathrm{Var}\left[\hat{G}_b(x)\right] &= \frac{1}{(n-b)^2}\sum_{k=2}^{q}\mathrm{Var}[\mathbf{1}\{T^0_{b,k}\le x\}] \\
&+ \frac{2}{(n-b)^2}\sum_{j=1}^{n-b}\sum_{k=2}^{n-b-j}\mathrm{Cov}\left(\mathbf{1}\{T^0_{b,k}\le x\},\mathbf{1}\{T^0_{b,k+j}\le x\}\right). \\
&=: S_1 + S_2.
\end{aligned}$$

Clearly $S_1 \to 0$ as $n \to \infty$. We split S_2 into two sums

$$S_{21} = \frac{2}{(n-b)^2}\sum_{j=1}^{b}\sum_{k=2}^{n-b-j}\mathrm{Cov}\left(\mathbf{1}\{T^0_{b,k}\le x\},\mathbf{1}\{T^0_{b,k+j}\le x\}\right);$$

$$S_{22} = \frac{2}{(n-b)^2}\sum_{j=b+1}^{n-b}\sum_{k=2}^{n-b-j}\mathrm{Cov}\left(\mathbf{1}\{T^0_{b,k}\le x\},\mathbf{1}\{T^0_{b,k+j}\le x\}\right).$$

The assumption $b/n \to 0$, yields immediately $S_{21} \to 0$. We now focus on S_{22}. Using the notation introduced in Lemma B.3, we set

$$\delta_n = \max_{1<k\le n-b+1}[P(|R^N_{b,k}| \ge \varepsilon) + P(|xR^D_{b,k}| \ge \varepsilon)]. \tag{B.20}$$

We have verified in Lemma B.3 that

$$\limsup_{n\to\infty}\delta_n = 0. \tag{B.21}$$

Define also

$$p_n^+ = P(T^N_{b,1} \le xT^D_{b,1} + 2\varepsilon), \qquad p_n^- = P(T^N_{b,1} \le xT^D_{b,1} - 2\varepsilon),$$
$$p^+_{n,j} = P(T^N_{b,1} \le xT^D_{b,1} + 2\varepsilon, T^N_{b,1+j} \le xT^D_{b,1+j} + 2\varepsilon)$$

and analogously $p_{n,j}^-$. Note that (c.f. (B.12) and (B.13))

$$p_n^- - \delta_n \le P(T_{b,k}^0 \le x) \le p_n^+ + \delta_n$$

and

$$p_{n,j}^- - 2\delta_n \le P(T_{b,k}^0 \le x, T_{b,k+j}^0 \le x) \le p_{n,j}^+ + 2\delta_n.$$

The above inequalities yield

$$p_{n,j}^- - 2\delta_n - (p_n^+ + \delta_n)^2 \le \operatorname{Cov}\left(\mathbf{1}\{T_{b,k}^0 \le x\}, \mathbf{1}\{T_{b,k+j}^0 \le x\}\right) \le p_{n,j}^+ + 2\delta_n - (p_n^- - \delta_n)^2.$$

Since for $j \ge b+1$, $(T_{b,1}^N, T_{b,1}^D)$ and $(T_{b,1+j}^N, T_{b,1+j}^D)$ are independent, we conclude that for $j \ge b+1$, $p_{n,j}^+ = (p_n^+)^2$ and $p_{n,j}^- = (p_n^-)^2$ and so for these j

$$(p_n^-)^2 - (p_n^+)^2 - 2\delta_n(p_n^+ + \delta_n) \le \operatorname{Cov}\left(\mathbf{1}\{T_{b,k}^0 \le x\}, \mathbf{1}\{T_{b,k+j}^0 \le x\}\right) \le (p_n^+)^2 - (p_n^-)^2 + 2\delta_n(p_n^- + \delta_n).$$

By (B.21), we therefore obtain

(B.22) $$\limsup_{n\to\infty} S_{22} \le 2[(p^+)^2 - (p^-)^2]$$

and

(B.23) $$\liminf_{n\to\infty} S_{22} \ge 2[(p^-)^2 - (p^+)^2],$$

where

$$p^+ = P\left(\frac{1}{2}[U^2(1) - V(1)] - x\int_0^1 U^2(t)dt \le 2\varepsilon\right)$$

and

$$p^- = P\left(\frac{1}{2}[U^2(1) - V(1)] - x\int_0^1 U^2(t)dt \le -2\varepsilon\right).$$

Since (B.22) and (B.23) hold for any $\varepsilon > 0$ and p^+ and p^- tend to the same limit as $\varepsilon \to 0$, we obtain $\lim_{n\to\infty} S_{22} = 0$, what completes the proof of the lemma.

Lemma B.5 *Under the assumptions of Theorem 3.1 b), for $k = 2, \ldots, n$,*

$$u_k^0 = u_k + \gamma_k + \xi(n),$$

where

(B.24) $$\gamma_k = \gamma_k(n) = (\hat{\rho}_n - \rho)y_{k-1}$$

and

(B.25) $$\xi(n) = -\frac{1}{n-1}\left\{\sum_{i=2}^n u_i - (\hat{\rho}_n - \rho)\sum_{i=2}^n y_{i-1}\right\}$$

satisfies

(B.26) $$\xi(n) = O_P\left(\frac{a_n}{n}\right).$$

PROOF. The decomposition of u_k^0 is a matter of simple algebra, so we only need to verify (B.26). By Lemma 3.1, $a_n^{-1}\sum_{i=1}^n u_i \xrightarrow{d} U(1)$ and by Theorem 4.1 of Davis and Resnick (1985), $a_n^{-1}\sum_{i=1}^n y_i \xrightarrow{d} (1-\rho)^{-1}U(1)$. Therefore the first term in the definition of $\xi(n)$ is $O_P(a_n n^{-1})$ and the second term $o_P(a_n n^{-1})$.

REMARK B.1 Unlike in Lemma B.2, the quantity $\max_k |\gamma_k(n)|$ is not of the order $O_P(a_n n^{-1})$ under H_A. Indeed, by Proposition 4.28 of Resnick (1987) $a_n^{-1}\max_{1\le k\le n} y_k$ converges in distribution, so $\max_k |\gamma_k(n)|$ is at least of the order $a_n^{-1}\tilde{a}_n$ which diverges to infinity, see Davis and Resnick (1986), relation (3.3).

LEMMA B.6 *Under the assumptions of Theorem 3.1 b), for every real x*

$$\lim_{n\to\infty} \frac{1}{n-b}\sum_{k=2}^{n-b+1} P(T_{b,k}^0 \le x) \to G(x).$$

PROOF. The proof is exactly the same as for Lemma B.3 except that now we must use Lemma B.5 instead of Lemma B.2. The additional complication comes from the term $\gamma_k(n)$, see Remark B.1. For this reason, a detailed verification of (B.14) is needed.

Observe that now, c.f. (B.18),

$$\text{(B.27)}\qquad R_{b,k}^N = U_{b,k}(1)a_b^{-1}\sum_{i=k}^{k+b-1}(\xi(n)+\gamma_i) + \frac{1}{2}a_b^{-2}\left[\sum_{i=k}^{k+b-1}(\xi(n)+\gamma_i)\right]^2$$

$$+a_b^{-2}\sum_{i=k}^{k+b-1}u_i(\xi(n)+\gamma_i) - \frac{1}{2}a_b^{-2}\sum_{i=k}^{k+b-1}(\xi(n)+\gamma_i)^2.$$

Expanding the terms in (B.27), we see that

$$R_{b,k}^N = R_{b,k}^{N,1} + R_{b,k}^{N,2} + R_{b,k}^{N,3},$$

where

$$R_{b,k}^{N,1} = U_{b,k}(1)a_b^{-1}\sum_{i=k}^{k+b-1}\xi(n) + \frac{1}{2}a_b^{-2}\left[\sum_{i=k}^{k+b-1}\xi(n)\right]^2 + a_b^{-2}\sum_{i=k}^{k+b-1}u_i\xi(n) - \frac{1}{2}a_b^{-2}\sum_{i=k}^{k+b-1}\xi^2(n),$$

$$R_{b,k}^{N,2} = -a_b^{-2}\sum_{i=k}^{k+b-1}\xi(n)\gamma_i - \frac{1}{2}a_b^{-2}\sum_{i=k}^{k+b-1}\gamma_i^2,$$

$$R_{b,k}^{N,3} = a_b^{-1}\sum_{i=k}^{k+b-1}\xi(n)\gamma_i + \frac{1}{2}a_b^{-1}\sum_{i=k}^{k+b-1}\gamma_i^2 + U_{b,k}(1)a_b^{-1}\sum_{i=k}^{k+b-1}\gamma_i + a_b^{-2}\sum_{i=k}^{k+b-1}u_i\gamma_i.$$

The quantity $R_{b,k}^{N,1}$ is equal to $R_{b,k}^N$ in (B.18) with $r_i(n)$ replaced by $\xi(n)$. Since $\xi(n)$ and $\max_i r_i(n)$ are of the same order, c.f. (B.6) and (B.26), it follows as in Lemma B.3 that

$$\limsup_{n\to\infty}\max_{1<k\le n-b+1} P(|R_{b,k}^{N,1}| > \varepsilon) = 0.$$

The quantity $R_{b,k}^{N,2}$ is of a smaller order than the sum of the first two terms in $R_{b,k}^{N,3}$, so it remains to verify that

$$\limsup_{n\to\infty} \max_{1<k\le n-b+1} P\left(\left| a_b^{-1} \sum_{i=k}^{k+b-1} \xi(n)\gamma_i \right| > \varepsilon \right) = 0, \tag{B.28}$$

$$\limsup_{n\to\infty} \max_{1<k\le n-b+1} P\left(\left| a_b^{-1} \sum_{i=k}^{k+b-1} \gamma_i^2 \right| > \varepsilon \right) = 0, \tag{B.29}$$

$$\limsup_{n\to\infty} \max_{1<k\le n-b+1} P\left(\left| U_{b,k}(1) a_b^{-1} \sum_{i=k}^{k+b-1} \gamma_i \right| > \varepsilon \right) = 0, \tag{B.30}$$

and

$$\limsup_{n\to\infty} \max_{1<k\le n-b+1} P\left(\left| a_b^{-2} \sum_{i=k}^{k+b-1} u_i\gamma_i. \right| > \varepsilon \right) = 0. \tag{B.31}$$

Relations (B.28)–(B.31) follow from Lemma B.1 applied as follows: For (B.28), set $S_n = \xi(n)(\hat\rho_n - \rho)$, $U_{k,n} = a_b^{-1}\sum_{i=k}^{k+b-1} y_{i-1}$. For (B.29), observe that

$$a_b^{-1} \sum_{i=k}^{k+b-1} \gamma_i^2 \le \left(a_b^{-2} \sum_{i=k}^{k+b-1} y_{i-1}^2 \right) \left(a_n(\hat\rho_n - \rho)^2 \right)$$

and that

$$a_n(\hat\rho_n - \rho)^2 = n^{-1/\kappa}\tilde L(n) L^{-3}(n) D_n^2 \xrightarrow{P} 0,$$

because D_n is bounded in probability, c.f. (B.2). Therefore, we can use Lemma B.1 with $S_n = a_n(\hat\rho_n - \rho)^2$ and $U_{k,n} = a_b^{-2}\sum_{i=k}^{k+b-1} y_{i-1}^2$. For (B.30), use

$$U_{k,n} = (a_b^{-1} \sum_{i=k}^{k+b-1} u_i)(a_b^{-1} \sum_{i=k}^{k+b-1} y_{i-1})$$

and $S_n = \hat\rho_n - \rho$. Finally, for (B.31), use $S_n = a_b^{-2}\tilde a_b(\hat\rho_n - \rho)$ and $U_{k,n} = \tilde a_b^{-1}\sum_{i=k}^{k+b-1} u_i y_{i-1}$, see Lemma B.7.

Lemma B.7 *Under the assumptions of Theorem 3.1 b),*

$$\tilde a_n^{-1} \sum_{i=1}^{n} u_i y_{i-1} = O_P(1).$$

Proof. The idea of the proof is standard for the sums of regularly varying random variables: we truncate at the level $\tilde a_n$ such that $nP(|u_1 y_0| > \tilde a_n) \sim n^{-1}$ and use Karamata's Theorem, see e.g. Theorem 0.6 of Resnick (1987). We will thus verify that

$$\tilde a_n^{-1} \sum_{i=1}^{n} u_i y_{i-1} \mathbf{1}\{|u_i y_{i-1}| \le \tilde a_n\} = O_P(1) \tag{B.32}$$

and

$$\text{(B.33)} \qquad \tilde{a}_n^{-1} \sum_{i=1}^{n} u_i y_{i-1} \mathbf{1}\{|u_i y_{i-1}| > \tilde{a}_n\} = O_P(1).$$

By Chebyshev's inequality, to verify (B.32) it suffices to show that

$$\limsup_{n\to\infty} \tilde{a}_n^{-2} n E[u_1^2 y_0^2 \mathbf{1}\{|u_1 y_0| \le \tilde{a}_n\}] = 0.$$

Since u_1 and y_0 are independent and satisfy Assumption 3.1, the function $t \mapsto P(u_1^2 y_0^2 > t)$ is regularly varying with index $-\kappa/2 > -1$, see Theorem 3.3 of Cline (1983). We therefore obtain from Karamata's theorem

$$E[u_1^2 y_0^2 \mathbf{1}\{|u_1 y_0| \le \tilde{a}_n\}] = \int_0^{\tilde{a}_n^2} P(u_1^2 y_0^2 > t)dt \sim c_\kappa \tilde{a}_n^2 P(u_1^2 y_0^2 > \tilde{a}_n^2).$$

To deal with (B.33), we use Markov's inequality and Karamata's theorem:

$$\begin{aligned} \tilde{a}_n^{-1} n E[|u_1 y_0| \mathbf{1}\{|u_1 y_0| > \tilde{a}_n\}] &= \tilde{a}_n^{-1} n \left\{ \tilde{a}_n P(|u_1 y_0| > \tilde{a}_n) + \int_{\tilde{a}_n}^{\infty} P(|u_1 y_0| > t) dt \right\} \\ &\sim c_\kappa n P(|u_1 y_0| > \tilde{a}_n). \end{aligned}$$

Lemma B.8 *Under the assumptions of Theorem 3.1 b),* $\mathrm{Var}\left[\hat{G}_b(x)\right] \to 0$.

Proof. The proof is identical to the proof of Lemma B.4 which relies on relation (B.21). By Lemma B.6, (B.21) also holds uder the assumptions of Theorem 3.1 b).

Proof of Theorem 3.1. To show that $\hat{G}_b(x) \xrightarrow{P} G(x)$ it suffices to verify that $E\hat{G}_b(x) \to G(x)$ and $\mathrm{Var}[\hat{G}_b(x)] \to 0$. Thus part a) follows immediately from Lemmas B.3 and B.4 and part b) from Lemmas B.6 and B.8.

Proof of Theorem 4.1. We present only the outline of the argument. The first step is to show that Φ_n defined by (4.3) has a weak limit with a continuous distribution function under the null hypothesis (4.2).

Observe that the estimators $\hat{\mu}_n$ and $\hat{\rho}_n$ defined in (4.1) satisfy

$$\text{(B.34)} \quad \begin{bmatrix} n a_n^{-1} \hat{\mu}_n \\ n(\hat{\rho}_n - 1) \end{bmatrix} = \begin{bmatrix} 1 & n^{-1} a_n^{-1} \sum y_{j-1} \\ n^{-1} a_n^{-1} \sum y_{j-1} & n^{-1} a_n^{-2} \sum y_{j-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} a_n^{-1} \sum u_j \\ a_n^{-2} \sum y_{j-1} u_j \end{bmatrix}.$$

Therefore under the null hypothesis (4.2),

$$\text{(B.35)} \begin{bmatrix} n a_n^{-1} \hat{\mu}_n \\ n(\hat{\rho}_n - 1) \end{bmatrix} \xrightarrow{d} \begin{bmatrix} 1 & \int U(t)dt \\ \int U(t)dt & \int U^2(t)dt \end{bmatrix}^{-1} \begin{bmatrix} U(1) \\ (U^2(1) - V(1))/2 \end{bmatrix} =: \begin{bmatrix} M \\ T \end{bmatrix}.$$

By Lemma 3.1, under (4.2),

$$\text{(B.36)} \qquad a_n^{-2} RSS_0 = a_n^{-2} \sum_{k=2}^{n} u_k^2 \xrightarrow{d} V(1).$$

Observe that

$$\begin{aligned} na_n^{-2}(RSS_0 - RSS_1) &= [n(\hat{\rho}_n - 1)]^2 n^{-1} a_n^{-2} \sum y_{k-1}^2 - \frac{n-1}{n}[na_n^{-1}\hat{\mu}_n]^2 \\ &+ 2n(\hat{\rho}_n - 1)[a_n^{-2} \sum y_{k-1}u_k] + 2[na_n^{-1}\hat{\mu}_n]a_n^{-1}\sum u_k \\ &- 2n(\hat{\rho}_n - 1)[na_n^{-1}\hat{\mu}_n][n^{-1}a_n^{-1}\sum y_{k-1}]. \end{aligned}$$

Therefore by Lemma 3.1, (3.4), (3.5) and (B.35)

(B.37) $$na_n^{-2}(RSS_0 - RSS_1) \xrightarrow{d} T^2 \int U^2(t)dt - M^2$$

$$+T[U^2(1) - V(1)] + 2MU(1) - 2TM \int U(t)dt =: N.$$

By (B.36) and (B.37),

(B.38) $$\Phi_n \xrightarrow{d} [V(1)]^{-1} N.$$

(All convergences are joint because they follow from Lemma 3.1 and the continuous mapping theorem.)

The remainder of the proof is similar to the proof of Theorem 3.1; it involves tedious estimation of numerous remainder terms but no new ideas beyond those presented in the Lemmas leading to and following the proof of Theorem 3.1.

References

Arcones, M. and Giné, E. (1989). The bootstrap of the mean with arbitrary bootstrap sample size. *Annals of the Institute Henri Poincaré*, **22,** 457–481.

Athreya, K. B. (1987). Bootstrap of the mean in the infinite variance case. *The Annals of Statistics*, **15,** number 2, 724–731.

Basawa, I. V., Mallik, A. K., McCormick, W. P. and Taylor, R. L. (1991). Bootstrap test of significance and sequential bootstrap estimation for unstable first order autoregressive processes. *Communications in Statististics – Theory and Methods*, **20,** 1015–1026.

Berkes, I., Horváth, L. and Kokoszka, P. S. (2003). Estimation of the maximal moment exponent of a GARCH(1,1) sequence. *Econometricc Theory*, **19,** 565–586.

Chan, N. H. and Tran, L. T. (1989). On the first order autoregressive process with infinite variance. *Econometric Theory*, **5,** 354–362.

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

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. *The Annals of Statistics*, **14,** number 2, 533–558.

Dickey, D. A. and Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with unit root. *Econometrica*, **49,** 1057–1074.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance.* Springer–Verlag, Berlin.

Guillaume, D. M., Dacorogna, M. M., Dave, R. D., Müller, U. A., Olsen, R. B. and Pictet, O. V. (1997). From the bird's eye to the microscope: a survey of new stylized facts of the intra-daily foreign exchange markets. *Finance and Stochastics*, **1,** 95–129.

Hamilton, J. D. (1994). *Time Series Analysis.* Princeton University Press, Princeton, NJ.

Heimann, G. and Kreiss, J-P. (1996). Bootstrapping general first order autoregression. *Statistics and Probability Letters*, **30,** 87–98.

Horváth, L. and Kokoszka, P. S. (2003). A bootstrap approximation to a unit root test statistic for heavy-tailed observations. *Statistics and Probability Letters*, **62,** 163–173.

Knight, K. (1989). On the bootstrap of the sample mean in the infinite variance case. *The Annals of Statistics*, **17,** 1168–1175.

Ling, S. and Li, W. K. (2003). Asymptotic inference for unit root processes with GARCH(1,1) errors. *Econometric Theory*, **19,** 541–564.

Ling, S. and McAleer, M. (2003). On adaptive estimation in nonstationary ARMA models with GARCH errors. *Annals of Statistics*, **31,** 642–674.

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.

Mittnik, S. and Rachev, S. (2000). *Stable Paretian Models in Finance.* Wiley, New York.

Paparoditis, E. and Politis, D. N. (2003). Residual based block bootstrap for unit root testing. *Econometrica*, **71,** number 3, 813–855.

Philllips, P. C. B. and Xiao, Z. (1998). A primer on unit root testing. Technical Report 972. Cowles Foundation for Research in Economics. Available at `http://korora.econ.yale.edu/phillips.htm`.

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

Resnick, S. I. (1986). Point processes, regular variation and weak convergence. *Advances in Appllied Probability*, **18,** 66–138.

Resnick, S. I. (1987). *Extreme Values, Regular Variation and Point Processes.* Springer-Verlag, New York.

Swensen, A. R. (2003). A note on the power of bootstrap unit root tests. *Econometric Theory*, **19,** 32–48.

Wang, Q., Lin, Y-X. and Gulati, C. M. (2003). Asymptotics for general fractionally integerated processes with applications to unit root tests. *Econometric Theory*, **19,**; Forthcoming.

Zarepour, M. and Knight, K. (1999). Bootsrapping unstable first order autoregressive process with errors in the domain of attraction of stable law. *Communications in Statististics - Stochastic Models*, **15(1),** 11–27.