# City Research Online

## City, University of London Institutional Repository

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

**Permanent repository link:**  https://openaccess.city.ac.uk/id/eprint/13955/

**Link to published version**: https://doi.org/10.1016/j.jeconom.2015.08.006

# Testing for (In)Finite Moments

Lorenzo Trapani*

Cass Business School

August 28, 2015

**Abstract**

This paper proposes a test to verify whether the $k$-th moment of a random variable is finite. We use the fact that, under general assumptions, sample moments either converge to a finite number or diverge to infinity according as the corresponding population moment is finite or not. Building on this, we propose a test for the null that the $k$-th moment does not exist. Since, by construction, our test statistic diverges under the null and converges under the alternative, we propose a randomised testing procedure to discern between the two cases. We study the application of the test to raw data, and to regression residuals. Monte Carlo evidence shows that the test has the correct size and good power; the results are further illustrated through an application to financial data.

**JEL codes**: C12.

**Keywords**: finite moments, randomised tests, Chover-type Law of the Iterated Logarithm, Strong Law of Large Numbers.

---
*Cass Business School, Faculty of Finance, 106 Bunhill Row, London EC1Y 8TZ. Tel.: +44 (0) 207 0405260; email: L.Trapani@city.ac.uk.

# 1 Introduction

An assumption common to virtually all studies in statistics and econometrics is that the moments of a random variable are finite up to a certain order. Existence of population moments is naturally required when computing sample moments. Moment restrictions are also routinely assumed in the various statements of the Law of Large Numbers (LLN) and of the Central Limit Theorem (CLT), thus playing a crucial role in estimation and testing - we refer to Davidson (2002), *inter alia*, for a comprehensive treatment of asymptotic theory. In addition to statistics and econometric theory, several applications in economics and finance require the calculation (and, therefore, the finiteness) of moments. However, a well-known stylised fact, e.g. when using high frequency financial data, is that heavy tails are often encountered (see e.g. Phillips and Loretan, 1994, and a recent contribution by Linton and Xiao, 2013; see also the references therein). Hence the importance of verifying whether assumptions on the finiteness of moments are satisfied.

In order to formally illustrate the problem, let $X$ be a random variable with distribution $F(x)$, and consider the functional $\Xi_X^k(t) \equiv \int_{-t}^{a} |x|^k \, dF(x) + \int_{a}^{t} |x|^k \, dF(x)$, where $a \in (-t, t)$ is finite and the two integrals exist for any $a$. Then the raw absolute moment of order $k$ is defined as

$$E|X|^k \equiv \mu_k = \lim_{t \to \infty} \Xi_X^k(t). \tag{1}$$

It is well known that, when the support of $X$ is not bounded, the integral in (1) needs not be finite, which entails that the $k$-th moment (and of course also moments of order higher than $k$) does not exist. Testing procedures to check for the existence of moments are available, although not always employed. A typical approach (see e.g., in the context of testing for covariance stationarity, Phillips and Loretan, 1991, 1994 and 1995) is based on estimating the so-called "tail index". This usually requires some assumptions on $F(x)$ - typically, it is assumed that the tails of $F(x)$ can be approximated as $L(x) x^{-\gamma}$, where $L(x)$ is a slowly varying function. The parameter $\gamma$ is referred to as the "tail index", and

it is related to the highest finite moment of $X$ - formally, this means that

$$\lim_{t\to\infty} \Xi_X^k(t) \begin{cases} = \infty \\ < \infty \end{cases} \text{according as} \quad \begin{matrix} k \geq \gamma \\ k < \gamma \end{matrix} . \tag{2}$$

Hence, one could use an estimate of $\gamma$ in order to test for the null hypothesis that $\gamma > k$, which is tantamount to testing for $H_0 : E|X|^k < \infty$. A routinely employed technique is the Hill estimator (Hill, 1975), or some variants thereof; we refer to Embrechts, Kluppelberg and Mikosch (1997) and de Haan and Ferreira (2006) for excellent reviews which also consider several improvements of the original Hill estimator. In general, however, estimation of $\gamma$ is fraught with difficulties. Considering the Hill estimator as a leading example, it is well known that its rate of convergence may be relatively slow: indeed, this is a common feature to all tail index estimators. Moreover, the quality of the Hill estimator depends crucially on selecting the appropriate number of order statistics - see Section 3.2, for details, and in particular the discussion after equation (21). If this is not chosen correctly, the Hill estimator can yield very poor inference; Resnick (1997) provides an insightful discussion of the main pitfalls of the Hill estimator, and also several possible variants to overcome such pitfalls.

*Hypotheses of interest and the main result of this paper*

In this paper, we propose a test for the null that the $k$-th raw moment of $X$ does *not* exist; formally, we develop a test for

$$\begin{cases} H_0 : \lim_{t\to\infty} \Xi_X^k(t) = \infty \\ H_A : \lim_{t\to\infty} \Xi_X^k(t) < \infty \end{cases} . \tag{3}$$

We base our analysis on the divergent part of the Strong LLN (SLLN). Defining the $k$-th sample moment, based on the sample $\{x_i\}_{i=1}^n$, as

$$\hat{\mu}_k \equiv \frac{1}{n} \sum_{i=1}^n |x_i|^k , \tag{4}$$

3

as $n \to \infty$ it holds that, almost surely

$$
\hat{\mu}_k \to \begin{cases} \infty \\ < \infty \end{cases} \quad \text{according as} \quad \begin{array}{l} \lim_{t \to \infty} \Xi_X^k(t) = \infty \\ \lim_{t \to \infty} \Xi_X^k(t) < \infty \end{array} . \tag{5}
$$

Based on (5), we use $\hat{\mu}_k$ to test for $H_0 : \lim_{t \to \infty} \Xi_X^k(t) = \infty$ in (3).

The literature has proposed several contributions that use (5), both for the purpose of estimating $\gamma$ and for conducting hypothesis testing. As far as the former is concerned, Meerschaert and Scheffler (1998; see also the related contribution by McElroy and Politis, 2007, and the references therein) exploit the generalised version of the CLT to propose a moment-based estimator of $\gamma$. As far as the latter issue (hypothesis testing) is concerned, Fedotenkov (2013; see also the related papers by Fedotenkov, 2015a and 2015b) develops a bootstrap-based methodology whose main idea is closely related to the contribution of the present paper. In particular, Fedotenkov (2013) proposes comparing two statistics: the full-sample estimator of $\mu_k$, and a subsample based one. Under the null hypothesis that $\mu_k$ is finite, both statistics would converge to $\mu_k$ by virtue of (5). Conversely, under the alternative that $\mu_k$ is not finite, the two statistics diverge at a different rate. Building on this, the test proposed by Fedotenkov (2013) is essentially based on comparing (by means of the bootstrap) the two statistics, checking whether their difference is bounded or diverges.

In the context of this paper, (5) is employed in order to test for the null hypothesis that $\mu_k$ does *not* exist. From a technical point of view, however, (5) is not used directly; rather, the main results in the paper hinge on a version of the Law of the Iterated Logarithm (LIL) for random variables that do not admit a finite first absolute moment, known in the literature as the "Chover-type LIL" (Chover, 1966). Thus, an ancillary contribution of this paper is the development of a Chover-type LIL for dependent data. From a methodological point of view, the results in this paper share, with the works cited above, the (desirable) feature of not having to determine an optimal number of order statistics to carry out inference, which is one of the main problems of the Hill estimator. However, note that, under the null hypothesis of an infinite $k$-order moment, there is no random-

ness in (5): the statistic $\hat{\mu}_k$ does not converge to any distribution (it diverges to positive infinity), and it cannot be used directly in order to conduct the test. Consequently, we employ a randomised testing procedure, which builds on a contribution by Pearson (1950). From a conceptual point of view, such approach is based on the idea that, when a statistic does not have randomness under the null (e.g. because it diverges) or when it has a non standard limiting distribution, randomness can be added by the researcher. Corradi and Swanson (2006) and Bandi and Corradi (2014) have recently employed randomised testing procedures. In particular, Bandi and Corradi (2014) propose a test to evaluate rates of divergence, which, albeit in a very different context, is essentially the same problem investigated in this paper. As far as conducting inference is concerned, we follow the approach used in Corradi and Swanson (2006), where randomisation is employed in conjunction with sample conditioning. This entails adding randomness to the basic statistic, and then deriving the asymptotics conditional on the sample, showing that limiting distribution and consistency hold for all samples save for a set of zero measure. Such approach is somehow akin to bootstrap based inference, which is also carried out conditional on the sample - although using bootstrap in this context would be problematic, e.g. due to the difficulties in extending the theory to the case of data with infinite first moment (see Cornea-Madeira and Davidson, 2014). A key difference with bootstrap-based inference is the interpretation that the notion of test size has in this context. Indeed, it is well known that, in a classical hypothesis testing context, the level $\alpha$ of a test means that, if a researcher applies the test $B$ times and the null is valid, then (s)he will reject the null with frequency $\alpha$ - that is, (s)he will be wrong $\alpha B$ times. Conversely, as illustrated by Corradi and Swanson (2006), in this context $\alpha$ is interpreted thus: out of $J$ researchers who apply the test, $\alpha J$ of them will reject the null when this is true. Despite such interpretational difference, as we show in Section 2, using this approach we overcome the issue of $\hat{\mu}_k$ diverging under the null, and we obtain a test statistic which, for a given level $\alpha$, rejects the null with probability $\alpha$ when true, and with probability 1 when false.

The remainder of the paper is organised as follows. In Section 2, we discuss the test, its theoretical properties (null distribution and consistency), and possible extensions to

regression residuals (Section 2.1). Section 3 contains, in addition to a set of guidelines on how to use the test (Section 3.1), a Monte Carlo exercise (Section 3.2), and an application (Section 3.3). Section 4 concludes. Proofs are in Appendix.

NOTATION We denote the ordinary limits as "$\rightarrow$"; convergence in distribution as "$\overset{d}{\rightarrow}$"; convergence in probability and almost surely as "$\overset{p}{\rightarrow}$" and "$\overset{a.s.}{\rightarrow}$" respectively. We use "a.s." as short-hand for "almost surely", "i.o." for "infinitely often", and "$\equiv$" for definitional equality. Finite constants that do not depend on the sample size are denoted as $M$, $M'$, ..., etc. Other relevant notation is introduced in the remainder of the paper.

## 2   The test

This section contains a description of how the test statistic is constructed, and its theoretical properties (reported in Theorems 1 and 2). In Section 2.1, we study the application of the test to regression residuals.

We start by reporting the testing procedure as a four step algorithm.

**Step 1** Compute $\hat{\mu}_k$.

**Step 2** Randomly generate an *i.i.d.* $N(0,1)$ sample of size $r$, say $\{\xi_j\}_{j=1}^{r}$, and define the sample $\left\{\sqrt{e^{\hat{\mu}_k}} \times \xi_j\right\}_{j=1}^{r}$.

**Step 3** Generate the sequence $\{\zeta_{j,n}(u)\}_{j=1}^{r}$ as

$$\zeta_{j,n}(u) \equiv I\left[\sqrt{e^{\hat{\mu}_k}} \times \xi_j \leq u\right], \tag{6}$$

for all $j$, where $u \neq 0$ is any real number and $I[\cdot]$ is the indicator function. The values of $u$ can be selected from a density $\varphi(u)$ on a bounded support $U = [\underline{u}, \bar{u}]$.

**Step 4** For each $u \in U \backslash \{0\}$, define

$$\vartheta_{nr}(u) \equiv \frac{2}{\sqrt{r}} \sum_{j=1}^{r} \left[\zeta_{j,n}(u) - \frac{1}{2}\right], \tag{7}$$

6

and the test statistic

$$\Theta_{nr} \equiv \int_{\underline{u}}^{\overline{u}} \vartheta_{nr}^2 (u) \, \varphi (u) \, du. \tag{8}$$

The following remarks contain comments on the specifications of the test, and a heuristic preview of how the test works; the choice of the artificial sample size, $r$, defined in Step 2, is discussed after Theorems 1 and 2, and in Section 3.1.

Note also that the statistic $\hat{\mu}_k$ is bound to be sensitive to the unit of measurement; thus, a scale invariant transformation thereof should be employed instead - for the sake of a concise discussion, we assume henceforth that $\hat{\mu}_k$ is scale-free; in Section 3.1, we explore ways in which scale invariance can be obtained.

**2.1** A heuristic description of the main idea of Step 3 is the following. Under the null hypothesis $H_0$ that $E\,|X|^k$ does not exist, as $n \to \infty$, $\sqrt{e^{\hat{\mu}_k}} \times \xi_j$ should follow a normal distribution with mean zero and infinite variance. This entails that, under $H_0$ with $n \to \infty$, for any real number $u$, the random variable $\zeta_{j,n}(u)$ has a Bernoulli distribution with $P\left[\zeta_{j,n}(u) = 1\right] = \frac{1}{2}$. Therefore, under $H_0$ as $n \to \infty$, $\zeta_{j,n}(u)$ has mean $\frac{1}{2}$ and variance $\frac{1}{4}$. Conversely, under the alternative that $E\,|X|^k < \infty$, $e^{\hat{\mu}_k}$ converges to a finite value. Hence, $\sqrt{e^{\hat{\mu}_k}} \times \xi_j$ should follow a normal distribution with mean zero and finite variance, so that, for any $u \neq 0$, $E\left[\zeta_{j,n}(u)\right] \neq \frac{1}{2}$.

**2.2** Step 4 follows directly from Step 3, and it is an application of the CLT. It can be expected that, under the null with $n \to \infty$ and $r \to \infty$, $\vartheta_{nr}(u) \xrightarrow{d} N(0,1)$ for every choice of $u$. Conversely, under the alternative that $E\,|X|^k < \infty$, the $\zeta_{j,n}(u)$s do not have mean $\frac{1}{2}$ and therefore a CLT does not apply to (7). Given that in (7) there is a sum involving a sequence with non-zero mean, it can be expected that $\vartheta_{nr}(u)$ diverges at a rate $\sqrt{r}$. This ensures the consistency of the test under $H_A$.

**2.3** In Step 3, we consider the possibility that several values of $u$ could be tried. The definition of $\Theta_{nr}$ in Step 4 is based on combining a continuous set of values of $\vartheta_{nr}(u)$, attaching a different weight to each $u$ according to some density $\varphi(u)$. Monte Carlo

7

evidence (Section 3.2) shows that choosing $U = \{-1, 1\}$ with equal probability works well under any scenario. From a theoretical point of view, it can be expected that the width of $U$ is positively related to both power and size: as it increases, the power versus the alternative that $E|X|^k < \infty$ will increase, but there will also be some size distortion under the null.

We now lay out the main assumptions on dependence and tail behaviour. Prior to that, recall the definition of *uniform mixing* (see e.g. Davidson, 2002, p. 209). Let $(\Omega, \mathcal{F}, P)$ denote the probability space on which $\{x_i\}_{i=1}^n$ is defined, and let $\mathcal{F}_j^{j+k} \equiv \sigma(X_i, j \le i \le j + k)$ and the sets $A \in \mathcal{F}_1^k$ and $B \in \mathcal{F}_{k+m}^\infty$; finally, define

$$\phi_m \equiv \sup_{k \in \mathbb{N}} |P(B|A) - P(A)|. \tag{9}$$

Then, $\{x_i\}_{i=1}^n$ is said to be *uniformly mixing* if $\phi_m = 0$ as $m \to \infty$.

Consider the following assumption.

**Assumption 1.** *(i)* $\{x_i\}_{i=1}^n$ is a uniformly mixing sequence with mixing numbers $\phi_m = Mm^{-\theta}$ for some $\theta > 0$; *(ii)* the $x_i$s have common distribution $F(x)$ which belongs in the domain of attraction of a stable distribution with tail exponent $\gamma \in (0, k]$, viz.

$$1 - F(x) = \frac{c_1(x) + o(1)}{x^\gamma} L(x) \text{ and } F(-x) = \frac{c_2(x) + o(1)}{x^\gamma} L(x), \tag{10}$$

as $x \to \infty$, with: $L(x) \ge 0$ slowly varying at infinity in the Karamata sense, $c_i(x) \ge 0$ and $c_1 + c_2 > 0$ where $c_i = \lim_{x \to \infty} c_i(x)$ for $i = 1, 2$.

Part *(i)* of the assumption imposes some structure on the dependence of the data - having $\theta > 0$ is a very mild requirement on the amount of memory allowed in the data. As mentioned in the Introduction, the main tool employed in the proofs - which is also one of the contributions of this paper - is a Chover-type LIL for uniformly mixing sequences. In addition to considering dependence (with a quite flexible amount of memory, since only $\theta > 0$ is required) uniform mixing affords great analytical tractability. Other forms of

dependence could also be considered, as long as Chover's LIL holds - see e.g. the results in Trapani (2014) for the case of strongly mixing data.

Part *(ii)* of the assumption contains the null hypothesis, represented by having $\gamma \leq k$. The exact specification of the slowly varying function $L(x)$ is not required, and thus, basically, part *(ii)* is needed only to ensure that some moments of $X$ exist.

Defining $P^*$ as the probability law of $\{\zeta_{j,n}(u)\}_{j=1}^r$ conditional on the sample, let "$\xrightarrow{d^*}$" denote convergence in distribution according to $P^*$.

The limiting distribution of $\Theta_{nr}$ under the null is given in the following Theorem.

**Theorem 1** *Let Assumption 1 hold. Under $H_0 : \lim_{t\to\infty} \Xi_X^k(t) = \infty$, as $(n, r) \to \infty$ with*

$$\frac{r}{\exp\left(n^{\frac{k}{\gamma}-1} \ln^{\frac{k}{\gamma}} n\right)} \to 0, \tag{11}$$

*it holds that $\Theta_{nr} \xrightarrow{d^*} \chi_1^2$ a.s. conditionally on the sample.*

Theorem 1 stipulates that $\Theta_{nr}$ has, under the null, a chi-squared distribution with one degree of freedom, as can be expected from the discussion above. Convergence to the null limiting distribution requires both $n \to \infty$ and $r \to \infty$. As far as the latter is concerned, $r$ needs to pass to infinity subject to (11), in order to ensure that a CLT holds. Since the test statistic is based on an *i.i.d.* sequence of uniformly distributed random variables, it can be expected that convergence should be quite fast, and therefore in practice $r$ is not needed to be too large.

We now consider the consistency of the test versus the alternative that the $k$-th moment exists.

**Theorem 2** *Let Assumption 1 hold with $\gamma = k + \varepsilon$ for some $\varepsilon > 0$ in part (ii). Define $c_\alpha$ as $P^*[\Theta_{nr} \leq c_\alpha] = \alpha$ under $H_0$. Under $H_A : \lim_{t\to\infty} \Xi_X^k(t) < \infty$, as $(n, r) \to \infty$ with the restriction (11), it holds that $P^*[\Theta_{nr} > c_\alpha] = 1$ a.s. conditionally on the sample if*

$$\lim_{(n,r)\to\infty} \frac{r}{\exp(\mu_k)} = \infty. \tag{12}$$

9

Theorem 2 states that tests based on $\Theta_{nr}$ have non trivial power versus the alternative that $E|X|^k$ exists. In the proof, we show that, under $H_A$, $\vartheta_{nr}(u)$ has a non-centrality parameter proportional to $\sqrt{\frac{r}{\exp(\mu_k)}}$, whence (12).

## 2.1 Application to regression residuals

We now turn to discussing the application of the test to regression residuals. Indeed, this is a natural application, since the existence of moments up to a certain order is routinely assumed for the error term in the regression model

$$y_i = \alpha + \beta'x_i + \epsilon_i; \tag{13}$$

for example, a typical assumption in (13) is that the $\epsilon_i$s have finite second moment, i.e. $E|\epsilon_i|^2 < \infty$. In order to verify the validity of this (and similar) assumptions, we show under which conditions the test developed above can be applied to the OLS residuals computed from (13).

Henceforth, we denote the distribution of $\epsilon_i$ as $F_\epsilon(x)$, and define the OLS residuals $\hat{\epsilon}_i$. Also, let $k \geq 1$ and $\Xi_\epsilon^k(t) \equiv \int_{-t}^a |x|^k dF_\epsilon(x) + \int_a^t |x|^k dF_\epsilon(x)$ with both integrals existing for any finite $a \in (-t, t)$; we propose to test for

$$\begin{cases} H_0 : \lim_{t\to\infty} \Xi_\epsilon^k(t) = \infty \\ H_A : \lim_{t\to\infty} \Xi_\epsilon^k(t) < \infty \end{cases}, \tag{14}$$

by using

$$\hat{\mu}_k^\epsilon \equiv \frac{1}{n}\sum_{i=1}^n |\hat{\epsilon}_i|^k,$$

or a scale invariant transformation. Using $\sqrt{e^{\hat{\mu}_k^\epsilon}}$, we define the corresponding test statistics as $\Theta_{nr}^\epsilon$.

Consider the following extension of Assumption 1.

**Assumption 1\*.** *(i)* Assumption 1 holds for $\{\epsilon_i\}_{i=1}^n$, with tail index $\gamma_\epsilon$; *(ii)* Assumption 1 holds for $\{x_i\}_{i=1}^n$, with tail index $\gamma_x = k + \varepsilon_x$ for some $\varepsilon_x > 0$; *(iii)* $\{x_i\}_{i=1}^n$ and

$\{\epsilon_i\}_{i=1}^n$ are two mutually independent groups; *(iv)* $x_i \neq 0$ a.s. for $1 \leq i \leq n$.

It holds that:

**Corollary 1** *Let $k \geq 1$. Under $H_0 : \lim_{t\to\infty} \Xi_\epsilon^k(t) = \infty$, if Assumption 1\*(i) holds, then as $(n, r) \to \infty$ with (11), it holds that $\Theta_{nr}^\epsilon \xrightarrow{d^*} \chi_1^2$. Under the alternative $H_A : \lim_{t\to\infty} \Xi_\epsilon^k(t) < \infty$ - corresponding to the case $\gamma_\epsilon = k + \varepsilon_\epsilon$ with $\varepsilon_\epsilon > 0$ - if Assumption 1\* holds, then, as $(n, r) \to \infty$ with (11) and (12), it holds that $P^*[\Theta_{nr}^\epsilon > c_\alpha] = 1$.*

Corollary 1 shows that the test can be extended to regression residuals, with the same null distribution and consistency properties as above. The only proviso is that the regressors have enough moments: in essence, the corollary requires that $E|x_i|^k < \infty$. As shown in the proof, this is needed in order for the test to have power: the regression residuals $\hat{\epsilon}_i$ contain $x_i$, and therefore even when $E|\epsilon_i|^k < \infty$, if $E|x_i|^k = \infty$ this may entail that $\hat{\mu}_k^\epsilon$ diverges, thereby yielding zero power. From a methodological point of view, therefore, the test ought to be applied to residuals after checking whether $E|x_i|^k = \infty$ or not.

Note that the test is designed for $k \geq 1$ only. When $k < 1$, the test may still work, but this depends on the maximum moments of both $x_i$ and $\epsilon_i$; the reason why the test may fail is that, under the alternative, the OLS estimator of $\beta$ in (13) may be inconsistent, and even diverge to positive infinity - this is e.g. the case if $E|x_i|^{2+\varepsilon} < \infty$ for some $\varepsilon > 0$; see also Cline (1989).

# 3 Applying the test: guidelines, simulations and empirical applications

In this section, we study three inter-related issues concerning how to apply the test. Firstly, we discuss how to make the test statistic scale invariant - Section 3.1. Secondly, we evaluate the properties of our test through a simulation exercise, also proposing some guidelines as to how to choose some of the test specifications - Section 3.2. Finally, we illustrate the implementation of the test and the interpretation of its results by means of an application to financial data - Section 3.3.

## 3.1 Scale invariance

As pointed out in Section 2, the raw absolute moment of order $k$, $\hat{\mu}_k$, cannot be employed directly as it is not scale invariant. Hence, the need for a re-scaling of the statistic $\hat{\mu}_k$: although several alternatives can be proposed, as a general rule it is necessary to have a scaling factor that does not diverge (or diverges at a slower rate than $\hat{\mu}_k$) under the null, and is bounded under the alternative.

Several choices are possible; in particular, based on Lemma 1 in Appendix, we propose the following family of scaling factors

$$\hat{\mu}_\psi = \frac{1}{n} \sum_{i=1}^{n} |x_i|^\psi, \tag{15}$$

where $\psi$ is chosen such that $\psi \in (0, k)$; thence, the re-scaled test statistic is

$$\hat{\mu}_k^* = \frac{\hat{\mu}_k}{\hat{\mu}_\psi^{k/\psi}}. \tag{16}$$

The test could then be carried out exactly as in Section 2, using $\hat{\mu}_k^*$ instead of $\hat{\mu}_k$. Although in principle any choice of $\psi \in (0, k)$ can be employed, "natural" choices are $\psi = 1$ or $\psi = 2$ (i.e., using the variance of the data) - in the simulations, we employ $\psi = 2$, which proves to be a good choice in terms of empirical rejection frequencies; only minor changes are noted, anyway, when setting $\psi = 1$. Whilst $\hat{\mu}_k^*$ fulfills the purpose of making $\hat{\mu}_k$ scale invariant, in the simulations we also found that, at least in finite samples, better results are achieved by further rescaling $\hat{\mu}_k^*$ by the value it would have if $X$ were normally distributed. Letting $\mu_k^{(N)}$ be the $k$-th absolute moment of a standard normal, the test statistic is thus based on

$$\hat{\mu}_k^{**} = \hat{\mu}_k^* \times \frac{\left[\mu_\psi^{(N)}\right]^{k/\psi}}{\mu_k^{(N)}}. \tag{17}$$

The rationale for employing $\hat{\mu}_\psi$ can be illustrated as follows. Based on Lemma 1 in

12

Appendix, and modulo a slowly varying function, it holds that, under $H_0$

$$\hat{\mu}_\psi = \begin{cases} O_{a.s.}(1) & \psi < \gamma \\ O_{a.s.}\left(n^{\frac{\psi}{\gamma}-1} \ln^{\frac{\psi}{\gamma}} n\right) & \psi \geq \gamma \end{cases} \quad \text{according as} \quad ;$$

therefore

$$\hat{\mu}_\psi^{k/\psi} = \begin{cases} O_{a.s.}(1) & \psi < \gamma \\ O_{a.s.}\left[n^{\frac{k}{\psi}\left(\frac{\psi}{\gamma}-1\right)} \ln^{\frac{k}{\gamma}} n\right] & \psi \geq \gamma \end{cases} \quad \text{according as} \quad .$$

As can be seen, in either case the rate of divergence of the scaling factor $\hat{\mu}_\psi^{k/\psi}$ is always smaller than that of $\hat{\mu}_k$. Since, under $H_0$, $\hat{\mu}_k = O_{a.s.}\left(n^{\frac{k}{\gamma}-1} \ln^{\frac{k}{\gamma}} n\right)$, we have

$$\hat{\mu}_k^* = \begin{cases} O_{a.s.}\left[n^{\frac{k}{\gamma}-1} \ln^{\frac{k}{\gamma}} n\right] & \psi < \gamma \\ O_{a.s.}\left[n^{\frac{k}{\psi}-1}\right] & \psi \geq \gamma \end{cases} \quad \text{according as} \quad , \tag{18}$$

and the same holds for $\hat{\mu}_k^{**}$. Under the alternative, $\hat{\mu}_\psi$ is bounded by the SLLN, so that $\hat{\mu}_k^*$ is also bounded.

Finally, we return to the issue of selecting the size of the artificial sample, $r$. Equations (11) and (18) can be combined in order to choose $r$; after standard algebra, it can be verified that it must hold

$$\begin{cases} \dfrac{r}{\exp\left(n^{\frac{k}{\gamma}-1} \ln^{\frac{k}{\gamma}} n\right)} \to 0 & \psi < \gamma \\ \dfrac{r}{\exp\left(n^{\frac{k}{\psi}-1}\right)} \to 0 & \psi \geq \gamma \end{cases} \quad \text{according as} \quad . \tag{19}$$

When $k > \gamma$, any choice such that $r$ is a polynomial function of $n$ will be appropriate. The case $k = \gamma$ is of special interest: in such case, (19) boils down to requiring $\frac{r}{n} \to 0$. Thus, choosing $r = o(n)$ always satisfies (19).

## 3.2 Monte Carlo simulations

We generate $n + 1000$ datapoints from an $ARMA(1,1)$ process, discarding the first 1000 observations to avoid dependence on initial conditions:

$$y_i = \rho y_{i-1} + \epsilon_i + \theta \epsilon_{i-1}, \tag{20}$$

with $(\rho, \theta) = \{(0,0), (0.5,0), (0,0.5), (0,-0.5)\}$; the sample sizes considered are $n \in \{100, 250, 500, 1000, 10000, 100000\}$.

The innovation $\epsilon_i$ is generated according to two schemes. In the first set of experiments, we generate data according to a Student t distribution with $\upsilon$ degrees of freedom $(t_\upsilon)$; such distribution is often found to be good at capturing the features of financial data (see e.g. Hurst and Platen, 1997; and Markowitz and Usmen, 1996a, 1996b), and therefore the results from this experiment should provide a set of guidelines for the applied user; indeed, we analyse the impact, on the test, of several specifications under this distributional design. We also consider a second set of experiments, where $\epsilon_i$ is generated as having a power law, as a robustness check to assess how the test responds to a different distribution; data are generated according to standard procedures, and we refer to e.g. Clauset, Shalizi and Newman (2009).

As far as the testing problem in concerned, without loss of generality, we consider testing for the existence of the fourth moment, viz.

$$\begin{cases} H_0 : \lim_{t \to \infty} \Xi_X^4(t) = \infty \\ H_A : \lim_{t \to \infty} \Xi_X^4(t) < \infty \end{cases}.$$

Let $\gamma$ denote the degree of freedom of the Student t distribution data are drawn from in the first set of experiments; and the tail index of the power law from which data are generated in the second set of experiments. We set $\gamma \in \{2, 3, 4, 5, 6\}$: the first three values are used to assess the size of the test, and the last two values are used in order to assess the power.

Based on Corollary 1, we apply the test to pre-whitened data, by fitting an $AR(7)$

14

model to $y_i$ and then applying the test to the residuals; unreported simulations show that, when applying the test directly to the raw data, this results in a massive oversizement when there is dependence - hence, a guideline is that the test ought to be applied to pre-whitened data. In order to make the test statistic scale invariant, we use, as suggested in Section 3.1

$$\hat{\mu}_4^{**} = \frac{1}{3} \times \frac{\hat{\mu}_4}{\hat{\mu}_2^2}.$$

As far as the test specifications are concerned, on account of Theorems 1 and 2, it can be expected that the empirical rejection frequencies will be affected by the size of the artificial sample, $r$, and by the values of $u$ employed (that is, by the set $U$). As regards the former, in the proof of Theorem 2 we show that, under the alternative, the test statistic has a noncentrality parameter proportional to $\sqrt{\frac{r}{\exp(\mu_k)}}$: thus, a large $r$ is bound to increase the power of the test. This is in a trade-off with (11), which indicates that a large value of $r$ will yield size distortion; of course, this is valid for finite samples, since asymptotically the test will have the correct size as long as (11) is satisfied. Similarly, the width of $U$ also has an impact on power and size. Indeed, as shown in the proof of Theorem 2, under the alternative the test has a noncentrality parameter which increases as the width of $U$ increases: hence, using large values of $u$ should boost the power, at the expense of size. In order to analyse the impact of $r$ and $U$ on the size and power of the test, we run four different experiments for the leading case of Student t data. In the first, benchmark case, we set $r = n^{\frac{4}{5}}$ and $U = \{-1, 1\}$, with each value drawn with equal probability of $\frac{1}{2}$; this proved to be the best choice for all cases considered. We also consider the cases $r = n^{\frac{1}{2}}$ and $U = \{-1, 1\}$, and $r = n^{\frac{4}{5}}$ and $U = \{-2, 2\}$. In the former case, the test is expected to be less powerful, whereas in the latter higher power should be observed, in presence of size distortion, at least for small samples. Finally, we also report the empirical rejection frequencies for the intermediate case $r = n^{\frac{1}{2}}$ and $U = \{-2, 2\}$. Results should look similar as $n \to \infty$, and in this respect having $n = 100000$ in the simulations should shed some light on the asymptotic performance of the test.

By way of comparison, we also report a set of experiments to determine the size and power of a test for $H_0 : \gamma \leq 4$ based on a direct estimate of the tail index $\gamma$. Based on Hill

(1975), the estimator of $\gamma$, say $\hat{\gamma}_{Hill}$, is calculated as

$$\hat{\gamma}_{Hill} = \left[ \frac{1}{h} \sum_{i=1}^{h} \ln \frac{\hat{\epsilon}_{(n-h+i)}}{\hat{\epsilon}_{(n-h)}} \right]^{-1}, \tag{21}$$

where $\hat{\epsilon}_{(s)}$ denotes the $s$-th order statistic (in descending order) of the sample of the absolute values of the residuals from fitting an $AR(7)$ model to the $y_i$s, i.e. from $\{|\hat{\epsilon}_i|\}_{i=1}^{n}$. The estimator is applied to the residuals of the $AR(7)$ model, rather than directly to the data: according to Embrechts, Kluppelberg and Mikosch (1997; see in particular Figure 5.5.4 on p. 270), this is an effective way of attenuating the impact of serial dependence in the data. As far as the threshold $h$ is concerned, we use $h = \lfloor n^{3/4} / \ln n \rfloor$. We found this to be the best choice for $h$; note that, in the second set of experiments, data are generated according to a strict Pareto model for the tails, and therefore in that case we can expect the Hill estimator to be unbiased. Alternatively, one could employ data-driven rules to select the optimal $h$ - we refer to e.g. Drees and Kaufmann (1998) and the references therein. Since the data can be expected to be still dependent even after pre-whitening, we base the test on the statistic $\sqrt{h} \frac{\hat{\gamma}_{Hill} - 4}{s}$, where $s^2$ is the inverse of the long-run variance estimator proposed in Hill (2010); we refer thereto for details on the implementation, and point out here that we have employed the Bartlett kernel, and set the bandwidth to $b = \lfloor n^{1/5} \rfloor$, based on the condition $b = o\left(h n^{-1/2}\right)$. Under the null, it holds that

$$\sqrt{h} \frac{\hat{\gamma}_{Hill} - 4}{s} \xrightarrow{d} N(0, 1). \tag{22}$$

We point out that the test based on $\hat{\gamma}_{Hill}$ is not meant to be the only possible alternative to our approach. Indeed, the performance of the Hill estimator can be quite poor, and various improvements have been suggested - see de Haan and Ferreira (2006). Rather, we would suggest to interpret the test based on $\hat{\gamma}_{Hill}$ as a naive benchmark. It is however worth noting that the applied literature customarily uses this approach to verify whether the data have finite moments of a certain order or not (e.g. Phillips and Loretan, 1991, 1994 and 1995).

Tables 1a-1c contains empirical rejection frequencies for the cases of data having a

16

Student t and a power law distribution, respectively. The number of simulations has been set equal to 1000, so that, when evaluating the size of the test, the empirical rejection frequencies should lie in the interval $[0.036, 0.064]$.

**[Insert Tables 1a-1c somewhere here]**

The effect of pre-whitening (in both cases: Student t and, as shown in Table 1c below, power law) is that the test is nearly unaffected by the presence of serial correlation: size and power almost do not change across the various combinations of $(\rho, \theta)$.

As far as size is concerned, consider first the benchmark case where $r = n^{\frac{4}{5}}$ and $U = \{-1, 1\}$ - Table 1a. The test has the correct size in both non-boundary cases $\gamma = 2$ and $\gamma = 3$. As could be expected, the test exhibits higher empirical rejection frequencies when $\gamma = 4$; this, however, attenuates when $n \geq 500$, with the test having the correct size again in all cases considered. Turning to the power (cases $\gamma = 5$ and $\gamma = 6$), this is higher than 50% for all cases considered when $n \geq 500$, and anyway higher than the 50% threshold for $n \geq 250$ in the non-boundary case $\gamma = 6$: the power increases monotonically with $n$ and $\gamma$, both features being in line with what can be expected. As mentioned above, these specifications (for $r$ and $U$) correspond to the best results under all scenarios considered: thus, a guideline from Table 1a and, in general, from this section is to choose $r$ quite close to $n$, and use $U = \{-1, 1\}$.

The other cases, displayed in Table 1b, complement the conclusions above. Results were rather similar across the various combinations of $(\rho, \theta)$, and therefore only the case with no serial dependence is reported for all three set-ups. The results in the table are in line with the theory: chiefly in the case $\gamma = 4$, decreasing $r$ reduces the empirical rejection frequency. The power, on the other hand, is reduced quite substantially, and the test has power higher than 50% only for $n \geq 10000$. Conversely, increasing the width of $U$ boosts the power, but the test appears massively oversized, especially for the case of $\gamma = 4$, where the correct size is attained only when $n \geq 100000$. The intermediate case $r = n^{\frac{1}{2}}$ and $U = \{-2, 2\}$ is the most similar to the case of $r = n^{\frac{4}{5}}$ and $U \in \{-1, 1\}$, although the power is slightly lower and the size, when $\gamma = 4$, is never correct unless $n \geq 10000$.

Turning to the case of data following a power law (Table 1c), there are few instances of oversizement when $n$ is small ($n = 100$) and $\gamma$ is equal to 3, but such tendency is relatively infrequent and it disappears for larger sample sizes. When $\gamma = 4$, the empirical rejection frequencies are higher than in the Student t case, with oversizement attenuating when $n \geq 1000$. Further, as far as power is concerned, the test is less powerful, for large samples, than with Student t data. This can be further considered in conjunction with the high rejection rates when $\gamma = 4$.

Note, finally, that tests based on the Hill estimator have lower power, even for relatively large $n$ - the power does increase above 50% for $n \geq 10000$ when data follow a Student t distribution, whereas it tends to be lower when the data follow a power law.

## 3.3 Application

In this section we illustrate how the test works through an empirical application to financial data. We consider daily returns from 3rd January, 2008, until 30th September, 2013, which corresponds to a sample size of $n = 1499$. We consider two groups of stocks from the FTSE 100: the banking sector (5 stocks) and the financial services sector (4 stocks), for a total of 9 stocks. A list of the constituents is in Tables 2a-2b.

We test for the existence of the first, second, third and fourth moment. In particular, letting $y_{j,i}$ be the return on stock $j$ at day $i$, we use the following test statistics to verify the existence of the first four moments

$$
\begin{aligned}
\hat{\mu}_{1,j} &= \frac{1}{n} \sum_{i=1}^{n} |y_{j,i}|, \\
\hat{\mu}_{2,j}^{**} &= \left(\sqrt{\frac{2}{\pi}}\right)^{2} \times \frac{\frac{1}{n}\sum_{i=1}^{n} y_{j,i}^{2}}{\left[\frac{1}{n}\sum_{i=1}^{n} |y_{j,i}|\right]^{2}}, \\
\hat{\mu}_{k,j}^{**} &= \frac{1}{\mu_{k}^{(Y)}} \times \frac{\frac{1}{n}\sum_{i=1}^{n} |y_{j,i}|^{k}}{\left[\frac{1}{n}\sum_{i=1}^{n} y_{j,i}^{2}\right]^{k/2}}, \text{ for } k = 3, 4.
\end{aligned}
$$

Based on the results from the simulations, the test is applied to pre-whitened data, using $U = \{-1, 1\}$.

Results are reported in Tables 2a-2b:

Tables 2a-2b contain, in addition to the tests, the first four sample moments for each stock, showing large values of the kurtosis for all the stocks considered. We also report an estimate of the tail index $\gamma$ based on (21). The point estimate, $\hat{\gamma}_{Hill,j}$ is generally around 3, with some exceptions (Ashmore and Standard Chartered) for which it is higher. In view of the poor performance of tests based on $\hat{\gamma}_{Hill,j}$ (as appears from the Monte Carlo exercise), we report $\hat{\gamma}_{Hill,j}$ only as a preliminary indicator: its computation is conducted in the same way as in the Monte Carlo section. Based on the $\hat{\gamma}_{Hill,j}$s, it can be expected that all series have finite variances, whilst further testing is required for the third and fourth moments.

Tests are reported in the lower halves of Tables 2a and 2b. The tables show that, for all stocks considered, mean and variance exist, which is consistent with other studies in the literature (see e.g. Phillips and Loretan, 1991). As far as higher moments are concerned, all stocks in the banking sector have infinite third (and, therefore, fourth) moment. Results are less clear-cut for the financial sector. Two stocks appear to have finite third moment (Aberdeen and Ashmore), although the null that the fourth moment is infinite is accepted. Indeed, in both cases there is not an overwhelming amount of evidence in favour of the null (e.g., in the case of Aberdeen, the null is accepted at 5% level, but it would be rejected at 10%). Heuristically, this is in line with the descriptive statistics: both stocks have kurtosis around 9, which (were one to assume a Student t distribution for the data) would correspond to a degree of freedom of 5, thus admitting finite fourth moments. The other stocks in the financial sector have the same behaviour as observed for the banking sector, namely infinite third moments. Again, these results are reinforced by the estimated values of the kurtosis, which are similar to the ones found in the banking sector.

# 4 Discussion and conclusions

This paper proposes a test for the null that the $k$-th moment of a random variable does not exist. The test uses the SLLN, which stipulates that sample moments diverge or converge according as their population counterparts are infinite or finite. Since, under the null, sample moments diverge to infinity and therefore have no randomness, we propose a randomised testing procedure. From a methodological point of view, this approach to testing for the finiteness of moments avoids having to estimate the tail index, which is known to be fraught with difficulties. Our simulations show that the test has the correct size and good power. It is important to point out that the test can be applied to verify the existence or not of any moment, including fractional moments - i.e., the case where $k$ in $E\left|X\right|^k$ is not an integer number. This is bound to prove useful e.g. when evaluating estimators or predictors which are of infinite variance, and whose properties, therefore, cannot be studied in terms of the customarily employed $L_2$-norm or its derivatives (details, and a comprehensive literature review, can be found in Matsui and Pawlas, 2015).

A natural question that stems from this paper is whether it is possible to derive an estimate of $\gamma$. This contribution is focused on providing a test for the null that $E\left|X\right|^k$ does not exist - this can be of relevance e.g. when computing descriptive statistics; or when employing a theory that requires the existence of moments up to a certain order. It would be possible to use a sequential approach, based on testing for the existence of consecutive, (possibly) integer values of $k$, although in general this methodology would have to take into account the risk of a high procedure-wise rejection frequency (see in particular a very insightful paper by Fedotenkov, 2015b). On the other hand, the approach proposed in this paper complements the estimator of $\gamma$ suggested by Meerschaert and Scheffler (1998), who propose

$$\frac{1}{\hat{\gamma}} = \frac{\ln \sum_{i=1}^{n} x_i^2}{2 \ln n},$$

showing that it is consistent, albeit at the slow rate $O_p\left(\frac{1}{\ln n}\right)$.

Finally, a word of warning on the meaning of the hypothesis testing framework is in order. Indeed, testing whether a quantity is passing to infinity, when samples are naturally

finite, is bound to be conceptually unclear. In order to understand the rationale of the test, note that the null hypothesis is tested for by evaluating the rate of divergence of a sample moment (or, rather, of a scale invariant transformation thereof) - this lends itself to being put into an asymptotic setup. However, despite such asymptotic characterisation, the purpose of the analysis is to test for the *magnitude* of a sample moment, rather than for its actual behaviour at infinity. In this respect, the approach suggested in this paper is strongly related to the contribution of Bandi and Corradi (2014), who also propose a test for rates of divergence: as the authors put it, "evaluating magnitudes is essential to a variety of econometric problems". Thus, the purpose of this paper is to propose a procedure to allow the researcher to decide whether the moments of a random variable are "small enough" to be able to assume that the underlying distribution admits such moments, or not.

# Appendix: Proofs and Derivations

Recall that $P^*$ is the probability law of the $\{\xi_j\}_{j=1}^r$ conditional on the sample; henceforth, we let $E^*$ and $V^*$ denote the expected value and the variance calculated with respect to $P^*$.

We start with a preliminary Lemma, which contains a Law of the Iterated Logarithm (LIL).

**Lemma 1**. *Under Assumption 1, it holds that*

$$\limsup_{n\to\infty} \left( \frac{1}{n^{\frac{k}{\gamma}}} \sum_{i=1}^{n} |x_i|^k \right)^{1/\ln\ln n} = e^{\frac{k}{\gamma}} \ a.s.$$

**Proof**. The result in the lemma is a LIL for $\gamma$-stable processes, and it is also known as a Chover-type LIL (Chover, 1966). Several results on Chover-type LILs are available in the literature (see e.g. Cai, 2006; Wu and Jiang, 2010; Trapani, 2014); thus, when possible, passages in the proof are omitted to save space.

Let $y_i = |x_i|^k$; on account of Assumption 1*(i)* and of Theorem 14.1 in Davidson (2002, p. 210), the non negative sequence $y_i$ is also uniformly mixing with mixing numbers of the same size. Define also $y_i^{(n)} = y_i \, I\,[|y_i| < a_n]$, and $a_n = [nL\,(n)\,f\,(n)]^{k/\gamma}$ with $f\,(n)$ a function such that

$$\limsup_{t\to\infty} \sup_{0\le t'\le t} \frac{f\,(t')}{f\,(t)} < \infty.$$

Further, define $S_j = \sum_{i=1}^{j} y_i$ and $S_j^{(n)} = \sum_{i=1}^{j} \left[ y_i^{(n)} - E\left( y_i^{(n)} \right) \right]$.

We start by showing the upper half of the LIL, i.e.

$$\limsup_{n\to\infty} \left( \frac{\sum_{i=1}^{n} y_i}{n^{k/\gamma}} \right)^{1/\ln\ln n} \le e^{(1+\eta)k/\gamma} \ a.s., \tag{23}$$

for every $\eta > 0$. This requires showing that

$$\sum_{n=1}^{\infty} \frac{1}{n} P\left[ \max_{1\le j\le n} |S_j| > \varepsilon a_n \right] < \infty, \tag{24}$$

for some $\varepsilon > 0$. The passages are very similar e.g. to those in the proof of Theorem 2.3 in Cai (2006). Indeed

$$P\left[\max_{1\leq j\leq n}|S_j| > \varepsilon a_n\right] \leq P\left[\max_{1\leq j\leq n} y_j > a_n\right] + P\left[\max_{1\leq j\leq n}\left|S_j^{(n)}\right| > \varepsilon a_n - \max_{1\leq j\leq n}\left|\frac{1}{n}\sum_{i=1}^{j}E\left(y_i^{(n)}\right)\right|\right].$$
(25)

Under Assumption 1, as $n \to \infty$ we have

$$\frac{1}{a_n}\max_{1\leq j\leq n}\left|\frac{1}{n}\sum_{i=1}^{j}E\left(y_i^{(n)}\right)\right| = 0;$$
(26)

the proof is in Cai (2006). Combining (25) and (26)

$$\sum_{n=1}^{\infty}\frac{1}{n}P\left[\max_{1\leq j\leq n}|S_j| > \varepsilon a_n\right] \leq \sum_{n=1}^{\infty}\frac{1}{n}\sum_{j=1}^{n}P\left[y_j > a_n\right] + \sum_{n=1}^{\infty}\frac{1}{n}P\left[\max_{1\leq j\leq n}\left|S_j^{(n)}\right| > \varepsilon' a_n\right],$$

for some $0 < \varepsilon' < \varepsilon$ and $n$ large enough. Assumption 1$(ii)$ entails that $\sum_{j=1}^{n} n^{-1}P\left[y_j > a_n\right]$ $< \infty$. Finally, consider

$$\sum_{n=1}^{\infty}\frac{1}{n}P\left[\max_{1\leq j\leq n}\left|S_j^{(n)}\right| > \varepsilon' a_n\right] \leq M\sum_{n=1}^{\infty}\frac{1}{na_n^2}E\left[\max_{1\leq j\leq n}\left|S_j^{(n)}\right|^2\right]\frac{1}{a_n^2} \leq M'\sum_{n=1}^{\infty}\frac{1}{na_n^2}\sum_{i=1}^{n}E\left[\left(y_i^{(n)}\right)^2\right],$$

which follows from Markov's inequality and Rosenthal's inequality (see Lemma 1.3 and Theorem 2.1 in Xuejun, Shuhe, Yan and Wenzhi, 2009). Hence, exploiting the definition of $a_n$

$$\sum_{n=1}^{\infty}\frac{1}{n}P\left[\max_{1\leq j\leq n}\left|S_j^{(n)}\right| > \varepsilon' a_n\right]$$

$$\leq M''\sum_{n=1}^{\infty}\frac{1}{a_n^2}E\left[\left(y_i^{(n)}\right)^2\right] = M''\sum_{n=1}^{\infty}\frac{1}{a_n^2}\int_{|x|\leq a_n^{1/k}}|x|^{2k}\,dF(x)$$

$$= M''\sum_{n=1}^{\infty}\sum_{h=1}^{n}\frac{1}{a_n^2}\int_{a_{h-1}^{1/k}<|x|\leq a_h^{1/k}}|x|^{2k}\,dF(x) \leq M'''\sum_{h=1}^{\infty}a_h^2\left[\int_{a_{h-1}^{1/k}<|x|\leq a_h^{1/k}}dF(x)\right]\sum_{n=h}^{\infty}\frac{1}{a_n^2}$$

$$\leq M''''\sum_{h=1}^{\infty}h\left[\int_{a_{h-1}^{1/k}<|x|\leq a_h^{1/k}}dF(x)\right] \leq M^{v}\int_{1}^{\infty}\frac{dx}{xf(x)} < \infty.$$

23

This completes the proof of (24). Equation (23) follows from (24) upon choosing $f_n = \ln^{1+\eta} n$ - see e.g. the proof of Corollary 2.4 in Cai (2006).

We now turn to the lower half of the LIL, i.e.

$$\limsup_{n \to \infty} \left( \frac{\sum_{i=1}^n y_i}{n^{k/\gamma}} \right)^{1/\ln \ln n} \geq e^{(1-\eta)k/\gamma} \text{ a.s.,} \tag{27}$$

for any $\eta \in (0, 1)$. This requires showing that if, for some sequence $b_n$ of positive numbers, $\sum_{n=1}^{\infty} P[y_n > b_n] = \infty$, then

$$\limsup_{n \to \infty} \frac{|\sum_{i=1}^n y_i - d_n|}{b_n} = \infty \text{ a.s.,} \tag{28}$$

for any non decreasing sequence $d_n$ of positive numbers. To show this, recall that the sequence $\{y_i\}$ is uniformly mixing of the same size as $x_i$. Hence, the second Borel-Cantelli Lemma holds - see Lemma 1.1.2 in Iosifescu and Theodorescu (1969). Thus, a 0/1 law can be shown, yielding $P[y_n > b_n \text{ i.o.}] = 0$ or 1 according as $\sum_{n=1}^{\infty} P[y_n > b_n] < \infty$ or $= \infty$; see e.g. Lemma 4(ii) in Wu and Jiang (2010). Hence, equation (28) can be shown by contradiction. Assuming that (28) does not hold when $\sum_{n=1}^{\infty} P[y_n > b_n] = \infty$, this entails stating that there is a $d_0 \in [0, +\infty)$ such that $\limsup_{n \to \infty} b_n^{-1} |\sum_{i=1}^n y_i - d_n| = d_0$ almost surely. Clearly $\limsup_{n \to \infty} b_n^{-1} |y_n - (d_n - d_{n-1})| \leq 2d_0 + \limsup_{n \to \infty} b_n^{-1} |d_n - d_{n-1}|$. However, since $b_n^{-1} y_n \xrightarrow{p} 0$, we also have $\limsup_{n \to \infty} b_n^{-1} |d_n - d_{n-1}| = 0$, so that $\limsup_{n \to \infty} b_n^{-1} y_n \leq 2d_0$ a.s.; hence, $P[y_n > M b_n \text{ i.o.}] = 0$ for some $M \leq 2d_0 < \infty$. By the Borel-Cantelli Lemma, this entails that $\sum_{n=1}^{\infty} P[y_n > M b_n] < \infty$. But this contradicts the initial statement. Thus, (28) holds for any sequence $d_n$. Now, given a non decreasing sequence of positive numbers $f_n$, by Assumption 1(ii) we have $\sum_{n=1}^{\infty} P\left[y_n > (n f_n)^{k/\gamma}\right] = \infty$, as long as $\sum_{n=1}^{\infty} (n f_n)^{-1} = \infty$. In such case, it follows that $\limsup_{n \to \infty} (n f_n)^{-k/\gamma} \sum_{i=1}^n y_i = \infty$ a.s.; hence, equation (27) can be proved following exactly the same passages as in the proof of Theorem 2 in Wu and Jiang (2010), setting $f_n = \ln^{1-\eta} n$. Combining (23) and (27), the Lemma follows. QED

**Proof of Theorem 1**. We prove the theorem for a more general set-up, defining $Q(\hat{\mu}_k)$ be a continuous, positive, monotonically increasing transformation of $\hat{\mu}_k$, with

24

$\lim_{t \to +\infty} Q(t) = +\infty$. By Lemma 1, under $H_0$ we have $\hat{\mu}_k = O_{a.s.} \left( n^{k/\gamma - 1} \ln^{k/\gamma} n \right)$; thus, by continuity, $Q(\hat{\mu}_k) = O_{a.s.} \left[ Q \left( n^{k/\gamma - 1} \ln^{k/\gamma} n \right) \right]$. Henceforth, the proof of Theorem 1 is based on a similar logic to the proof of Theorem 5 in Bandi and Corradi (2014), and we only report the main passages. Let $\varsigma_{j,n} \equiv Q^{1/2}(\hat{\mu}_k) \xi_j$; conditional on the sample, $\varsigma_{j,n} \sim N[0, Q(\hat{\mu}_k)]$. Define now the set

$$\Omega_n \equiv \left\{ \omega : \frac{Q(\hat{\mu}_k)}{Q\left(n^{k/\gamma-1}\right)} > \varepsilon > 0 \right\},$$

for any $\varepsilon > 0$. Lemma 1 entails that, under $H_0$, $P\left[\lim_{n \to \infty} \Omega_n\right] = 1$. All the passages below are reported conditional on $\omega \in \Omega_n$. For each $u$ we have

$$
\begin{aligned}
\vartheta_{nr}(u) &= \frac{2}{\sqrt{r}} \sum_{j=1}^{r} \left\{ \zeta_{j,n}(u) - E^* \left[ \zeta_{j,n}(u) \right] \right\} + \frac{2}{\sqrt{r}} \sum_{j=1}^{r} \left\{ E^* \left[ \zeta_{j,n}(u) \right] - \frac{1}{2} \right\} \quad (29) \\
&= I + II,
\end{aligned}
$$

with

$$E^* \left[ \zeta_{j,n}(u) \right] = \frac{1}{2} + \frac{1}{\sqrt{2\pi T(\hat{\mu}_k)}} \int_0^u \exp\left[ -\frac{1}{2} \frac{t^2}{Q(\hat{\mu}_k)} \right] dt, \quad (30)$$

where we consider the case of $u > 0$ without loss of generality. Consider first $II$ in (29); based on (30), we have

$$
\begin{aligned}
II &\leq 2\sqrt{r} \frac{1}{\sqrt{2\pi T(\hat{\mu}_k)}} \int_0^u \exp\left[ -\frac{1}{2} \frac{t^2}{Q(\hat{\mu}_k)} \right] dt \quad (31) \\
&\leq \frac{2\sqrt{r}}{\sqrt{2\pi}} \int_0^{u/\sqrt{Q(\hat{\mu}_k)}} \exp\left[ -\frac{1}{2} s^2 \right] ds \leq M \sqrt{\frac{2}{\pi}} \sqrt{\frac{r}{Q(\hat{\mu}_k)}},
\end{aligned}
$$

where the third inequality is based on Taylor's expansion around $\frac{u}{\sqrt{Q(\hat{\mu}_k)}}$. By Lemma 1, it follows that $II$ in (29) is $O_p\left[ rT^{-1} \left( n^{k/\gamma - 1} \ln^{k/\gamma} n \right) \right]$; this is $o_p(1)$ uniformly in $u$ if (11) holds.

Turning to $I$ in (29), note that $\bar{\zeta}_{j,n}(u) \equiv \zeta_{j,n}(u) - E^* \left[ \zeta_{j,n}(u) \right]$ is an $i.i.d.$ sequence

with mean zero; it holds that

$$
\begin{aligned}
V^* \left[ \frac{2}{\sqrt{r}} \sum_{j=1}^{r} \bar{\zeta}_{j,n}(u) \right] &= 4 E^* \left[ \frac{1}{r} \sum_{j=1}^{r} \sum_{k=1}^{r} \bar{\zeta}_{j,n}(u) \bar{\zeta}_{k,n}(u) \right] = 4 \frac{1}{r} \sum_{j=1}^{r} E^* \left[ \bar{\zeta}_{j,n}^2(u) \right] \\
&= 1 + O_p \left[ \frac{1}{Q \left( n^{k/\gamma - 1} \right)} \right],
\end{aligned}
$$

where the second equality comes from the fact that $\bar{\zeta}_{j,n}(u)$ is generated independently across $j$, and the last equality comes from (30) and the passages thereafter. This holds uniformly in $u$ by the same passages as above. Thus, a CLT can be applied to $I$, so that, as $(n,r) \to \infty$, $I \xrightarrow{d^*} N(0,1)$. Putting everything together, as $(n,r) \to \infty$ with (11), $\vartheta_{nr}(u) \xrightarrow{d^*} N(0,1)$ uniformly in $u$. This entails that $\lim_{n,r \to \infty} \Theta_{nr} = \lim_{n,r \to \infty} \int_{\underline{u}}^{\overline{u}} \vartheta_{nr}^2(u) \varphi(u) \, du \xrightarrow{d^*} \int_{\underline{u}}^{\overline{u}} \chi_1^2 \varphi(u) \, du = \chi_1^2$. QED

**Proof of Theorem 2.** Under Assumption 1, with part *(ii)* modified so as to allow for $\gamma = k + \varepsilon$, it holds that $E\left[|Y|\ln|Y|\right] < \infty$. Since uniform mixing implies strong mixing, we can apply a SLLN for strong mixing sequences (e.g. Rio, 1995): thus, $\hat{\mu}_k \xrightarrow{a.s.} \mu_k < \infty$, so that $Q(\hat{\mu}_k) \xrightarrow{a.s.} Q(\mu_k) < \infty$. Similarly to the proof of Theorem 1, define

$$
\Omega_n^+ \equiv \{\omega : Q(\hat{\mu}_k) < M < \infty\},
$$

such that under $H_A$ we have $P\left[\lim_{n \to \infty} \Omega_n^+\right] = 1$. All the passages below are reported conditional on $\omega \in \Omega_n^+$.

Consider (29). Term $I$ still satisfies a CLT by construction, so that, under $H_A$, $I \xrightarrow{d^*} N(0,1)$. As far as $II$ in (29) is concerned, by (30) we have (considering the case of $u > 0$)

$$
II = 2\sqrt{r} \frac{1}{\sqrt{2\pi T(\hat{\mu}_k)}} \int_0^u \exp\left[-\frac{1}{2} \frac{t^2}{Q(\hat{\mu}_k)}\right] dt \leq M \sqrt{\frac{r}{Q(\hat{\mu}_k)}}. \tag{32}
$$

Note that the non-centrality parameter $II$ increases with the width of $U$. These passages entail that $\vartheta_{nr}(u)$ has a non-centrality parameter proportional to $\sqrt{\frac{r}{Q(\mu_k)}}$; therefore, $\Theta_{nr}$ has a noncentrality parameter that diverges under (12), giving the desired result. QED

**Proof of Corollary 1.** The proof is based on combining the proofs of Theorems 1 and 2. Define the $n$-dimensional vectors $\epsilon \equiv [\epsilon_1, ..., \epsilon_n]'$ and $\hat{\epsilon} \equiv [\hat{\epsilon}_1, ..., \hat{\epsilon}_n]'$. By construction, $\hat{\epsilon} = M\epsilon$, with $M \equiv I_n - W(W'W)^{-1}W'$, with $I_n$ the $n$-dimensional identity matrix and $W$ the design matrix.

In order to study the behaviour under the null, it suffices to show that when $E|\epsilon_1|^k = \infty$, Lemma 1 holds for $\hat{\mu}_k^{\epsilon}$. Let $\|b\|_k$ denote the $L_k$-norm of an $n$-dimensional vector $b$, i.e. $\|b\|_k = \left[\sum_{j=1}^n |b_j|^k\right]^{1/k}$; we can write $\hat{\mu}_k^{\epsilon} = n^{-1}\|\hat{\epsilon}\|_k^k = n^{-1}\|M\epsilon\|_k^k$. Based on Lemma 2.2 in Grcar (2003), there exists a finite and strictly positive constant, say $C_M$, such that $n^{-1}\|M\epsilon\|_k^k \geq C_M n^{-1}\|\epsilon\|_k^k$, so that $\hat{\mu}_k^{\epsilon} \geq C_M n^{-1}\|\epsilon\|_k^k$. By applying Lemma 1, $n^{-1}\|\epsilon\|_k^k = O_{a.s.}\left(n^{k/\gamma-1}\ln^{k/\gamma}n\right)$ under $H_0$. The same arguments as in the proof of Theorem 1 yield the asymptotics under the null of $\Theta_{nr}^{\epsilon}$.

As far as the behaviour of $\Theta_{nr}^{\epsilon}$ under alternatives is concerned, it suffices to show that, under $H_A$, $\hat{\mu}_k^{\epsilon}$ converges to a finite limit as $n \to \infty$. Let $\hat{\beta} - \beta = \left[\sum_{i=1}^n x_i^2\right]^{-1}\left[\sum_{i=1}^n x_i\epsilon_i\right]$; by applying the $C_r$-inequality we have

$$\frac{1}{n}\sum_{i=1}^n |\hat{\epsilon}_i|^k \leq 2^{k-1}\frac{1}{n}\sum_{i=1}^n |\epsilon_i|^k + 2^{k-1}\left|\hat{\beta} - \beta\right|^k \frac{1}{n}\sum_{i=1}^n |x_i|^k = 2^{k-1}(I + II).$$

As far as $I$ is concerned, using Assumption 1*(i) under the alternative, the SLLN entails that $n^{-1}\|\epsilon\|_k^k$ converges a.s. to a finite limit, so that $I = O_{a.s.}(1)$. Further

$$\begin{aligned}
II &= \left[\frac{1}{n}\sum_{i=1}^n x_i^2\right]^{-k}\left[\frac{1}{n}\sum_{i=1}^n x_i\epsilon_i\right]^k\left[\frac{1}{n}\sum_{i=1}^n |x_i|^k\right] \\
&\leq M\left[\frac{1}{n}\sum_{i=1}^n |x_i|^k\right]\left[\frac{1}{n}\sum_{i=1}^n x_i\epsilon_i\right]^k \leq M'\left[\frac{1}{n}\sum_{i=1}^n |x_i|^k\right]\left[\frac{1}{n}\sum_{i=1}^n |x_i\epsilon_i|\right]^k;
\end{aligned}$$

note that it is not important whether $n^{-1}\sum_{i=1}^n x_i^2$ converges or not in order to prove that $II$ is bounded: what matters is that it is bounded away from zero - part *(iii)* of Assumption 1* rules this out. By virtue of Assumption 1*(ii), the SLLN yields $n^{-1}\sum_{i=1}^n |x_i|^k = O_{a.s.}(1)$. Consider the remaining term; by virtue of the independence between $x_i$ and $\epsilon_i$, the sequence $\{x_i\epsilon_i\}_{i=1}^n$ is also $\phi$-mixing (see Theorem 5.2 in Bradley, 2005). Under the alternative that $\gamma_{\epsilon} > k$, and on account of Assumption 1*(ii), the SLLN can be employed

again, yielding $n^{-1} \sum_{i=1}^{n} |x_i \epsilon_i| = O_{a.s.}(1)$; this is not necessarily the sharpest bound, but it suffices for our purposes. Hence, $II = O_{a.s.}(1)$; this yields that, under the alternative, $\hat{\mu}_k^{\epsilon} = O_{a.s.}(1)$. The proof henceforth is the same as in Theorem 2. QED

**References**

Bandi, F., Corradi, V., 2014. Nonparametric nonstationarity tests. Econometric Theory, 30, 127–149.

Cai, G.-H., 2006. Chover-type laws of the iterated logarithm for weighted sums of $\rho^*$-mixing sequences. Journal of Applied Mathematics and Stochastic Analysis, 2006:65023.

Chover, J., 1966. A law of the iterated logarithm for stable summands. Proceedings of the American Mathematical Society, 17, 441–443.

Clauset, A., Shalizi, C.R., Newman, M.E.J., 2009. Power-law distributions in empirical data. SIAM Review, 51, 661–703.

Cline, D.B.H., 1982. Consistency for least squares regression estimators with infinite variance data. Journal of Statistical Planning and Inference, 23, 163–179.

Cornea-Madeira A., Davidson, R., 2014. A parametric bootstrap for heavy-tailed distributions. Econometric Theory, 31, 1–22.

Corradi, V., Swanson, N.R., 2006. The effects of data transformation on common cycle, cointegration, and unit root tests: Monte Carlo and a simple test. Journal of Econometrics, 132, 195–229.

Davidson, J., 2002. Stochastic limit theory. Oxford: Oxford University Press.

Drees, H., Kaufmann, E., 1998. Selecting the optimal sample fraction in univariate extreme value estimation. Stochastic Processes and Applications, 75, 149–172.

Embrechts, P, Kluppelberg, C., Mikosch, T., 1997. Modelling extremal events for insurance and finance. New York: Springer.

Fedotenkov I., 2013. A bootstrap method to test for the existence of finite moments. Journal of Nonparametric Statistic, 25, 315–322.

Fedotenkov I., 2015a, A simple nonparametric test for the existence of finite moments, Mimeo.

Fedotenkov I., 2015b. Bootstrap for testing the existence of finite moments: a caution for possible misapplication, Mimeo.

Grcar, J.F., 2003. A matrix lower bound. Lawrence Berkeley National Laboratory report LBNL-50635.

de Haan, L., Ferreira, A., 2006. Extreme value theory: an introduction. New York: Springer.

Hill, B.M., 1975. A simple general approach to inference about the tail of a distribution. The Annals of Statistics, 3, 1163–1174.

Hill, J.B., 2010. On tail index estimation for dependent, heterogeneous data. Econometric Theory, 26, 1398–1436.

Hurst, S. R., Platen, E. 1997. The marginal distributions of returns and volatility. In Y. Dodge (Ed.), $L_1$-Statistical Procedures and Related Topics, Volume 31 of IMS Lecture Notes - Monograph Series, pp. 301-314. Institute of Mathematical Statistics Hayward, California.

Iosifescu, M., Theodorescu, R., 1969. Random Processes and Learning. Berlin: Springer.

Linton, O.B., Xiao, Z., 2013. Estimation of and inference about the expected shortfall for time series with infinite variance. Econometric Theory, 29, 771–807.

Markowitz, H., Usmen, N. 1996a. The likelihood of various stock market return distributions, Part 1: Principles of inference. Journal of Risk & Uncertainty, 13, 207-219.

Markowitz, H., Usmen, N. 1996b. The likelihood of various stock market return distributions, Part 2: Empirical results. Journal of Risk & Uncertainty, 13, 221-247.

Matsui, M., Pawlas, Z., 2015. Fractional absolute moments of heavy tailed distributions. Forthcoming in Brazilian Journal of Probability and Statistics.

McElroy T., Politis D.N., 2007. Moment-based tail index estimation. Journal of Statistical Planning and Inference, 137, 1389–1406.

Meerschaert M.M., Scheffler H-P., 1998. A simple robust estimation method for the thickness of heavy tails. Journal of Statistical Planning and Inference, 71, 19–34.

Pearson, E.S., 1950. On questions raised by the combination of tests based on discon-

tinuous distributions. Biometrika, 37, 383–398.

Phillips, P.C.B., Loretan, M., 1991. The Durbin–Watson ratio under infinite variance errors. Journal of Econometrics, 47, 85–114.

Phillips, P.C.B., Loretan, M., 1994. Testing the covariance stationarity of heavy-tailed time series. Journal of Empirical Finance, 1, 211–248.

Phillips, P.C.B., Loretan, M., 1995. On the theory of testing covariance stationarity under moment condition failure. In G. S. Maddala, P.C.B. Phillips and T.N. Srinivasan (eds.), Advances in Econometrics and Quantitative Economics: Essays in Honor of Professor C.R. Rao. Basil Blackwell.

Resnick, S.I., 1997. Heavy tail modeling and teletraffic data. Annals of Statistics, 25, 1805–1869.

Rio, E., 1995. A maximal inequality and dependent Marcinkiewicz-Zygmund strong laws. Annals of Probability, 23, 918–937.

Trapani, L., 2014. Chover-type Laws of the $k$-Iterated Logarithm for strongly mixing sequences. Journal of Mathematical Analysis and Applications, 420, 908–916.

Xuejun, W., Shuhe, H., Yan, S., Wenzhi, Y., 2009. Moment inequality for $\varphi$-mixing sequences and its applications. Journal of Inequalities and Applications, 2009:379743.

Wu, Q., Jiang, Y., 2010. Chover-type laws of the $k$-iterated logarithm for $\tilde{\rho}$-mixing sequences of random variables. Journal of Mathematical Analysis and Applications, 366, 435–443.

| n | $(\rho,\theta)$ Test specifics $v$ | (0, 0) 2 | 3 | 4 | 5 | 6 | (0.5, 0) 2 | 3 | 4 | 5 | 6 | (0, 0.5) 2 | 3 | 4 | 5 | 6 | (0. − 0.5) 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 100 | $PRE-W$ | 0.060 | 0.062 | 0.104 | 0.320 | 0.446 | 0.052 | 0.059 | 0.099 | 0.331 | 0.445 | 0.052 | 0.061 | 0.095 | 0.336 | 0.443 | 0.049 | 0.060 | 0.105 | 0.316 | 0.451 |
| | $Hill$ | 0.057 | 0.103 | 0.120 | 0.172 | 0.185 | 0.063 | 0.100 | 0.143 | 0.174 | 0.198 | 0.043 | 0.099 | 0.124 | 0.168 | 0.183 | 0.050 | 0.100 | 0.118 | 0.163 | 0.182 |
| 250 | $PRE-W$ | 0.055 | 0.056 | 0.098 | 0.422 | 0.513 | 0.069 | 0.059 | 0.101 | 0.434 | 0.512 | 0.069 | 0.059 | 0.094 | 0.420 | 0.502 | 0.059 | 0.057 | 0.098 | 0.419 | 0.503 |
| | $Hill$ | 0.040 | 0.108 | 0.169 | 0.216 | 0.256 | 0.042 | 0.110 | 0.163 | 0.218 | 0.266 | 0.042 | 0.110 | 0.148 | 0.218 | 0.245 | 0.046 | 0.096 | 0.164 | 0.231 | 0.256 |
| 500 | $PRE-W$ | 0.055 | 0.056 | 0.063 | 0.549 | 0.668 | 0.059 | 0.058 | 0.066 | 0.528 | 0.688 | 0.049 | 0.051 | 0.064 | 0.513 | 0.678 | 0.056 | 0.057 | 0.061 | 0.545 | 0.698 |
| | $Hill$ | 0.047 | 0.115 | 0.183 | 0.229 | 0.276 | 0.050 | 0.121 | 0.195 | 0.240 | 0.293 | 0.045 | 0.119 | 0.170 | 0.247 | 0.290 | 0.045 | 0.118 | 0.178 | 0.258 | 0.295 |
| 1000 | $PRE-W$ | 0.041 | 0.056 | 0.061 | 0.640 | 0.788 | 0.053 | 0.057 | 0.057 | 0.637 | 0.766 | 0.041 | 0.056 | 0.058 | 0.663 | 0.780 | 0.047 | 0.063 | 0.061 | 0.649 | 0.780 |
| | $Hill$ | 0.043 | 0.105 | 0.234 | 0.312 | 0.383 | 0.038 | 0.129 | 0.219 | 0.310 | 0.367 | 0.035 | 0.120 | 0.212 | 0.305 | 0.366 | 0.034 | 0.122 | 0.209 | 0.304 | 0.387 |
| 10000 | $PRE-W$ | 0.050 | 0.055 | 0.060 | 0.796 | 0.936 | 0.044 | 0.051 | 0.062 | 0.783 | 0.933 | 0.046 | 0.043 | 0.059 | 0.708 | 0.923 | 0.046 | 0.048 | 0.057 | 0.704 | 0.934 |
| | $Hill$ | 0.018 | 0.126 | 0.281 | 0.443 | 0.624 | 0.020 | 0.123 | 0.274 | 0.463 | 0.603 | 0.023 | 0.120 | 0.298 | 0.474 | 0.599 | 0.025 | 0.104 | 0.297 | 0.443 | 0.628 |
| 100000 | $PRE-W$ | 0.050 | 0.051 | 0.060 | 0.806 | 0.996 | 0.050 | 0.051 | 0.059 | 0.806 | 1.000 | 0.050 | 0.051 | 0.059 | 0.806 | 1.000 | 0.050 | 0.051 | 0.059 | 0.806 | 1.000 |
| | $Hill$ | 0.012 | 0.096 | 0.238 | 0.596 | 0.806 | 0.011 | 0.096 | 0.338 | 0.596 | 0.821 | 0.011 | 0.095 | 0.339 | 0.597 | 0.804 | 0.011 | 0.095 | 0.338 | 0.596 | 0.804 |

Table 1a. (BENCHMARK CASE) Empirical rejection frequencies at a nominal 5% value; data are generated as Student t with $v$ degrees of freedom. The main specifications of the test are: $r = n^{4/5}$ and $U = \{-1, 1\}$, with both values chosen with equal probability. Each block corresponds to a combination of $n$ and $(\rho, \theta)$ contains 5 different tests: two based on using raw data (entries labelled "RAW"), two based on using pre-whitened data by fitting an $AR(7)$ model (entries labelled "PRE-W") and the last one is based on the Hill estimator as discussed in this section. For each block, each column corresponds to a different value of the tail index $\gamma$; in the first two columns ($\gamma = 2$ and $\gamma = 3$), the empirical rejection frequency represents the size of the test, whilst in the last columns ($\gamma = 5$ and $\gamma = 6$) it represents the power of the test.

| $n$ | $(\rho, \theta)$ $\upsilon$ | | | (0, 0) | | |
|---|---|---|---|---|---|---|
| | | 2 | 3 | 4 | 5 | 6 |
| | *Test specifics* | | | | | |
| 100 | $r = n^{1/2}, U \in \{-1, 1\}$ | 0.031 | 0.026 | 0.027 | 0.077 | 0.104 |
| | $r = n^{4/5}, U \in \{-2, 2\}$ | 0.088 | 0.124 | 0.254 | 0.444 | 0.493 |
| | $r = n^{1/2}, U \in \{-2, 2\}$ | 0.031 | 0.048 | 0.069 | 0.307 | 0.412 |
| 250 | $r = n^{1/2}, U \in \{-1, 1\}$ | 0.039 | 0.039 | 0.044 | 0.232 | 0.351 |
| | $r = n^{4/5}, U \in \{-2, 2\}$ | 0.059 | 0.094 | 0.123 | 0.587 | 0.716 |
| | $r = n^{1/2}, U \in \{-2, 2\}$ | 0.039 | 0.046 | 0.082 | 0.426 | 0.464 |
| 500 | $r = n^{1/2}, U \in \{-1, 1\}$ | 0.060 | 0.049 | 0.062 | 0.367 | 0.432 |
| | $r = n^{4/5}, U \in \{-2, 2\}$ | 0.062 | 0.082 | 0.097 | 0.696 | 0.841 |
| | $r = n^{1/2}, U \in \{-2, 2\}$ | 0.061 | 0.054 | 0.086 | 0.473 | 0.597 |
| 1000 | $r = n^{1/2}, U \in \{-1, 1\}$ | 0.059 | 0.061 | 0.070 | 0.395 | 0.526 |
| | $r = n^{4/5}, U \in \{-2, 2\}$ | 0.041 | 0.076 | 0.091 | 0.783 | 0.895 |
| | $r = n^{1/2}, U \in \{-2, 2\}$ | 0.059 | 0.061 | 0.072 | 0.565 | 0.668 |
| 10000 | $r = n^{1/2}, U \in \{-1, 1\}$ | 0.032 | 0.051 | 0.055 | 0.512 | 0.651 |
| | $r = n^{4/5}, U \in \{-2, 2\}$ | 0.042 | 0.057 | 0.079 | 0.898 | 1.000 |
| | $r = n^{1/2}, U \in \{-2, 2\}$ | 0.032 | 0.051 | 0.063 | 0.584 | 0.680 |
| 100000 | $r = n^{1/2}, U \in \{-1, 1\}$ | 0.055 | 0.048 | 0.054 | 0.584 | 0.874 |
| | $r = n^{4/5}, U \in \{-2, 2\}$ | 0.055 | 0.055 | 0.061 | 1.000 | 1.000 |
| | $r = n^{1/2}, U \in \{-2, 2\}$ | 0.055 | 0.048 | 0.045 | 0.687 | 0.995 |

Table 1b. Empirical rejection frequencies at a nominal 5% value; data are generated as Student t with $\upsilon$ degrees of freedom for the case $(\rho, \theta) = (0, 0)$. For each sample size, the entries refer to: (a) the case where $r = n^{\frac{1}{2}}$ and $U = \{-1, 1\}$; (b) the case $r = n^{\frac{4}{5}}$ and $U = \{-2, 2\}$; (c) the case $r = n^{\frac{1}{2}}$ and $U = \{-2, 2\}$.

| n | Test specifics | (0, 0) 2 | 3 | 4 | 5 | 6 | (0.5, 0) 2 | 3 | 4 | 5 | 6 | (0, 0.5) 2 | 3 | 4 | 5 | 6 | (0, −0.5) 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 100 | PRE − W | 0.062 | 0.065 | 0.154 | 0.394 | 0.416 | 0.058 | 0.063 | 0.146 | 0.413 | 0.423 | 0.061 | 0.064 | 0.162 | 0.382 | 0.403 | 0.060 | 0.065 | 0.148 | 0.412 | 0.415 |
| | Hill | 0.031 | 0.051 | 0.064 | 0.072 | 0.067 | 0.032 | 0.043 | 0.066 | 0.055 | 0.073 | 0.031 | 0.054 | 0.058 | 0.066 | 0.064 | 0.032 | 0.055 | 0.056 | 0.065 | 0.073 |
| 250 | PRE − W | 0.063 | 0.061 | 0.126 | 0.497 | 0.561 | 0.061 | 0.060 | 0.132 | 0.516 | 0.577 | 0.060 | 0.061 | 0.119 | 0.476 | 0.548 | 0.057 | 0.059 | 0.121 | 0.483 | 0.572 |
| | Hill | 0.028 | 0.039 | 0.059 | 0.083 | 0.081 | 0.022 | 0.048 | 0.057 | 0.071 | 0.095 | 0.025 | 0.042 | 0.063 | 0.085 | 0.089 | 0.030 | 0.044 | 0.073 | 0.074 | 0.085 |
| 500 | PRE − W | 0.050 | 0.049 | 0.098 | 0.527 | 0.613 | 0.057 | 0.062 | 0.105 | 0.541 | 0.625 | 0.051 | 0.049 | 0.100 | 0.518 | 0.611 | 0.051 | 0.050 | 0.104 | 0.536 | 0.607 |
| | Hill | 0.032 | 0.060 | 0.083 | 0.116 | 0.124 | 0.033 | 0.056 | 0.082 | 0.122 | 0.115 | 0.031 | 0.056 | 0.082 | 0.100 | 0.127 | 0.032 | 0.060 | 0.073 | 0.014 | 0.124 |
| 1000 | PRE − W | 0.044 | 0.058 | 0.068 | 0.626 | 0.783 | 0.039 | 0.059 | 0.071 | 0.634 | 0.794 | 0.040 | 0.051 | 0.069 | 0.636 | 0.792 | 0.043 | 0.053 | 0.070 | 0.641 | 0.782 |
| | Hill | 0.028 | 0.060 | 0.090 | 0.142 | 0.154 | 0.028 | 0.059 | 0.105 | 0.136 | 0.150 | 0.035 | 0.063 | 0.097 | 0.121 | 0.160 | 0.029 | 0.058 | 0.100 | 0.129 | 0.153 |
| 10000 | PRE − W | 0.043 | 0.041 | 0.054 | 0.687 | 0.836 | 0.043 | 0.041 | 0.056 | 0.708 | 0.852 | 0.043 | 0.040 | 0.044 | 0.704 | 0.848 | 0.042 | 0.041 | 0.050 | 0.692 | 0.831 |
| | Hill | 0.014 | 0.071 | 0.159 | 0.244 | 0.303 | 0.015 | 0.062 | 0.147 | 0.229 | 0.324 | 0.019 | 0.065 | 0.143 | 0.235 | 0.322 | 0.012 | 0.073 | 0.163 | 0.254 | 0.331 |
| 100000 | PRE − W | 0.043 | 0.040 | 0.052 | 0.760 | 0.903 | 0.042 | 0.041 | 0.058 | 0.798 | 0.911 | 0.040 | 0.041 | 0.041 | 0.773 | 0.912 | 0.043 | 0.040 | 0.040 | 0.766 | 0.895 |
| | Hill | 0.004 | 0.074 | 0.213 | 0.377 | 0.534 | 0.006 | 0.066 | 0.206 | 0.375 | 0.560 | 0.006 | 0.072 | 0.248 | 0.394 | 0.550 | 0.007 | 0.065 | 0.229 | 0.388 | 0.570 |

Table 1c. Empirical rejection frequencies at a nominal 5% value; data are generated from a power law with tail index $\gamma$.

| Company | Barclays | HSBC | Lloyds | RBS | Std. Chart. |
|---|---|---|---|---|---|
| **Descriptive Statistics** | | | | | |
| Mean | $-0.0004$ | $-0.0001$ | $-0.0008$ | $-0.0016$ | $0.0000$ |
| Std. Dev. | $0.0418$ | $0.0216$ | $0.0438$ | $0.0516$ | $0.0297$ |
| Skewness | $1.2506$ | $-0.3093$ | $-0.8238$ | $-6.7949$ | $0.3129$ |
| Kurtosis | $28.675$ | $14.691$ | $24.156$ | $145.216$ | $13.395$ |
| $\hat{\gamma}_{Hill}$ | $3.008$ | $3.486$ | $3.546$ | $2.256$ | $6.226$ |
| **Tests** | | | | | |
| $\mu_1$ | $2.61 \times 10^{-2}$ [0.000] | $1.40 \times 10^{-2}$ [0.000] | $2.70 \times 10^{-2}$ [0.000] | $2.87 \times 10^{-2}$ [0.000] | $1.90 \times 10^{-2}$ [0.000] |
| $\mu_2$ | $17.46 \times 10^{-4}$ [0.000] | $4.68 \times 10^{-4}$ [0.000] | $19.22 \times 10^{-4}$ [0.000] | $26.66 \times 10^{-4}$ [0.001] | $8.81 \times 10^{-4}$ [0.000] |
| $\mu_3$ | $277.27 \times 10^{-6}$ [0.316] | $30.24 \times 10^{-6}$ [0.107] | $318.74 \times 10^{-6}$ [0.225] | $1152.87 \times 10^{-6}$ [0.554] | $77.78 \times 10^{-6}$ [0.139] |
| $\mu_4$ | $8736.86 \times 10^{-8}$ [0.485] | $321.95 \times 10^{-8}$ [0.233] | $8945.16 \times 10^{-8}$ [0.555] | $103630.23 \times 10^{-8}$ [0.420] | $1040.76 \times 10^{-8}$ [0.597] |

Table 2a. Results for companies in the banking sector - the stocks considered are, respectively, Barclays plc, HSBC Holdings plc, Lloyds Banking Group plc, The Royal Bank of Scotland Group plc, Standard Chartered plc. For each stock, we report the mean, the standard deviation (computed after adjusting for the degree of freedom), the skewness and the kurtosis (both unadjusted). Hill estimator of the tail index has been computed based on (21), using the same specifications as in the Monte Carlo section. In the "Tests" section, we report the values taken by the $k$-th absolute sample moment; number in square brackets are p-values.

| | Company | Aberdeen | Ashmore | ICAP | Schroders |
|---|---|---|---|---|---|
| **Descriptive Statistics** | | | | | |
| Mean | | 0.0001 | 0.0002 | $-0.0004$ | 0.0005 |
| Std. Dev. | | 0.0238 | 0.0264 | 0.0304 | 0.0268 |
| Skewness | | 0.3224 | $-0.1938$ | $-0.0941$ | $-0.0649$ |
| Kurtosis | | 8.946 | 9.179 | 13.467 | 22.357 |
| $\hat{\gamma}_{Hill}$ | | 3.393 | 5.434 | 3.043 | 2.472 |
| **Tests** | | | | | |
| $\mu_1$ | | $1.63 \times 10^{-2}$ | $1.75 \times 10^{-2}$ | $1.99 \times 10^{-2}$ | $1.75 \times 10^{-2}$ |
| | | [0.000] | [0.000] | [0.000] | [0.000] |
| $\mu_2$ | | $5.65 \times 10^{-4}$ | $6.98 \times 10^{-4}$ | $9.22 \times 10^{-4}$ | $7.15 \times 10^{-4}$ |
| | | [0.000] | [0.000] | [0.000] | [0.000] |
| $\mu_3$ | | $32.93 \times 10^{-6}$ | $46.81 \times 10^{-6}$ | $81.21 \times 10^{-6}$ | $64.52 \times 10^{-6}$ |
| | | [0.034] | [0.041] | [0.282] | [0.284] |
| $\mu_4$ | | $286.11 \times 10^{-8}$ | $446.85 \times 10^{-8}$ | $1146.78 \times 10^{-8}$ | $1143.70 \times 10^{-8}$ |
| | | [0.098] | [0.111] | [0.424] | [0.707] |

Table 2b. Results for companies in the financial services sector - the stocks considered are, respectively, Aberdeen Asset Management plc, Ashmore Group plc, ICAP plc and Schroders plc. For each stock, we report the mean, the standard deviation (computed after adjusting for the degree of freedom), the skewness and the kurtosis (both unadjusted). Hill estimator of the tail index has been computed based on (21), using the same specifications as in the Monte Carlo section. In the "Tests" section, we report the $k$-th absolute sample moment; number in square brackets are p-values.