Tail structural change in small samples

Tail Structural Change in Small Samples

Stefan Straetmans *   Bertrand Candelon †‡

September 1, 2008

Abstract

The tail of financial returns typically declines with a power law ("fat tails"). However, less is known on the temporal stability of the parameter governing the tail decline, i.e. the so-called tail index α. Using the Hill estimator for the tail index, we study the small sample properties of some recently proposed endogenous tests for structural change in α. We find that small sample critical values can widely differ across different Data Generating Processes due to the bias in the Hill estimator and small sample estimation risk. Given the fact that the critical values seem data specific, we recommend a bootstrap-based version of the structural change test. Applying this stability test to a large cross-section of stock indices, we are hardly able to detect any temporal shift in the tail index. This is reassuring news for the proponents of Extreme value techniques (EVT) who assume stationary tail behavior over long time spans when applying tail index and extreme quantile estimation.

J.E.L. Codes: F31, G15, G19, C49
Keywords: Heavy tails; Tail Index; Extreme Value Analysis; Endogenous stability test; Finite sample properties; Bootstrap.

---

*Corresponding author: Limburg Institute of Financial Economics (LIFE), Economics Faculty, Maastricht University, PoBOX 616, MD 6200, Maastricht, The Netherlands ; e-mail: s.straetmans@finance.unimaas.nl.

†Maastricht University, Department of Economics.

‡We benefited from suggestions and comments by participants at the Netherlands Econometrics Study Group (NESG) in Tilburg and seminar participants at IESE (Barcelona) and the Universite d'Orleans. The usual disclaimers apply.

1 Introduction

The 1997 Asian crisis, the LTCM debacle or the recent subprime credit crunch and the resulting financial market turmoil have increased the awareness of both academics and practitioners on the importance of accurately assessing the likelihoods of such extreme events. However, the academic interest into large tail events is far from new, see e.g. Mandelbrot (1963) as a seminal reference. He was one of the first to acknowledge that overnight financial market turbulence cannot be described by the normal distribution paradigm. More specifically, tail probabilities seem to exhibit a polynomial tail decay ("heavy" tails) in contrast to the exponential tail decays of so-called "thin-tailed" models like the normal df. This "heavy tail" characteristic has been detected for most financial asset classes. Numerous empirical studies subsequently focused on identifying the degree of probability mass in the tail by estimating the so called tail index α.[1] Loosely speaking this parameter reflects the number of bounded distributional moments that are still finite.

Much less attention has been paid to the possibility and consequences of a nonconstant tail index α.[2] Conditional volatility models such as the ARCH- and GARCH-type class (see Engle (1982) and Bollerslev (1986)) reconcile a stationary unconditional distribution function (constant α) with clusters of high and low volatility in the conditional df. However,

[1] Jansen and de Vries (1991), Longin (1996), Lux (1996) and Hartmann, Straetmans and de Vries (2004) investigated the probability mass in the tails of stock market returns; whereas fat tails in foreign exchange returns have been considered, inter alia, by Koedijk, Schafgans and de Vries (1990), Koedijk, Stork and de Vries (1992), Hols and de Vries (1991) and Hartmann, Straetmans and de Vries (2003). Bond extremes have been rather neglected in the empirical literature; De Haan, Jansen, Koedijk and de Vries (1994) and Hartmann et al. (2004) constitute two notable exceptions.

[2] Exceptions constitute Phillips and Loretan (1990); Koedijk et al. (1990, 1992); Jansen and de Vries (1991) or Pagan and Schwert (1990). These studies all perform "exogenous" tests of breakpoint detection by imposing the candidate-breakpoint.

the question arises whether it is realistic to assume that the tail of the unconditional df (and thus measures of long-run risk like unconditional quantiles) remains invariant over very long time periods. In other words: can highly volatile periods like the Asian crisis or the Mexican Tequila crisis and periods of market quiescence both be explained by one and the same stationary distributional model or do crisis periods induce structural shifts in α?

Testing for structural change in the tail behavior of the unconditional distribution is relevant from both a statistical and policy perspective. First, whether Extreme Value Theory (EVT) or e.g. the cited ARCH and GARCH models are applicable depends on the stationarity assumption for the unconditional tail.[3] Also, a nonconstant α implies a violation of covariance stationarity which invalidates standard statistical inference based on regression analysis. From a policy perspective, quantifying the correct level of the tail index is of potential importance to risk managers and financial regulators because it is a necessary ingredient for calculating the unconditional Value-at-Risk (VaR) very far into the distributional tail for e.g. the sake of stress testing. For example, if a decrease in α (and thus a rise in the magnitude and frequency of extreme events) is not properly recognized by risk managers and regulators, unconditional VaR quantiles will most probably be underestimated resulting in insufficient capital buffers. This can ultimately jeopardize the stability of the financial system.

The scant empirical literature on the constancy issue mainly focuses on testing for a single known (i.e. exogenously selected) breakpoint in α and provides only weak evidence for structural change in tail behavior.[4] Quintos, Fan and Phillips (2001) (QFP from here

[3]Starica (2003) focuses on this potential weakness of GARCH-type models, especially when fitted over long time periods.

[4]Tests for structural change in α have to be distinguished from cross sectional equality tests (see e.g. Koedijk et al.(1990) on exchange rates or Jondeau and Rockinger (2003) on stock markets) or asymmetry tests between left and right tails of the same series (see e.g. Jansen and de Vries (1991) or Lux (1996) on

on) proposed a battery of stability tests based on the Hill estimator for α. They also derived limiting asymptotic distributions for these tests and briefly studied their small sample behavior.

The main contribution of this paper is twofold: we extend the QFP Monte Carlo study in several directions and propose a bootstrap-based version of the stability test for empirical applications. Whereas the QFP small sample analysis uses fixed (percentual) fractions of tail observations for different sample sizes, we select the Hill estimator's nuisance parameter by minimizing its Asymptotic Mean Squared Error (AMSE). The latter approach constitutes common practice in EVT whereas it can be argued that taking a fixed percentage of extremes leads to degenerate asymptotic limiting dfs for the Hill estimator and accompanying stability tests. Also, whereas the QFP simulations were solely based on the class of symmetric stable distributions we also consider Data Generating Processes (DGPs) that are more representative for financial data, e.g. models that allow for the existence of the unconditional variance ($\alpha > 2$) or that exhibit real-life data features such as temporal linear or nonlinear dependence (serial correlation and volatility clustering in returns).

Anticipating on our results, we find that size, (size-corrected) power and the ability to detect breaks in small samples varies a lot depending on the considered DGP. Small sample critical values seem to inherit the Hill bias. This bias in critical values differs across DGPs and persists in larger samples. This suggests that there is no globally valid set of critical values for temporal stability tests of the tail index. Instead, we propose to bootstrap the critical values in empirical applications for each data set separately. The size-corrected power properties and the ability to detect breaks in the Monte Carlo study suggest that a "recursive" version of the stability test is to be preferred provided the sample is sufficiently large ($n \geq 2,000$). Upon applying a bootstrap-based version of this test to a large cross section of developed

stock index tail asymmetries).

and emerging stock markets, we hardly detect any breaks in tail behavior.

The rest of the paper is organized as follows. Section 2 provides a refresher on the statistical theory of heavy tails and accompanying endogenous stability tests. Section 3 contains an elaborate Monte-Carlo investigation of the endogenous breakpoint tests' size, power and break date ability. Section 4 provides an empirical application. Conclusions are comprised in section 5.

2 Testing structural change in the tail behavior: theory

A short digression on the theory and estimation of the index of regular variation is provided in section 2.1 followed by a short discussion of the mechanics of some temporal stability tests for the tail index α (section 2.2).

2.1 Regular variation

We start from the empirical stylized fact that sharp fluctuations in financial market prices exhibit fat tails, see e.g. Mandelbrot (1963) for an early reference or the more recent monograph by Embrechts, Klüppelberg and Mikosch (1997). Without loss of generality, the presented estimation and testing procedures are expressed in terms of the right tail, i.e., the survivor function $P\{X \geq x\} := 1 - F(x)$. Our empirical application (Section 4) will focus on stock market returns' left tails. This requires taking the negative of a return series prior to applying the sketched framework. Under fairly general conditions the survivor function of heavy tailed (or "regularly varying") distributions can be approximated by the second order order Taylor expansion for large x :

$$1 - F(x) = ax^{-\alpha}(1 + bx^{-\beta} + o(x^{-\beta})), \tag{2.1}$$

with $a > 0$, $b \in \Re$, $\beta > 0$, see e.g. de Haan and Stadmuller (1996). The parameters β and b that govern the second order behavior in (2.1) reflect the deviation from pure Pareto behavior in the tail. As will be argued later on, those parameters also strongly influence the small sample properties of the Hill statistic and stability tests. The case $\beta = 0$ corresponds to the expansion $P\{X \geq x\} \simeq ax^{-\alpha}[1 + b\ln x]$. The tail specializes to an exact Pareto when $b = 0$.

The regular variation property implies that the (appropriately scaled) upper tail returns lie in the (maximum) domain of attraction of the Type-II extreme value ("Frechet") distribution. The tail index α reflects the speed at which the tail probability in (2.1) decays if x is increased. Clearly, the lower α the slower the probability decay and the higher the probability mass in the tail of X, *ceteris paribus* the level of x. The regular variation property, *inter alia*, implies that all distributional moments higher than α, i.e. $E[X^r]$, $r > \alpha$, are unbounded, signifying "fat tails". Fat tailed models of the unconditional distribution include the Student-t, symmetric stable, Burr, and Frechet dfs as well as the GARCH class of conditional volatility models.[5] As for the tail of the standard normal distribution, a popular tail approximation expresses the survivor function $1 - \Phi(x)$ in terms of the density $\phi(x)$:

$$\begin{aligned} 1 - \Phi(x) &\simeq \frac{\phi(x)}{x}, \quad x \text{ large} \\ &= (2\pi x)^{-1} \exp\left(-\frac{1}{2}x^2\right), \end{aligned}$$

which clearly describes an exponentially declining tail, see Feller (1971a, VII.1). Distributions characterized by this tail decay are dubbed "thin tailed" because the tail probabilities

[5] Hall (1982) imposes the more stringent condition $\alpha = \beta$ on the tail expansion. This covers certain distributions like the stable laws and the type II extreme value distribution (Frechet); but it does not apply to e.g. the Student-t or the Burr df. For the Student-t df the tail expansion (2.1) holds, though, with α equal to the degrees of freedom ($\alpha = v$) and $\beta = 2$. As for the Burr df, the second order parameter is not restricted. The value of β is unknown for the GARCH class.

decline much faster to zero as under (2.1); but these distributions possess all moments, and hence do not capture what is typically observed in financial data.

The focus of the paper will be on the small sample properties of temporal stability tests for α-estimators. The investigated test statistics use Hill's (1975) estimator for α as an input. Let $X_{1,n} \leq X_{2,n} \leq \cdots \leq X_{n,n}$ represent the ascending order statistics that correspond with the returns series X for a sample of size n. Then Hill's estimator boils down to:

$$\widehat{\alpha} = \left(\frac{1}{m} \sum_{j=0}^{m-1} \ln \left(\frac{X_{n-j,n}}{X_{n-m,n}} \right) \right)^{-1}, \tag{2.2}$$

with m the number of highest order statistics used in the estimation. The convergence in distribution of the Hill statistic critically depends on the rate at which the nuisance parameter m grows with the total sample size n. The main convergence in distribution result for $\widehat{\alpha}$ is summarized in the following theorem:

Theorem 1 *(Asymptotic normality) Assume that* $1 - F(x)$ *obeys (2.1). If* $m,n \longrightarrow \infty$ *we distinguish two cases:*

(A) If $m = o\left(n^{2\beta/2\beta+\alpha}\right)$ *then* $\sqrt{m}\left(\widehat{\alpha} - \alpha\right) \xrightarrow{d} N\left(0, \eta\alpha^2\right)$.

(B) If $m = cn^{2\beta/2\beta+\alpha}$ *then* $\sqrt{m}\left(\widehat{\alpha} - \alpha\right) \xrightarrow{d} N\left(\varphi\alpha, \eta\alpha^2\right)$ *for strictly positive and finite* $c = \left(\frac{a^{2\beta/\alpha}(\alpha+\beta)^2\alpha}{2b^2\beta^3}\right)^{\frac{\alpha}{2\beta+\alpha}}$ *and* $\varphi = sign(b)\left(2\beta/\alpha\right)^{-1/2}$.

see e.g. Hall (1982) or Haeusler and Teugels (1985) for the i.i.d. case ($\eta = 1$). More recently, convergence in distribution for the Hill estimator has also been proven in the presence of temporal dependencies ($\eta \neq 1$).[6]

[6] Hsing (1991) derives consistency and asymptotic normality under serially dependent data. Quintos et al. (2001) extend this result to stationary GARCH processes with conditionally normal innovations. Finally, Drees (2002) derives convergence in distribution for stationary time series processes exhibiting general forms of linear and nonlinear dependence. All these studies conclude that the asymptotic variance for dependent data differs from the i.d.d. variance (α^2). This explains the $\eta-$factor in Theorem 1.

Loosely speaking, Theorem 1 implies that proper convergence to a nondegenerate df requires m to rise with n at a "sufficiently slow" speed, i.e., $m, n \to \infty$ but $m/n \to 0$. This, however, does not hold when selecting a fixed fraction of extremes $\kappa = m/n$ (see e.g. Dumouchel (1983)). Previous studies - including QFP - have argued that this simple rule-of-thumb performs well in small samples but its lack of asymptotic justification constitutes a major problem. We will therefore renege from using this criterion.

Condition (B) of the convergence theorem provides an alternative way to selecting the nuisance parameter. It can be easily shown that the expression for the nuisance parameter m under (B) minimizes the Asymptotic Mean Squared Error (AMSE) for $\widehat{\alpha}$, see e.g. Danielsson and de Vries (1997). This AMSE minimization principle is exploited in virtually all empirical EVT studies to pick m^* and we will therefore use this criterion in the rest of the paper.[7] Theorem 1 also shows that the AMSE criterion induces a bias in the Hill statistic, i.e., $E(\widehat{\alpha} - \alpha) \sim m^{-1/2}\varphi\alpha$. We will thoroughly document the small sample consequences of this bias effect on the accompanying stability tests in the Monte Carlo simulation section. For more elaborate expositions on heavy tails and tail index estimators, see e.g. the monograph by Embrechts et al. (1997).

2.2 Structural change in tail behavior

We reconsider the small sample properties of a recursive, rolling and sequential procedure for detecting single unknown breaks in the Hill statistic $\widehat{\alpha}$ earlier introduced by Quintos et

[7] Condition (A) could also be used as a selection criterion. For example, choose $\delta > 0$ in $m^* = cn^{\frac{2\beta}{2\beta+\alpha} - \delta}$. However, although this criterion guarantees asymptotic unbiasedness, the bias is still present in small samples. Moreover, the small sample standard deviation increases with δ. If one cares more about bias than variance, (A) may be an interesting criterion for selecting m. In practice, however, researchers care both about bias and variance and prefer to trade-off bias and variance such as under condition (B). Finally, it is unclear how to choose δ.

al. (2001). Let t denote the endpoint of a subsample of size $w_t < n$. The recursive estimator uses subsamples $[1;t] \subset [1;n]$ and boils down to:

$$\widehat{\alpha}_t = \left(\frac{1}{m_t} \sum_{j=0}^{m_t - 1} \ln \left(\frac{X_{t-j,t}}{X_{t-m_t,t}} \right) \right)^{-1}, \tag{2.3}$$

with $m_t = ct^{\frac{2\beta}{2\beta+\alpha}}$. The rolling estimator is conditioned on a fixed subsample size $w^* < n$; the tail index is estimated by rolling over the subsample, i.e., the subsample is shifted through the full sample by eliminating past observations and adding future observations whilst keeping the subsample size constant at w^*.

$$\widehat{\alpha}_t^* = \left(\frac{1}{m_{w^*}} \sum_{j=0}^{m_{w^*} - 1} \ln \left(\frac{X_{w^*-j,w^*}}{X_{w^*-m_{w^*},w^*}} \right) \right)^{-1}, \tag{2.4}$$

with $m_{w^*} = c\,(w^*)^{\frac{2\beta}{2\beta+\alpha}}$ .Finally, the sequential estimator (denoted by $\widehat{\alpha}_{2t}$) is identical to the recursive estimator in (2.3) but calculated in reverse calendar time, i.e., using the more recent observations first.

The three tests can now be constructed using the sequences:

$$Y_n^2(r) = \left(\frac{tm_t}{n} \right) \left(\frac{\widehat{\alpha}_t}{\widehat{\alpha}_n} - 1 \right)^2, \tag{2.5}$$

$$V_n^2(r) = \left(\frac{w^* m_{w^*}}{n} \right) \left(\frac{\widehat{\alpha}_t^*}{\widehat{\alpha}_n} - 1 \right)^2, \tag{2.6}$$

$$Z_n^2(r) = \left(\frac{tm_t}{n} \right) \left(\frac{\widehat{\alpha}_t}{\widehat{\alpha}_{2t}} - 1 \right)^2, \tag{2.7}$$

with $r = t/n$ representing a fraction of the sample. Expressions (2.5) and (2.6) measure the fluctuation in the recursive and rolling values of the Hill statistic relative to their full sample counterpart $\widehat{\alpha}_n$ whereas the sequential test uses (2.7) to compare the fluctuations of the recursive with the reverse recursive estimator. The null hypothesis of interest is that the tail index α does not exhibit any temporal changes. The null hypothesis of constancy then takes the form

$$H_0 : \alpha_{[nr]} = \alpha, \quad \forall r \in R_\varepsilon \ = [\varepsilon; 1-\varepsilon] \subset [0;1] \ , \tag{2.8}$$

with $[nr]$ representing the integer value of nr. Without prior knowledge on the direction of a break, one is interested in testing the null against the two-sided alternative hypothesis $H_A : \alpha_{[nr]} \neq \alpha$. For practical reasons the above test is calculated over compact subsets of $[0;1]$, i.e., t equals the integer part of nr for $r \in R_\varepsilon = [\varepsilon; 1-\varepsilon]$ and for small $\varepsilon > 0$. Sets like R_ε are often used in the construction of parameter constancy tests (see, e.g., Andrews, 1993).[8] Conform with Quandt's (1960) seminal work on structural change tests for linear regression models, the candidate break date r is selected where the testing sequences (2.5), (2.6) and (2.7) reach their supremum. This renders the most likely time point for the constancy hypothesis to be violated.

3 Monte Carlo experiments

We investigate the small sample behavior of the recursive, rolling and sequential test for a variety of stochastic models - both for the conditional and the unconditional df - used in the modelling of financial time series. Each model exhibits regularly varying tails and obeys the asymptotic second order expansion in eq.(2.1). The number of upper order extremes for the Hill statistic minimizes the Asymptotic Mean Squared Error of the Hill estimator. We calculate (size-corrected) small sample power against a variety of realistic break scenarios

[8]The restricted choice of r implies that $\varepsilon n \leq t \leq (1-\varepsilon)\, n$. When the lower bound would be violated the recursive estimates might become too unstable and inefficient because of too small sub-sample sizes. On the other hand, the test will never find a break for t equal or very close to n, because the test value (2.5) is close to zero in that latter case. Thus, for computational efficiency one might stop calculating the tests beyond the upper bound of $(1-\varepsilon)\, n < n$. Conform with Andrews (1993), we set the trimming value $\varepsilon = 0.15$.

as alternative hypotheses. Last but not least, we report simulated break estimates averaged over the statistically significant breaks at the 95 percent significance level.

A short description of the main data generating processes is provided in subsection (3.1). The analytic derivation of the nuisance parameters for these DGP's is discussed in subsection (3.2). Finite sample critical values and size-corrected power properties are discussed in subsection (3.3) and (3.4), respectively. Finally, small sample bias in breakpoint estimates when the true break lies at different breakpoint locations is evaluated in (3.5).

3.1 Data generating processes

We choose a variety of heavy tailed data generating processes (DGP's) and accompanying parameter configurations (a, b, α, β) in eq. (2.1). A Monte Carlo experiment is based on the symmetric stable df, Student-t, Frechet df (Type-II extreme value), i.e. $P\{X > x\} = \exp(-x^{-\alpha})$, Burr df, i.e. $P\{X > x\} = \left(1 + x^{\beta}\right)^{-\alpha/\beta}$, AR(1) with stable innovations, GARCH(1,1) with conditionally normal errors and a rudimentary Stochastic Volatility model. Thus, we distinguish between i.i.d. models and dependent models.

Symmetric stable draws are obtained using the algorithm proposed by Samorodnitsky and Taqqu (1994):

$$X_{stable} = \frac{\sin \alpha\gamma}{(\cos\gamma)^{1/\alpha}} \left(\frac{\cos(1-\alpha)\gamma}{W} \right), \tag{3.1}$$

where $0 < \alpha < 2$ represents the tail index. The parameter γ is drawn uniformly on $[-\pi/2\ \pi/2]$ whereas W is exponentially distributed with mean 1.[9] As for the the tail index

[9]The symmetric stable df has some drawbacks if one wants to use it as a model for financial asset returns. First, their additivity feature (sums of stable dfs remain identically distributed after appropriate scaling) seems overly restrictive: Feller (1971, ch. VIII.8) shows that the class of regularly varying dfs only exhibit additivity in the tail area. Also, less heavy tailed models may be preferable to capture financial asset tails given the fact that the heavy tailed symmetric stable class fails to have a finite variance ($\alpha < 2$), Finally,

of the Student-t, Frechet and Burr dfs, it is varied between 2 and 4 which is the range one typically observes for stock market tails, see e.g. Jansen and de Vries (1991) or Hartmann et al. (2004).

We also allow for models that exhibit linear and nonlinear dependence because the i.i.d. assumption is too restrictive for financial return data. An AR(1) process with first order autocorrelation $\theta = 0.1$ and with symmetric stable innovations is used to generate serially correlated data.[10] The choice for stable innovations is motivated by their invariance property under addition.[11] In order to generate persistence in volatility, we use two distinct models. First, we implement a simple stochastic volatility model from Danielsson et al. (2001):

$$\begin{aligned} Y_t &= U_t \sqrt{\frac{v}{\chi^2(v)}} H_t, \qquad P\{U_t = -1\} = P\{U_t = 1\} = 0.5 \\ H_t &= \beta Q_t + \theta H_{t-1}, \quad Q_t \sim N(0,1), \quad \beta = 0.1, \ \theta = 0.9 \end{aligned} \tag{3.2}$$

The unconditional df is Student-t with $\alpha = v$ degrees of freedom. The multiplicative factor U_t guarantees the fair game property $E_{t-1}(Y_t) = 0$ (without this factor, the model both exhibits dependence in the first and the second moment). Second, we also simulate from

the normal df can act as a "local alternative" for the stable model. Indeed, let α stand for the tail index as defined earlier. For $\alpha < 2$, α determines the maximal number of bounded moments up to α but when $\alpha = 2$ (the case of the normal df in (3.1)) all moments exist and α is not interpretable as the tail index. Thus, stable processes with α close to 2 can only be distinguished from a normal df on the basis of α-estimates in very large samples.

[10] At low return frequencies (daily, weekly) empirical studies typically do not find a statistically significant autocorrelation in financial return series which is consistent with the weak version of the market efficiency hypothesis. On the contrary, market microstructure effects in high frequency data might induce statistically significant first order serial correlations, see e.g. Andersen and Bollerslev (1997).

[11] An AR(1) process $X_t = \theta X_{t-1} + u_t$ with first order serial correlation $0 < \theta < 1$ is equivalent to the MA(∞) process $X_t = \sum_{i=0}^{\infty} \theta^i u_{t-i}$. If the innovations u_{t-i} are i.i.d. symmetric stable, it follows from Feller (1971, ch. VIII.8) that $X_t \overset{d}{=} \left(1 + \theta^\phi + \theta^{2\phi} + \cdots\right)^{1/\phi} u_t = \left(\frac{1}{1-\theta^\phi}\right)^{1/\phi} u_t$. Thus the $AR(1)$ dependent stable draws X_t and the i.i.d. stable innovations u_t exhibit the same distribution upon some scaling constant.

a GARCH(1,1) model with conditionally normal innovations. The sum of the GARCH volatility parameters $\theta = \beta_0 + \beta_1$ is chosen such that the tail index of the corresponding unconditional df either equals 4 (details on the relation between the GARCH parameters and the tail index are provided in Appendix B).

3.2 Choice of optimal number of extremes

Tail index estimators like the Hill statistic imply a bias/variance trade-off, i.e., the more data used from the distributional center the smaller will be the variance of the estimator but the more bias will be introduced. Goldie and Smith (1987) therefore proposed to select the "optimal" m in (2.2) by minimizing the asymptotic Mean Squared Error (AMSE) of the Hill statistic. Using the second order expansion for regularly varying dfs in (2.1), Danielsson and de Vries (1997) derived an expression of the AMSE of $\widehat{\alpha}_{HILL}$ in terms of the second order expansion parameters:

$$AMSE(\widehat{\alpha}_{HILL}, m) = a^{-2\beta/\alpha} \frac{1}{\alpha^2} \frac{\beta^2 b^2}{(\alpha+\beta)^2} (\frac{m}{n})^{\frac{2\beta}{\alpha}} + \frac{1}{\alpha^2 m}, \tag{3.3}$$

where the first part is the squared bias and the second part is the variance. The above expression shows that the second order parameters b and β are responsible for the bias in the Hill statistic, i.e., if either b or β equals zero, the bias term disappears and the distributional tail in (2.1) specializes to an exact Pareto.

Minimizing (3.3) w.r.t. m renders the optimal number $m^*$of highest order statistics

$$m^* = cn^{2\beta/2\beta+\alpha}, \qquad c = \left(\frac{\alpha(\alpha+\beta)^2}{2\beta^3 b^2} a^{2\beta/\alpha} \right)^{\frac{\alpha}{2\beta+\alpha}}. \tag{3.4}$$

which is the same expression as under condition (B) of Theorem 1. The optimal threshold path implies that smaller fractions of upper order extremes m/n will be selected when the sample size n grows large.

Table 1 reports the parameter vectors (a, b, α, β) for the tail expansions of all considered DGP's. The values can be used to derive the analytic expressions for m^*(see eq.(3.4)).

[Insert Table 1]

Details on the accompanying tail expansion derivations are provided in Appendix A. A number of interesting observations can be made from Table 1. First, the popular restriction $\alpha = \beta$ does not seem to hold for all dfs. The second order parameter β for the Student-t equals 2 regardless the degrees of freedom parameter. As for the Burr df, β can be varied and chosen independent from α and is therefore suited as a vehicle to study the impact of changing higher order tail behavior in (2.1). The table also reveals that b can both be positive and negative which implies that the sign of the asymptotic Hill bias, $sign(b)$, differs across different models. In the next subsection, the relation between the bias and the sign of this parameter will be further clarified when discussing the simulation results.

The Monte Carlo study for small sample critical values will be calculated using the analytic expressions for m^*. However, if analytic expressions for m^*do not exist (e.g. the GARCH model, real data or processes with breaks when evaluating power and break date ability), we minimize the Beirlant, Dierckx, Goegebeur and Matthys (1999) sample equivalent of the Asymptotic Mean Squared Error (AMSE) in order to select the optimal m^*. These authors derived an exponential regression model for the log-spacings of upper order statistics from regularly varying tails:

$$j\left(\ln X_{n-j+1,n} - \ln X_{n-j,n}\right) \sim \left(\gamma + d_{n,m}\left(\frac{j}{m+1}\right)^{-\rho}\right) f_j, \quad 1 \leq j \leq m. \tag{3.5}$$

Here $\gamma = 1/\alpha$, $\rho = -\beta/\alpha$, $(f_1, f_2, \cdots, f_m)$ is a vector of independent standard exponential random variables, and $d_{n,m}$ stands for $d\left(\frac{n+1}{m+1}\right)$, $3 \leq m \leq n/3$. The asymptotic variance and the asymptotic bias for the inverse of the Hill statistic $\widehat{\gamma} = 1/\widehat{\alpha}$ can be approximated by

$\sigma^2(\widehat{\gamma}) \sim \gamma^2/m$ and $E(\widehat{\gamma}-\gamma) \sim \frac{d_{n,m}}{1-\rho}$. The Asymptotic Mean Squared Error (AMSE) for the Hill statistic $\widehat{\gamma}$ can now be estimated for different values of m :

$$AMSE(\widehat{\gamma}) = \left(\frac{d_{n,m,LS}}{1-\rho_{LS}}\right)^2 + \frac{\gamma_{LS}^2}{m},$$

which is typically U-shaped as a function of m due to the bias-variance tradeoff. The estimators γ_{LS}, ρ_{LS} and $d_{n,m,LS}$ refer to Ordinary Least Squares estimators of the corresponding parameters in the nonlinear model (3.5). The optimal sample fraction m^* is then estimated as the one where AMSE reaches its minimum, i.e., $m^* = \arg\min_m [AMSE(\widehat{\gamma})]$.[12]

For sake of convenience, we make two additional assumptions when implementing the Beirlant et al. (1999) criterion. First, we impose the restriction $\alpha = \beta$ ($\rho \equiv -\beta/\alpha = -1$) on the parameters of the tail expansions; this circumvents the need of separate β-estimation. Obtaining stable and accurate estimates of the second order parameter has been shown to be notoriously difficult (see, e.g., Gomes, de Haan and Peng (2002); Gomes and Martins(2002)) which makes it beneficial to impose the restriction. Moreover, simulations have shown that the Beirlant criterion still performs well in determining m^*, even for DGP's where $\alpha \neq \beta$. Second, we do not apply the Beirlant optimization criterion on each recursive, rolling or sequential subsample considered in (2.5)-(2.6)-(2.7) separately. Instead, we determined the full sample $\widehat{m}$ which implies that the full sample scaling constant c in (3.3) follows by $\widehat{c} = \widehat{m}/n^{2/3}$. Upon extrapolating the optimal path for m to the subsamples defined by the stability tests (and using the notation from section 2.2), we obtain $\widehat{m}_t = \widehat{c}t^{2/3}$ for the recursive and sequential tests and $\widehat{m}_{w^*} = \widehat{c}(w^*)^{2/3}$ for the rolling test, respectively. Thus, for sake of simplicity we assume that c does not change across subsamples and that it can be set equal to its full sample value.

[12]Subsample bootstrap algorithms (see, e.g., Danielsson, de Haan, Peng, de Vries (2001) to select m^* by means of AMSE minimization constitute an alternative way; but these are only applicable for sample sizes that are larger than the ones we employ in the Monte Carlo section and the empirical application.

3.3 Monte Carlo results

From Theorem 1 it straightforwardly follows that the Hill statistic - when applied to regularly varying tails and conditioned on a nuisance parameter m that minimizes AMSE - exhibits bias and estimation risk in small samples. In this section we investigate to what extent bias and variance properties are transmitted to the small sample critical values and power of the considered stability tests based on $\widehat{\alpha}$. We also want to establish whether the bias problem erodes the ability of the tests to rightly locate break dates. To this aim, we simulate from the set of models that have been introduced in the previous section.

3.3.1 Small sample properties of the Hill statistic

As a benchmark for comparison with the rest of the simulations, we start by shortly reconsidering the small sample behavior of the Hill estimator. Table 2 reports simulated means and standard deviations of the Hill statistic for differing sample sizes n. Averages and standard deviations are calculated over 5,000 replications.

[Insert Table 2]

Table 2 is divided into a left/right panel and an upper/lower panel. The left panel contains Hill estimates for "optimal" m in minimal AMSE sense whereas the right panel conditions the Hill statistic on a fixed percentage of tail observations (i.e., Dumouchel's rule). We further distinguish between models that either generate dependent or independent draws (lower and upper panel, respectively). In the upper panel, we let α and β vary; in the lower panel, the degree of serial correlation or volatility clustering is manipulated *ceteris paribus* α and β.

First and foremost, one can see that deviations from unbiasedness $|E(\widehat{\alpha}) - \alpha|$ only decrease when m^* is chosen "optimally" in the sense that AMSE is minimized. Indeed, the right hand side (RHS) panel estimates ($m = 0.1n$) seem to diverge away from the true underlying value of α when the sample size is increased. This should not surprise given that Dumouchel's rule does not guarantee proper convergence in distribution of the Hill statistic (see Theorem 1). The optimal m^* results in the left panel show a large heterogeneity in small sample bias and estimation accuracy across different distributions. As predicted by theorem 1, the sign of the bias corresponds with the sign of b. Indeed, from Table 1, we know that b is only positive for the stable class which explains the positive Hill bias in Table 2 for stable draws and the negative bias for all other classes of dfs. One also observes that the deviation from unbiasedness as well as the corresponding standard deviation of the Hill statistic is smaller for heavier tails (lower values of α). The intuition behind this result is that lighter tails are closer to a thin tailed local alternative like the normal distribution that does not satisfy (2.1). This decreases the accuracy - both in terms of bias and standard deviation - of tail estimation techniques that assume regular variation as a starting point. It is also worth noticing what happens when the second order parameter β changes for given values of α. Only in the Burr distribution case, we can let β evolve independently from α. The Burr outcomes reveal that the bias and standard error of $\widehat{\alpha}$ decrease for higher values of β, i.e., the closer the tail expansion (2.1) approximates a pure Pareto tail the smaller will be the bias and estimation risk. The lower table panel reports the impact of temporal dependence on bias and variance properties of the Hill statistic. Both higher serial correlation in the AR(1) processes as well as a higher persistence in volatility clustering (Stochastic Volatility and GARCH model class) seem to increase the deviation from unbiasedness as well as the standard deviation.

Next we investigate to what extent the Hill bias is transferred into the size and power properties of the stability tests as well as their ability to accurately identify break dates, i.e., are the stability test properties very different for the high bias/variance cases as compared to the low bias/variance cases? Tables 3 and 4 report simulated critical values for i.i.d. models and models that exhibit temporal dependence, respectively. Each table is further split in three panels containing the small sample distributional quantiles for the recursive, rolling and sequential tests presented in (2.5)-(2.6)-(2.7). The quantiles of the test statistics are calculated as follows. For samples of size T=500, 2000 we first generate 20,000 simulations from the considered DGP's. These Monte-Carlo simulations are used to obtain estimates of the 90th, 95th and 99th percentile of the stability tests' small sample distribution. In order to further reduce the variability of the simulated quantiles, this process is repeated 10 times. The final critical value estimates are the averages over these 10 replications. The average quantile estimates are reported in the tables (corresponding standard deviations between brackets).[13]

[Insert Tables 3 and 4]

The heterogeneity in the small sample critical values across different DGP's is nearly one-to-one with the preceding table results on bias and estimation risk for the Hill estimator: critical values and their estimation risk are higher for those cases that exhibit a higher bias in the Hill estimator. More specifically, higher values of the tail index α and the persistence parameter θ (either standing for serial correlation or volatility persistence) increase the critical values whereas higher values of the second order parameter β (cf. Burr df) decrease the critical values. Thus, the tables provide convincing evidence that the bias in the Hill estimator is transferred into the critical values. In other words, "globally valid" critical values that can be applied to all financial return tails do not seem to exist, even for medium size samples ($n = 2,000$). The critical values for $\rho = -5$ actually come close to the asymptotically

[13]Reported are the standard deviations of the quantile levels q (and not the averages $\overline{q}$). Assuming for sake of simplicity that the 10 simulated quantile estimates are independent, the reduction in estimation risk is given by $\sigma(\overline{q}) = \sigma(q)/\sqrt{10}$.

unbiased critical values reported in the Quintos et al. (2001) paper. This is not so surprising because the Burr tail becomes indistinguishable from a pure Pareto df in the latter case (Hill estimators are unbiased with Pareto data). But the table also convincingly shows that using asymptotically unbiased critical values leads to large overrejection of the null of parameter constancy in the majority of cases, i.e., when the true critical values are upward biased due to the asymptotic Hill bias.

Next, Tables 5-7 and 8-10 report small sample power and estimates of the break points for the recursive, rolling and sequential stability test, respectively. We consider sudden upward and downwards jumps in α of different magnitudes and at different points in time (r=0.25, 0.50, 0.75). The power is based on 20,000 replications and is size-adjusted using the adequate small sample critical values from the previous table. The breakpoint estimates are also based on 20,000 replications but the estimated breaks are averaged over statistically significant values of $\widehat{r}$ only, significance being determined by the 95% small sample critical values from Tables 2-3.[14]

[Insert Tables 5 to 10]

The recursive and rolling test both exhibit satisfactory power if α decreases;however, the power of the rolling test is larger in detecting an increase in α. The latter result can be understood by observing that eq.(2.2) is based on the m largest observations so that extremal returns occurring in the initial recursive sample will partly remain in the selection of the m highest order statistics when the sample size is increased. This initial extremes dominance when $\alpha_1 < \alpha_2$ does not occur for the rolling test since the influence on $\widehat{\alpha}$ of extremal behavior that occurs in the initial sample gradually drops out when the rolling

[14]For break scenarios (α_1, α_2) we calculate the power and break estimates using the 95% small sample critical value that corresponds with $\min(\alpha_1, \alpha_2)$.

window is shifted through the total sample. The sequential test seems to do poorly, although the power differs quite a lot depending on the location of the break and the direction of the change in α. As concerns the ability to date breaks, the recursive test clearly outperforms the other ones for most considered DGP's provided the break scenario implies an increase in tail fatness ($\alpha_1 > \alpha_2$).[15] However, the recursive test's inability to detect breaks when $\alpha_1 < \alpha_2$ is more apparent than real. Indeed, if one lacks prior knowledge on the direction of the jump in the tail index (as is the case in most empirical applications), the recursive test can be performed both in calendar time ("forward" recursive test) as well as by inverting the sample ("backward" or "reverse" recursive test). A decrease in the tail index - if present in the data - should then be signaled by the forward version of the recursive test whereas an increase should be detected by the backward version of the recursive test. This is also the strategy we will implement in the empirical application.

Sofar the general discussion on power and break date ability. One can also often observe large differences in power results and break point detection across different DGP's. This heterogeneity can again be explained by the determinants of the Hill bias. More specifically, higher values of the persistence parameter θ (either standing for serial correlation or volatility persistence) increase the Hill bias and the bias in the estimated break dates and decrease the power. On the other hand, higher values of the second order parameter β (cf. Burr df) decrease the Hill bias and the bias in the estimated break dates and increase the power. Thus, the tables provide convincing evidence that the bias in the Hill estimator is also influencing the stability tests' power and ability to date breaks. Indeed, the power for the Burr case with $\rho = -5$ lies close to 100%, even in small samples whereas bias and estimation risk for $\widehat{r}$ are negligibly small.

[15]The power and break date results show that satisfactory power is a necessary but not sufficient condition for accurate breakpoint detection. The rolling test for $\alpha_1 < \alpha_2$ provides a nice illustration.

4 Empirical Application

We perform stability tests for a large cross section of 21 developed and emerging stock markets. Given the more frequent occurrence of institutional, political and financial crises in emerging markets, it is interesting to see whether the latter markets are more prone to structural change in the tail behavior than developed markets. Daily prices indices (excluding dividends) denominated in local currency were obtained from Datastream inc. Returns are calculated as log first differences and our sample ranges from January 1, 1988 until August 13, 2007 which amounts to $5,117$ daily prices.

Given the poor properties of the rolling and sequential test in terms of power and ability to detect breaks, we implement the recursive test in the empirical application. From the Monte Carlo investigation, we know that small sample critical values can differ considerably depending on the model of regular variation assumed for the tail. This is because the severity of the Hill bias and its resulting influence on the stability tests'small sample critical values largely differs across different parametrizations. In order to avoid this "model risk" related to a Monte Carlo simulation or a parametric bootstrap we opted for a bootstrap-based semi-parametric approach towards determining the critical values of the test.

The previous simulation section has also illustrated that temporal dependence in the data increases the critical values of the considered stability tests. Moreover, it is well known that stock returns exhibit nonlinear dependencies like GARCH effects (volatility clustering). Upon assuming that GARCH-type volatility clustering constitutes the main source of temporal dependence, we implement a GARCH-corrected version of the recursive test

$$Q_{r \in R_\tau} = \sup \widehat{\eta_t}^{-1} Y_n^2\left(t\right), \tag{4.1}$$

where $\widehat{\eta}_t$ is the estimate of the time varying scaling factor, see Quintos et al. (2001, p. 643). Subsequently, small sample critical values can be bootstrapped (CV_B) at the 90, 95 and 99 percent levels.[16] Upon assuming that the scaling factor corrects the test for any temporal dependence, the bootstrap does not have to take account of this data feature and we can resort to a "wild" version of the bootstrap instead of a block bootstrap.[17]

The simulation section already showed that the stability tests' critical values can be quite unstable in small samples. We try to reduce this sampling variability in the same way as we did in the Monte Carlo experiments: first, we bootstrap $20,000$ replication samples. Each of these samples render an estimate of the 90^{th}, 95^{th} and 99^{th} percentile of (4.1). Second, we repeat this bootstrap and estimation procedure 10 times. The final critical value estimates are the averages over the latter 10 replications.

Table 11 reports the testing outcomes and break dates (if any). The table further includes the forward (Q_F) and backward (Q_B) version of (4.1) which enables us to distinguish rises in $\widehat{\alpha}$ from drops - if present. The number of upper order extremes m^* is determined using the Beirlant method. Evidently, bootstrapped critical values are identical for the forward and backward test.

[Insert Table 11]

The null of parameter constancy is rejected if the sup-value exceeds the critical values, e.g. $Q > CV_B(p)$ with $p = 10\%$, 5% or 1%.

[16] Davidson and Flachaire (2007) argue that the asymptotic properties of the bootstrap in an EVT context may be hampered when time series do not possess a finite second moment ($\alpha < 2$). An example constitutes the class of symmetric stable distributions. However, the large majority of empirical studies on the magnitude of the tail index for stock markets find that $\widehat{\alpha}$ hovers around 3 and is significantly above 2, see e.g. Jansen and de Vries (1991) or Longin (1996).

[17] A bootstrap in blocks would be appropriate in case one would not have corrected the recursive test for temporal dependence effects like volatility clusters. However, to the best of our knowledge, there is no rule-of-thumb available yet for choosing the optimal block length in an Extreme Value Theory (EVT) framework. That is why we first scaled the test such that we are allowed to randomly reshuffle the data afterwards.

The table first of all shows that the critical values are very much stock market dependent which seems to confirm that the application of one set of critical values to test the null of parameter constancy would not be wise. Moreover, the emerging small sample critical values seem to dominate their developed counterparts. Most importantly, however, the empirical evidence for breaks in α is very limited: the backward recursive test outcomes suggest an increase in α (thinner tails) for Taiwan and Hong Kong but evidence lacks for changes in tail behavior for any other markets (either developed or emerging). The "thinning" of the tail for the two mentioned emerging markets may be due to a stabilization in the market brought by institutional reform and liberalization in the aftermath of the Asian crisis.

That the (limited) empirical evidence for structural change in α is generated by emerging market data is not surprising given the degree of institutional reform, liberalization experiments, regime changes in monetary and exchange rate policies and last but not least the vehement financial turmoil that has frequently hit these markets. On the other hand, most emerging markets - as well as all developed markets - exhibit stationary tail behavior. The empirical results leave us with the conviction that heavy tails are invariant over time, i.e., the tail index α constitutes a long run characteristic of financial series that can be used for the assessment of long run risk, stress testing and financial stability.

5 Conclusions

This paper provides a thorough study of the small sample behavior of some popular tests for detecting time variation in the tail index of financial returns. The tests are "endogenous" in the sense that they produce an estimate of the breakpoint location upon detection of a statistically significant break. Our Monte Carlo experiment determines critical values, size-corrected power and the ability to date breaks for a myriad of Data Generating Processes

(DGP's). The tests all use the Hill estimator for the tail index as an input. In line with the bulk of the empirical literature, the number of upper order extremes is selected by minimizing the (asymptotic or sample) Mean Squared error of the Hill statistic. The DGP's are chosen with an eye towards mimicking some popular empirical stylized facts of financial data like volatility clustering and serial dependence.

We find that the small sample critical values differ a lot across different distributional models and sample sizes. More specifically, critical values are higher when the bias in the Hill estimator is more severe. Moreover, the bias in the Hill estimator and the critical values persist in large samples and removing it can be shown to be notoriously difficult. Thus, applying the same critical values to different DGP's or empirical datasets seems wrong and almost surely implies the use of a size distorted structural change test. We therefore propose a bootstrap-based procedure to determine the critical values of the stability test when using real-life data. Using bootstrap-based small sample critical values for the stability tests - at least a recursive-based version - renders satisfactory power and ability to detect breaks. Applying the recursive-based test on a large set of emerging and developed stock market data, we are hardly able to detect breaks in the tail behavior. Otherwise stated, the tails of the unconditional distribution of stock market returns seem to be relatively unchanged over time. This seems to confirm most previous empirical applications on univariate EVT: the tail index of financial returns typically fluctuates between 2 and 4 regardless the asset, asset class or time period considered.

A Derivations of 2nd order expansion parameters

In Theorem 1 we argued that $m = cn^{2\beta/2\beta+\alpha}$ is the optimal nuisance parameter for the Hill statistic that minimizes the AMSE$(\widehat{\alpha})$. The scaling constant c in turn depends on the parameters (a, b, α, β) of the second order tail expansion (2.1). Thus, the parameters and the resulting m^* are uniquely determined upon knowledge of this tail expansion. To simplify their derivation it is instructive to re-express the tail expansion (2.1) for $p = x^{-1}$ close to zero:

$$1 - G(p) = ap^{\alpha}\left(1 + bp^{\beta} + o(p^{\beta})\right), \tag{A.1}$$

with $a > 0$, $b \in \Re$, $\beta > 0$ and $F(x) = G(p)$. In the Monte Carlo section we show that biases in the Hill estimator, the stability tests' critical values and the breakpoint estimates are critically determined by the level of b and β. The pure Pareto model ($b = 0$ and/or $\beta \sim \infty$) provides the benchmark case because it renders unbiased Hill estimates, test statistics and break point estimates.

The parameters a, b and β easily follow by either expanding the cumulative distribution $G(p)$ (c.d.f.) (if it exists in closed form) or the accompanying density around $p = 0$. The Frechet and Burr dfs have c.d.f.'s in closed form which implies that their respective second order Taylor expansions for p small (x large) straightforwardly follow as

$$\begin{aligned} 1 - G_{FRECHET}(p) &= 1 - \exp(-p^{\alpha}) \\ &\simeq p^{\alpha}\left(1 - \frac{1}{2}p^{\alpha}\right), \qquad p \text{ small} \\ &= x^{-\alpha}\left(1 - \frac{1}{2}x^{-\alpha}\right), \quad x \text{ large} \end{aligned}$$

and

$$(a, b, \beta)_{FRECHET} = (1, -1/2, \alpha)$$

As for the Burr distribution, the 2nd order expansion for the c.d.f. reads

$$\begin{aligned} 1 - G_{BURR}(p) &= \left(1 + p^{-\beta}\right)^{-\alpha/\beta} \\ &\simeq p^{\alpha}\left(1 - \frac{\alpha}{\beta}p^{\beta}\right), \qquad p \text{ small} \\ &= x^{-\alpha}\left(1 - \frac{\alpha}{\beta}x^{-\beta}\right), \quad x \text{ large} \end{aligned}$$

which implies

$$(a,b)_{BURR} = (1, -\alpha/\beta)$$

Clearly, whereas first order and higher order behavior are related ($\beta = \alpha$) in the Frechet case, the 2nd order parameter can be freely chosen in case of the Burr df. This implies that the Burr distribution becomes indistinguishable from a pure Pareto distribution for large β.

The other DGP's do not exhibit explicit c.d.f. in closed form which somewhat complicates the derivation of the second order parameters. For the symmetric stable class neither the c.d.f. nor the density exists in closed form but tail expansions are already readily availabe in the statistical literature:

$$1 - F(x) = \pi^{-1} \sum_{i=1}^{\infty} (-1)^i \frac{\Gamma(i\alpha)}{i! x^{i\alpha}} \sin\left(\frac{i\alpha\pi}{2}\right), \quad x \text{ large}$$

see (Ibragimov and Linnik (1971, ch. 2)). Only considering the expansion's first two terms renders the second order approximation:

$$1 - F(x) \simeq \frac{1}{\pi}\Gamma(\alpha)\sin\left(\frac{\alpha\pi}{2}\right) x^{-\alpha}\left(1 - \frac{\Gamma(2\alpha)\sin(\alpha\pi)}{2\Gamma(\alpha)\sin\left(\frac{\alpha\pi}{2}\right)} x^{-\alpha}\right),$$

and the parameter vector that we need for determining m^* easily follows:

$$(a;b;\beta)_{STABLE} = \left(\frac{1}{\pi}\Gamma(\alpha)\sin\left(\frac{\alpha\pi}{2}\right); -\frac{\Gamma(2\alpha)\sin(\alpha\pi)}{2\Gamma(\alpha)\sin\left(\frac{\alpha\pi}{2}\right)}; \alpha\right)$$

Notice that both the Frechet df and the symmetric stable class are characterized by the restriction $\alpha = \beta$.

In order to determine the tail expansion parameters for the Student-t, we need to expand the tail density $g(p)$ because the c.d.f. does not exist in closed form. The 2nd order Taylor expansion of the student-t density for large x (small p) boils down to:

$$
\begin{aligned}
f(x) &= \frac{\Gamma\left(\frac{\alpha+1}{2}\right)}{\Gamma\left(\frac{\alpha}{2}\right)\sqrt{\pi\alpha}}\left(1+\frac{x^2}{\alpha}\right)^{-\frac{\alpha+1}{2}} \\
&\simeq \frac{\Gamma\left(\frac{\alpha+1}{2}\right)}{\Gamma\left(\frac{\alpha}{2}\right)\sqrt{\pi\alpha}}\alpha^{\frac{\alpha-1}{2}}\alpha x^{-\alpha-1}\left[\alpha-\frac{\alpha^2(\alpha+1)}{2(\alpha+2)}(\alpha+2)x^{-3}\right], \qquad x \text{ large }, \qquad \text{(A.2)}
\end{aligned}
$$

As for the asymptotic expansion for the class of regularly varying densities, it easily follows from (A.1):

$$
\begin{aligned}
G'(p) &= g(p) \simeq a\alpha p^{\alpha+1} + ab(\alpha+\beta)p^{\alpha+\beta+1}, \qquad p \text{ small} \\
&= ax^{-\alpha-1}\left(\alpha+b(\alpha+\beta)x^{-\beta-1}\right), \qquad x \text{ large.} \qquad \text{(A.3)}
\end{aligned}
$$

Comparing (A.3) with (A.2) directly renders the parameter vector

$$
(a,b,\beta)_{STUDENT} = \left(\frac{\Gamma\left(\frac{\alpha+1}{2}\right)}{\Gamma\left(\frac{\alpha}{2}\right)\sqrt{\pi\alpha}}\alpha^{\frac{\alpha-1}{2}}; -\frac{\alpha^2(\alpha+1)}{2(\alpha+2)}, 2\right).
$$

It also follows from the second order term between brackets that the restriction $\beta = 2$ holds for the symmetric stable class.

Finally, it can be easily shown that the serially correlated stable draws (denoted by ARSTA in the tables) and the student-t draws that exhibit dependence in the second moment (SVSTU) in eq. (3.2) exhibit the same optimal m^* as their i.i.d. stable and student-t counterparts. The additivity property under addition for the symmetric stable df ensures that the serially dependent stable draws exhibit the same distribution as the i.i.d. symmetric stable upon some scaling constant. The Student-t draws that exhibit dependence in the second moment also exhibit the same distribution as an i.i.d. student-t process upon some scaling constant. In general, a linear transform $\widetilde{X} = tX$ that changes the scaling constant leaves the tail index and the optimal value of upper order extremes invariant. This directly follows from the tail expansion for $\widetilde{X}$:

$$\begin{aligned} P\left\{\widetilde{X} > x\right\} &= P\left\{X > t^{-1}x\right\} \\ &\simeq at^{\alpha}x^{-\alpha}\left(1 + bt^{\beta}x^{-\beta}\right) \end{aligned}$$

which implies that $\widetilde{a} = at^{\alpha}$ and $\widetilde{b} = bt^{\beta}$. The parameters α and β are left unchanged by the linear transform. Substituting $\widetilde{a}$ and $\widetilde{b}$ into $c = \left(\frac{\alpha(\alpha+\beta)^2}{2\beta^3 b^2} a^{2\beta/\alpha}\right)$ leaves the value of $m^* = cn^{\frac{2\beta}{2\beta+\alpha}}$ invariant.

Table 1 in the main body of the text summarizes the tail expansion parameters (a, b, α, β)

B Calibration of GARCH(1,1) parameters

In order to generate the clusters of volatility feature for the conditional df, we simulated, inter alia, from a GARCH(1,1) process with conditionally normal disturbances. Let X_t follow a GARCH(1,1) process, then

$$\begin{aligned} X_t &= \sigma_t Z_t \\ \sigma_t^2 &= \beta_0 + \beta_1 \sigma_{t-1}^2 + \lambda X_{t-1}^2 \\ Z_t &\sim i.i.d. \text{ N(0,1)} \end{aligned}$$

It can be shown that the GARCH scheme also induces the fat tail property on the unconditional distribution of the returns, see e.g. Embrechts et al. (1997). Also, the tail index α is a function of the parameters of the model. Given the normality of Z_t and provided $\beta_1 + \lambda < 1$, one can show that α is related to the parameters of the conditional df:

$$E\left(\lambda Z^2 + \beta_1\right)^{\alpha/2} = 1, \tag{B.1}$$

see e.g. Mikosch and Starica (1998). Empirical evidence suggests that $2 < \alpha < 4$ for most financial asset classes and we therefore use these boundary values in the Monte Carlo simulations. When $\alpha = 2$, eq. (B.1) implies that $\lambda + \beta_1 = 1$. This still leaves us with an infinite number of possible parameter combinations. For sake of simplicity, we will calibrate $(\lambda, \beta_1) = (1/2, 1/2)$ in the simulation section for the $\alpha = 2$ case. As for the upper bound value $\alpha = 4$, eq. (B.1) boils down to

$$3\lambda^2 + 2\beta\lambda + \beta^2 - 1 = 0 \tag{B.2}$$

Substituting $\beta_1 = c - \lambda$ $(c < 1)$ into (B.2), one obtains a quadratic equation in λ :

$$2\lambda^2 = 1 - c^2$$

It follows that for a given value of c, the clusters of volatility parameters in the GARCH(1,1) model are uniquely identified, i.e., $(\lambda, \beta_1) = \left(\sqrt{\frac{1-c^2}{2}}, c - \sqrt{\frac{1-c^2}{2}}\right)$. In empirical studies one often encounters $\beta_1 + \lambda$ close to 1. We therefore set c equal to 0.75, 0.85, or 0.95 in the Monte Carlo section to investigate the impact of different degrees of volatility persistence on the test statistics. The intercept β_0 is set to 10^{-6} which is in line with previous simulation studies, see e.g. Danielsson and de Vries (1999).

References

[1] Andersen, T., and Bollerslev, T. (1997), "Intraday periodicity and volatility persistence in financial markets," *Journal of Empirical Finance*, 4, 115-158.

[2] Andrews, D.W.K. (1993), "Tests for Parameter Stability and Structural Change with Unknown Change Point," *Econometrica*, 59, 817-858.

[3] Beirlant, J., Dierckx, G., Goegebeur, Y., and Matthys, G. (1999), "Tail index estimation and an exponential regression model," *Extremes*, 2, 177-200.

[4] Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307-327.

[5] Danielsson J., and de Vries, C.G. (1997), "Tail index and quantile estimation with very high frequency data," *Journal of Empirical Finance*, 4, 241-257.

[6] Danielsson, J., de Haan, L., Peng, L., and de Vries, C.G. (2001), "Using a bootstrap method to choose the sample fraction in tail index estimation," *Journal of Multivariate Analysis*, 76, 226-248.

[7] Davidson, R., and Flachaire, E. (2007), "Asymptotic and Bootstrap Inference for Inequality and Poverty Measures," *Journal of Econometrics*, 141(1), 141-166.

[8] de Haan, L., Jansen, D.W., Koedijk, K.G., and de Vries, C.G. (1994) Safety first portfolio selection, extreme value theory and long run asset risks. In: Galambos, J. (ed.), Proceedings from a Conference on Extreme Value Theory and Applications, Kluwer Press, 471-487.

[9] de Haan, L. and Stadtmüller, U. (1996),"Generalized regular variation of second order," *Journal of the Australian Mathematematical Society (Series A)* , 61, 381-385.

[10] Drees, H. (2003), "Extreme Quantile Estimation for Dependent Data with Applications to Finance," *Bernoulli*, 9, 617-657.

[11] Dumouchel, W.H. (1983), "Estimating the stable index α in Order to Measure Tail Thickness: A critique," *Annals of Statistics*, 11, 1019-1031.

[12] Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997), *Modelling Extremal Events*, Berlin (Springer-Verlag).

[13] Engle, R.F. (1982), "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of UK Inflation," *Econometrica*, 50, 9871008.

[14] Feller, W. (1971), *An introduction to Probability Theory and its Applications*, vol.2, New York-Wiley.

[15] Goldie, C.M., and Smith, R.H. (1987), "Slow variation with remainder: theory and applications," *Quarterly Journal of Mathematics* (Oxford 2nd series), 38, 45-71.

[16] Gomes, I., L. de Haan, and Peng, L. (2002), "Semi-parametric estimation of the second order parameter in statistics of extremes," *Extremes*, 5(4), 387-414.

[17] Gomes, I., and Martins, M.J. (2002), "Semi-parametric estimation of the second order parameter in statistics of extremes," *Extremes*, 5(4), 387-414.

[18] Haeusler E, and Teugels, J. (1985), "On asymptotic normality of Hill's estimator for the exponent of regular variation," *Annals of Statistics*, 13, 743-756.

[19] Hall, P. (1982), "On some simple estimates of an exponent of regular variation," *Journal of the Royal Statistical Society (Series B)*, 42, 37-42.

[20] Hartmann, P., Straetmans, S., and de Vries, C.G. (2003), *A global perspective on extreme currency linkages*. In: Hunter, W.C., Kaufman, G.G., Pomerleano, M. (Eds.), Asset Price Bubbles: The Implications for Monetary, Regulatory and International Policies. MIT Press, Cambridge (MA), pp. 361-382.

[21] Hartmann, P., Straetmans, S., and de Vries, C.G. (2004), "Asset market linkages in crisis periods," *Review of Economics and Statistics*, 86(1), 313-326.

[22] Hill, B.M. (1975), "A simple general approach to inference about the tail of a distribution," *The Annals of Statistics*, 3, 1163-1173.

[23] Hols, M.C.A.B., and de Vries, C.G. (1991), "The limiting distribution of extremal exchange rate returns," *Journal of Applied Econometrics*, 6, 287-302.

[24] Hsing, T. (1991), "Extremal Index Estimation for a Weakly Dependent Stationary Sequence," *The Annals of Statistics*, 21(4), 1547-1569.

[25] Ibragimov, I.A., and Linnik, Y.V. (1971), *Independent and stationary sequences of random variables*, in J.F.C. Kingman ed., Groningen (Wolters-Noordhoff).

[26] Jansen, D.W., and de Vries, C.G. (1991), "On the frequency of large stock returns: putting booms and busts into perspective," *Review of Economics and Statistics*, 73,19-24.

[27] Jondeau, E., and Rockinger, M. (2003), "Testing for differences in the tails of stock-market returns," *Journal of Empirical Finance*, 10, 559-581.

[28] Koedijk, K.G., Schafgans, M.M.A., and de Vries, C.G. (1990), "The tail index of exchange rate returns," *Journal of International Economics*, 29, 93-108.

[29] Koedijk, K.G., Stork, P.A., and de Vries, C.G. (1992), "Foreign exchange rate regime differences viewed from the tails," *Journal of International Money and Finance*, 11, 462-473.

[30] Longin, F.M. (1996), "The asymptotic distribution of extreme stock market returns," *Journal of Business*, 69, 383-408.

[31] Lux, T. (1996), "The stable Paretian Hypothesis and the frequency of large returns: an examination of major German stocks," *Applied Financial Economics*, 6, 463-475.

[32] Mandelbrot, B. (1963), "The variation of certain speculative prices," *Journal of Business*, 36, 394-419.

[33] Mikosch, T., and Starica, C. (1998), "Limit Theory for the Sample Autocorrelations and Extremes of a GARCH(1,1) process," working paper.

[34] Pagan, A.R., and Schwert, G.W. (1990), "Testing for Covariance Stationarity in Stock Market Data," *Economics Letters*, 33, 165-170.

[35] Phillips, P.C.B., and Loretan, M. (1990), "Testing Covariance Stationarity under Moment Condition Failure with an Application to Common Stock Returns", Cowles Foundation Discussion Paper nr. 947.

[36] Quandt, R. (1960), "Test of the Hypothesis that a Linear Regression obeys Two Separate Regimes," *Journal of the American Statistical Association*, 55, 324-330.

[37] Quintos, C., Fan, Z., and Phillips, P.C.B. (2001), "Structural Change Tests in Tail Behavior and the Asian Crisis," *Review of Economic Studies*, 68, 633-663.

[38] Samorodnitsky, G. and Taqqu, M. (1994), *Stable Non-Gaussian Random Processes*, New York (Chapmann and Hall).

[39] Starica, C. (2003), "Is Garch(1,1) as Good a Model as the Accolades of the Nobel Prize Would Imply?" mimeo.

Table 1: Tail expansion parameters

	a	b	β	$\rho=-\frac{\beta}{\alpha}$
Stable	$\pi^{-1}\Gamma(\alpha)\sin\frac{\alpha\pi}{2}$	$-\frac{\Gamma(2\alpha)\sin\alpha\pi}{2\Gamma(\alpha)\sin\frac{\alpha\pi}{2}}>0$	α	-1
Student	$\frac{\Gamma\left(\frac{\alpha+1}{2}\right)}{\Gamma\left(\frac{\alpha}{2}\right)\sqrt{\pi\alpha}}\alpha^{\frac{\alpha-1}{2}}$	$-\frac{\alpha^2(\alpha+1)}{2(\alpha+2)}<0$	2	$-\frac{2}{\alpha}$
Frechet	1	$-1/2<0$	α	-1
Burr	1	$-\rho^{-1}<0$	nonrestricted	

Table 2: Hill statistic: small sample bias and estimation risk models

	$\widehat{\alpha}$ (*s.e.*)					
	$m = m^*$			$m = 0.1n$		
	$n = 500$	$n = 2,000$	$n = 20,000$	$n = 500$	$n = 2,000$	$n = 20,000$
	Panel A: i.i.d. models					
Stable(α)						
1.2	1.23 (0.16)	1.24 (0.10)	1.23 (0.05)	1.25 (0.18)	1.23 (0.09)	1.22 (0.03)
1.5	1.79 (0.44)	1.66 (0.25)	1.57 (0.10)	1.77 (0.27)	1.74 (0.13)	1.73 (0.04)
Student(α)						
2	1.79 (0.28)	1.87 (0.19)	1.93 (0.09)	1.72 (0.23)	1.70 (0.11)	1.69 (0.03)
4	3.16 (0.74)	3.40 (0.56)	3.64 (0.34)	2.43 (0.31)	2.41 (0.15)	2.40 (0.05)
Frechet(α)						
2	1.88 (0.16)	1.92 (0.10)	1.96 (0.05)	1.98 (0.28)	1.96 (0.14)	1.95 (0.04)
4	3.75 (0.32)	3.84 (0.21)	3.93 (0.10)	3.97 (0.57)	3.91 (0.28)	3.90 (0.09)
Burr($\alpha, -\rho$)						
$(2, -0.5)$	1.68 (0.28)	1.77 (0.21)	1.87 (0.12)	1.59 (0.21)	1.57 (0.10)	1.57 (0.03)
$(2, -5)$	1.97 (0.11)	1.98 (0.06)	1.99 (0.02)	2.04 (0.29)	2.01 (0.14)	2.00 (0.04)
$(4, -0.5)$	3.37 (0.56)	3.56 (0.42)	3.73 (0.25)	3.18 (0.43)	3.14 (0.21)	3.13 (0.06)
$(4, -5)$	3.93 (0.22)	3.96 (0.12)	3.99 (0.04)	4.08 (0.59)	4.02 (0.29)	4 (0.09)
	Panel B: Models with temporal dependence in the first or second moment					
AR(α, θ)						
$(1.5, 0.2)$	1.82 (0.48)	1.68 (0.27)	1.57 (0.12)	1.78 (0.29)	1.74 (0.14)	1.74 (0.04)
$(1.5, 0.4)$	1.87 (0.55)	1.70 (0.32)	1.58 (0.14)	1.80 (0.34)	1.75 (0.17)	1.74 (0.05)
SVSTU(α, θ)						
$(4, 0.85)$	3.21 (0.76)	3.42 (0.57)	3.64 (0.34)	2.45 (0.33)	2.41 (0.16)	2.40 (0.05)
$(4, 0.95)$	3.27 (0.77)	3.43 (0.56)	3.64 (0.34)	2.52 (0.37)	2.43 (0.19)	2.40 (0.06)
GARCH(α, θ)						
$(4, 0.85)$	4.38 (1.77)	3.75 (0.82)	3.51 (0.26)	2.82 (0.45)	2.66 (0.23)	2.60 (0.07)
$(4, 0.95)$	4.74(1.99)	3.88 (0.92)	3.50 (0.32)	2.81 (0.50)	2.62 (0.27)	2.55 (0.09)

Note: Simulated average values and standard deviations are reported for the Hill statistic (20,000 replications) and for different sample sizes. The Hill statistic is conditioned on both a fixed fraction of extremes and the fraction that minimizes the asymptotic mean squared error. Parameters α and $\rho = -\beta/\alpha$ refer to the tail index and the ratio of the second order parameter to the tail index, respectively. The first order serial correlation of an AR(1) or the volatility persistence parameter in GARCH(1,1) models or stochastic volatility models with student-t innovations (SVSTU) are denoted by θ.

Table 3: Small sample critical values for recursive, rolling and sequential test: i.i.d. draws models

DGP	T=500			T=2,000		
	0.90	0.95	0.99	0.90	0.95	0.99
			Panel A: Recursive test			
Stable(α)						
1.2	1.97 (0.04)	2.78 (0.08)	5.22 (0.20)	2.00 (0.02)	2.67 (0.03)	4.64 (0.19)
1.5	5.12 (0.13)	8.39 (0.19)	20.37 (1.27)	3.41 (0.10)	4.97 (0.20)	9.61 (0.63)
Student(α)						
2	1.99 (0.05)	2.85 (0.06)	5.80 (0.26)	1.84 (0.02)	2.43 (0.04)	4.24 (0.15)
4	2.42 (0.08)	3.87 (0.21)	9.20 (0.81)	2.18 (0.04)	3.17 (0.08)	6.33 (0.34)
Frechet(α)	1.78 (0.03)	2.28 (0.05)	3.72 (0.14)	1.81 (0.02)	2.25 (0.03)	3.40 (0.09)
Burr(α, ρ)						
$(2, -0.2)$	2.48 (0.08)	3.95 (0.17)	9.71 (0.80)	2.47 (0.07)	3.77 (0.14)	8.43 (0.45)
$(2, -5)$	1.53 (0.02)	1.94 (0.03)	3.07 (0.11)	1.55 (0.03)	1.91 (0.04)	2.78 (0.05)
			Panel B: Rolling test ($\gamma = 0.2$)			
Stable(α)						
1.2	2.40 (0.07)	3.33 (0.08)	5.98 (0.22)	2.33 (0.05)	3.00 (0.10)	4.82 (0.18)
1.5	14.20 (0.54)	22.44 (1.18)	54.82 (4.69)	6.12 (0.13)	8.33 (0.26)	14.84 (0.79)
Student(α)						
2	2.87 (0.04)	4.10 (0.11)	7.97 (0.44)	2.15 (0.05)	2.87 (0.08)	4.84 (0.22)
4	4.81 (0.19)	7.46 (0.31)	17.66 (1.10)	3.06 (0.07)	4.38 (0.15)	8.40 (0.34)
Frechet(α)	1.64 (0.03)	2.17 (0.03)	3.60 (0.15)	1.63 (0.02)	2 (0.01)	3.04 (0.1)
Burr(α, ρ)						
$(2, -0.2)$	2.69 (0.08)	4.29 (0.17)	9.50 (0.57)	2.96 (0.09)	4.41 (0.13)	8.94 (0.61)
$(2, -5)$	1.66 (0.03)	2.09 (0.04)	3.25 (0.09)	1.52 (0.02)	1.82 (0.04)	2.57 (0.10)
			Panel C: sequential test			
Stable(α)						
1.2	21.67 (0.53)	31.73 (0.86)	59.01 (1.89)	16.21 (0.45)	22.54 (0.96)	40.38 (1.98)
1.5	24.33 (0.73)	39.03 (1.53)	87.89 (3.10)	16.51 (0.40)	24.13 (1.12)	48.81 (2.29)
Student(α)						
2	21.49 (0.34)	31.54 (1.04)	60.26 (3.62)	17.86 (0.43)	25.18 (0.80)	45.70 (1.22)
4	25.05 (0.47)	38.41 (0.77)	77.96 (3.55)	19.04 (0.67)	28.16 (1.06)	53.39 (2.55)
Frechet(α)	18.43 (0.33)	26 (0.60)	45.18 (1.21)	16. 76 (0.38)	23.09 (0.47)	39.16 (1.46)
Burr(α, ρ)						
$(2, -0.2)$	25.04 (0.89)	37.04 (1.95)	76.74 (4.49)	19.44 (0.54)	28.27 (0.81)	54.67 (3.98)
$(2, -5)$	19.97 (0.30)	27.75 (0.63)	49.44 (1.78)	19.75 (0.46)	26.20 (0.44)	43.10 (1.49)

Note: Critical values are reported for varying sample sizes (T), and different levels of significance. Critical values are averaged over 10 estimates of the appropriate quantiles. Each of the 10 estimates is simulated with 20,000 replications. Corresponding standard deviations for the critical values are reported between brackets (s.e.). The parameters α and $\rho = -\beta/\alpha$ refer to the tail index and the ratio of the second order parameter to the tail index, respectively.

Table 4: Small sample critical values for dependent models

DGP	$T = 500$			$T = 2,000$		
	0.90	0.95	0.99	0.90	0.95	0.99
				Panel A: Recursive test		
ARSTA(α, θ)						
$(1.2, 0.2)$	2.65 (0.07)	4.04 (0.13)	8.73 (0.71)	2.65 (0.05)	3.74 (0.07)	6.97 (0.27)
$(1.2, 0.4)$	4.01 (0.10)	6.43 (0.23)	15.05 (1.02)	4.01 (0.08)	5.97 (0.13)	11.79 (0.64)
SVSTU(α, θ)						
$(2, 0.85)$	2.25 (0.04)	3.27 (0.08)	6.67 (0.36)	1.92 (0.04)	2.56 (0.05)	4.59 (0.18)
$(2, 0.95)$	2.56 (0.06)	3.82 (0.12)	7.96 (0.47)	2.05 (0.05)	2.78 (0.06)	4.95 (0.23)
GARCH(α, θ)						
$(4, 0.85)$	3.41 (0.15)	6.08 (0.31)	20.46 (1.68)	2.63 (0.03)	3.42 (0.06)	7.25 (0.71)
$(4, 0.95)$	4.14 (0.14)	7.67 (0.30)	26.20 (2.50)	3.33 (0.07)	5.05 (0.21)	15.30 (1.63)
				Panel B: Rolling test		
ARSTA(α, θ)						
$(1.2, 0.2)$	3.16 (0.05)	4.40 (0.11)	8.26 (0.49)	3.12 (0.08)	4.07 (0.08)	6.54 (0.24)
$(1.2, 0.4)$	4.64 (0.07)	6.63 (0.15)	12.70 (0.46)	4.75 (0.08)	6.26 (0.09)	10.53 (0.34)
SVSTU(α, θ)						
$(2, 0.85)$	3.24 (0.10)	4.54 (0.20)	8.97 (0.42)	2.27 (0.04)	3.04 (0.05)	5.05 (0.24)
$(2, 0.95)$	3.73 (0.08)	5.22 (0.13)	9.87 (0.43)	2.46 (0.05)	3.25 (0.09)	5.48 (0.16)
GARCH(α, θ)						
$(4, 0.85)$	4.86 (0.08)	8.31 (0.26)	25.80 (1.84)	2.05 (0.03)	2.71 (0.06)	5.66 (0.26)
$(4, 0.95)$	5.75 (0.10)	9.81 (0.34)	29.06 (2.28)	2.89 (0.08)	4.31 (0.14)	10.31 (0.61)
				Panel C: sequential test		
ARSTA(α, θ)						
$(1.2, 0.2)$	26.86 (0.78)	40.68 (1.30)	85.68 (2.97)	21.41 (0.42)	31.06 (0.99)	60.6 (3.97)
$(1.2, 0.4)$	35.98 (0.64)	56.60 (2.13)	133.09 (5.63)	30.33 (0.45)	46.62 (0.77)	100.18 (4.54)
SVSTU(α, θ)						
SV$(2, 0.85)$	21.24 (0.68)	31.60 (0.95)	61.84 (2.80)	17.84 (0.33)	25.34 (0.63)	45.50 (2.34)
SV$(2, 0.95)$	21.25 (0.59)	31.54 (0.95)	60.88 (2.34)	17.72 (0.33)	25.15 (0.58)	45.62 (2.17)
GARCH(α, θ)						
$(4, 0.85)$	38.55 (1.11)	59.42 (2.31)	123.01 (3.55)	38.11 (0.88)	57.20 (1.01)	117.21 (4.69)
$(4, 0.95)$	36.76 (0.97)	57.25 (1.92)	119.74 (6.98)	214.97 (0.90)	264.68 (1.05)	396.78 (6.50)

Note: Critical values are reported for the recursive, rolling and sequential test and for varying sample size (T=500, 2000), and levels of significance (90 percent, 95 percent, 99 percent). The considered DGP's exhibit temporal dependence in the first moment (serial correlation) or the second moment (volatility persistence). The optimal number of highest order statistics m is determined analytically for all unconditional models by minimizing the Asymptotic Mean Squared Error (AMSE). For Garch(1,1) models we set m by applying the Beirlant et al. (1999) algorithm. Critical values are averaged over 10 estimates of the appropriate quantiles, where each of these quantile estimates is in turn based on 20,000 replications. Corresponding standard deviations for the critical values are reported between brackets (s.e.). The parameters α and $\rho = -\beta/\alpha$ refer to the tail index and the ratio of the second order parameter to the tail index, respectively. First order serial correlation of an autoregressive process with stable innovations (ARSTA), the volatility persistence parameter in GARCH(1,1) models or stochastic volatility models with student-t innovations (SVSTU) is always denoted by θ.

Table 5: Size-corrected small sample power of recursive test

DGP	$T = 500$			$T = 2,000$		
$(\alpha_1; \alpha_2)$	$r = 0.25$	$r = 0.5$	$r = 0.75$	$r = 0.25$	$r = 0.50$	$r = 0.75$
Stable						
$(1.5, 1.2)$	22	32	25	53	71	55
$(1.8, 1.2)$	50	73	57	98	100	99
$(1.2, 1.5)$	1.18	1.36	1.5	1.96	2.66	1.10
$(1.2, 1.8)$	3.4	3.56	1.48	4.2	2.5	1
Student						
$(4, 2)$	21	32	24	49	73	62
$(2, 4)$	0.5	0.7	2	2.54	0.94	1.18
Frechet						
$(4, 2)$	97	99	98	100	100	100
$(2, 4)$	24	8	1	93	30	1.6
Burr ($\rho = -0.2$)						
(4,2)	31	37	26	52	66	53
(2,4)	1.16	0.5	1.3	0.3	0.22	1.76
Burr ($\rho = -5$)						
(4,2)	99	100	99	100	100	100
(2,4)	98	99	65	100	100	100
ARSTA($\theta = 0.2$)						
$(1.5, 1.2)$	19	24	18	41	58	42
$(1.2, 1.5)$	0.7	0.2	1.2	0.38	0.26	1.00
SVSTU($\theta = 0.95$)						
$(4, 2)$	20	31	23	49	71	59
(2,4)	0.56	0.70	1.80	3.16	1.84	1.16
ARCH						
(4,2)	6.2	16.74	22.18	17.42	31.54	22.56
(2,4)	0.34	2.86	2.7	0.06	0.36	0.94

Note: The power is reported for different sample sizes (T=500, 2000), different locations of the (true) breakpoints (r=0.25, 0.50, 0.75) and different jump scenarios (α_1, α_2) for the tail index. The power is size-corrected using small sample critical values (see Tables 2-3) and is calculated as the rejection frequency under the null hypothesis of parameter constancy over 20,000 Monte Carlo replications. The optimal number of highest order statistics m is determined by applying the Beirlant et al. (1999) algorithm. The parameters α and $\rho = -\beta/\alpha$ refer to the tail index and the ratio of the second order parameter to the tail index, respectively. First order serial correlation of an autoregressive process with stable innovations (ARSTA), the volatility persistence parameter in GARCH(1,1) models and in stochastic volatility models with student-t innovations (SVSTU) is always denoted by θ.

Table 6: Size-corrected small sample power for rolling test test

DGP	T=500			T=2,000		
$(\alpha_1;\alpha_2)$	r=0.25	r=0.5	r=0.75	r=0.25	r=0.50	r=0.75
Stable						
$(1.5;1.2)$	7	8	5	32	38	22
$(1.8;1.2)$	21	22	8	94	94	58
$(1.2;1.5)$	5	9	7	22	37	30
$(1.2;1.8)$	7	21	22	57	93	94
Student						
$(4,2)$	6	6	4	27	32	15
$(2,4)$	4	6	5	15	31	27
Frechet						
$(4,2)$	95	97	83	100	100	100
$(2,4)$	84	97	95	100	100	100
Burr $(\rho=-0.2)$						
$(4,2)$	19	22	15	35	45	30
$(2,4)$	15	21	19	32	46	36
Burr $(\rho=-5)$						
$(4,2)$	100	100	95	100	100	100
$(2,4)$	96	100	100	100	100	100
ARSTA$(\theta=0.2)$						
(1.5,1.2)	6	7	5	25	27	15
(1.2,1.5)	4	7	5	16	28	24
SVSTU$(\theta=0.95)$						
(4,2)	12	22	16	43	66	55
(2,4)	0.2	0.26	0.66	1.28	0.74	0.8
ARCH						
(4,2)	2.7	3.62	4.16	7.94	15.86	22.16
(2,4)	2.82	1.8	1.2	5.4	3.44	2.06

Note: The power is reported for different sample sizes (T=500, 2000), different locations of the (true) breakpoints (r=0.25, 0.50, 0.75) and different jump scenarios (α_1, α_2) for the tail index. The power is size-corrected using small sample critical values (see Tables 2-3) and is calculated as the rejection frequency under the null hypothesis of parameter constancy over 20,000 Monte Carlo replications. The optimal number of highest order statistics m is determined by applying the Beirlant et al. (1999) algorithm. The parameters α and $\rho = -\beta/\alpha$ refer to the tail index and the ratio of the second order parameter to the tail index, respectively. First order serial correlation of an autoregressive process with stable innovations (ARSTA), the volatility persistence parameter in GARCH(1,1) models and in stochastic volatility models with student-t innovations (SVSTU) is always denoted by θ.

Table 7: Size-corrected small sample power for sequential test

DGP	T=500			T=2,000		
(α_1, α_2)	r=0.25	r=0.50	r=0.75	r=0.25	r=0.50	r=0.75
Stable						
$(1.5; 1.2)$	14	28	42	15	45	69
$(1.8; 1.2)$	21	61	86	52	99	100
$(1.2; 1.5)$	6	3	1	10	4	2
$(1.2; 1.8)$	6	1	0.3	7	48	74
Student						
$(4, 2)$	10	21	35	12	41	71
$(2, 4)$	3	1	0.4	0.6	0.1	0.6
Frechet						
$(4, 2)$	25	91	99	80	100	100
$(2, 4)$	2	18	34	83	99	100
Burr $(\rho = -0.2)$						
$(4, 2)$	17	41	57	14	45	68
$(2, 4)$	1	0.2	0.08	0.28	0.02	0.1
Burr $(\rho = -5)$						
$(4, 2)$	57	100	100	100	100	100
$(2, 4)$	27	87	99	100	100	100
ARSTA$(\theta = 0.2)$						
(1.5,1.2)	-	-	-	13	35	54
(1.2,1.5)	-	-	-	1.12	0.8	1.84
SVSTU$(\theta = 0.95)$						
(4,2)	10	22	36	12	43	71
(2,4)	1.82	1.1	1.98	0.34	0.20	0.64
ARCH						
$(4, 2)$	6.72	10	14.88	9.14	20.24	40.6
$(2, 4)$	1.30	0.9	1.74	0.32	0.08	0.54

Note: The power is reported for different sample sizes (T=500, 2000), different locations of the (true) breakpoints (r=0.25, 0.50, 0.75) and different jump scenarios (α_1, α_2) for the tail index. The power is size-corrected using small sample critical values (see Tables 2-3) and is calculated as the rejection frequency under the null hypothesis of parameter constancy over 20,000 Monte Carlo replications. The optimal number of highest order statistics m is determined by applying the Beirlant et al. (1999) algorithm. The parameters α and $\rho = -\beta/\alpha$ refer to the tail index and the ratio of the second order parameter to the tail index, respectively. First order serial correlation of an autoregressive process with stable innovations (ARSTA), the volatility persistence parameter in GARCH(1,1) models and in stochastic volatility models with student-t innovations (SVSTU) is always denoted by θ.

Table 8: Breakpoint estimates for recursive test

DGP	T=500			T=2,000		
$(\alpha_1; \alpha_2)$	breakpoints					
	r=0.25	r=0.50	r=0.75	r=0.25	r=0.50	r=0.75
Stable						
$(1.5, 1.2)$	0.42 (0.17)	0.53 (0.13)	0.64 (0.16)	0.33 (0.12)	0.50 (0.10)	0.66 (0.14)
$(1.8, 1.2)$	0.36 (0.14)	0.53 (0.09)	0.71 (0.10)	0.27 (0.06)	0.50 (0.04)	0.73 (0.06)
$(1.2, 1.5)$	0.62 (0.11)	0.55 (0.15)	0.48 (0.17)	0.48 (0.13)	0.48 (0.08)	0.48 (0.16)
$(1.2, 1.8)$	0.53 (0.11)	0.52 (0.08)	0.52 (0.20)	0.39 (0.08)	0.49 (0.06)	0.51 (0.15)
Student						
$(4, 2)$	0.40 (0.17)	0.53 (0.13)	0.67 (0.15)	0.33 (0.13)	0.51 (0.10)	0.70 (0.11)
$(2, 4)$	0.58 (0.20)	0.42 (0.22)	0.41 (0.17)	0.53 (0.11)	0.51 (0.12)	0.39 (0.15)
Frechet						
$(4, 2)$	0.26 (0.06)	0.49 (0.05)	0.71 (0.09)	0.25 (0.01)	0.49 (0.02)	0.73 (0.04)
$(2, 4)$	0.40 (0.09)	0.48 (0.08)	0.44 (0.13)	0.39 (0.07)	0.49 (0.08)	0.44 (0.11)
Burr($\rho = -0.2$)						
$(4, 2)$	0.38 (0.16)	0.51 (0.16)	0.61 (0.19)	0.30 (0.09)	0.50 (0.08)	0.70 (0.10)
$(2, 4)$	0.62 (0.16)	0.47 (0.22)	0.40 (0.19)	0.65 (0.16)	0.31 (0.19)	0.37 (0.14)
Burr($\rho = -5$)						
$(4, 2)$	0.26 (0.03)	0.50 (0.03)	0.73 (0.05)	0.25 (0.06)	0.48 (0.05)	0.72 (0.08)
$(2, 4)$	0.36 (0.08)	0.51 (0.04)	0.67 (0.09)	0.36 (0.05)	0.51 (0.03)	0.71 (0.06)
ARSTA($\theta = 0.2$)						
$(1.5, 1.2)$	0.45 (0.19)	0.54 (0.14)	0.66 (0.15)	0.35 (0.14)	0.50 (0.11)	0.65 (0.14)
$(1.2, 1.5)$	0.67 (0.15)	0.50 (0.23)	0.45 (0.15)	0.67 (0.12)	0.56 (0.23)	0.43 (0.16)
SVSTU($\theta = 0.95$)						
$(4, 2)$	0.40 (0.17)	0.52 (0.14)	0.64 (0.17)	0.35 (0.14)	0.51 (0.11)	0.56 (0.17)
$(2, 4)$	0.63 (0.16)	0.43 (0.21)	0.46 (0.19)	0.58 (0.14)	0.58 (0.13)	0.55 (0.17)
ARCH						
$(4, 2)$	0.41 (0.18)	0.52 (0.15)	0.65 (0.19)	0.37 (0.15)	0.60 (0.14)	0.75 (0.09)
(2,4)	0.35 (0.23)	0.71 (0.20)	0.67 (0.21)	0.71 (0.15)	0.82 (0.05)	0.82 (0.06)

Note: Estimated break dates are reported for different sample sizes (T=500, 2000), different locations of the (true) breakpoints (r=0.25, 0.50, 0.75) and different jump scenarios (α_1, α_2) for the tail index. "Candidate"-breakpoints are calculated over 20,000 Monte Carlo replications. Average break date estimates are obtained by averaging over the statistically significant "candidate"-breaks. using small sample critical values as cut-off point (cf. Tables 2-3). The optimal number of highest order statistics m is determined by applying the Beirlant et al. (1999) algorithm. The parameters α and $\rho = -\beta/\alpha$ refer to the tail index and the ratio of the second order parameter to the tail index, respectively. First order serial correlation of an autoregressive process with stable innovations (ARSTA), the volatility persistence parameter in GARCH(1,1) models and in stochastic volatility models with student-t innovations (SVSTU) is always denoted by θ.

Table 9: Breakpoint estimates for rolling test

DGP	T=500			T=2,000		
(α_1, α_2)	breakpoints					
	r=0.25	r=0.50	r=0.75	r=0.25	r=0.50	r=0.75
Stable						
$(1.5, 1.2)$	0.37 (0.22)	0.39 (0.14)	0.51 (0.18)	0.65 (0.10)	0.36 (0.11)	0.48 (0.18)
$(1.8, 1.2)$	0.26(0.10)	0.37(0.11)	0.48(0.18)	0.23 (0.03)	0.36 (0.10)	0.48 (0.18)
$(1.2, 1.5)$	0.68(0.17)	0.81(0.14)	0.86(0.20)	0.72 (0.17)	0.84 (0.11)	0.95 (0.09)
$(1.2, 1.8)$	0.72(0.17)	0.83(0.11)	0.95(0.08)	0.72 (0.17)	0.84 (0.10)	0.97 (0.02)
Student						
$(4, 2)$	0.37 (0.22)	0.39 (0.14)	0.49 (0.17)	0.26 (0.10)	0.37 (0.11)	0.49 (0.18)
$(2, 4)$	0.71 (0.18)	0.81 (0.14)	0.85 (0.20)	0.72 (0.17)	0.83 (0.11)	0.95 (0.08)
Frechet						
$(4, 2)$	0.23 (0.03)	0.35 (0.10)	0.48 (0.18)	0.23 (0.02)	0.35 (0.10)	0.48 (0.17)
$(2, 4)$	0.72 (0.18)	0.84 (0.10)	0.97 (0.03)	0.72 (0.18)	0.85 (0.10)	0.97 (0.02)
Burr($\rho = -0.2$)						
$(4, 2)$	0.28 (0.13)	0.37 (0.13)	0.48 (0.18)	0.26 (0.10)	0.37 (0.11)	0.48 (0.18)
$(2, 4)$	0.71 (0.18)	0.82 (0.12)	0.92 (0.12)	0.71 (0.17)	0.83 (0.11)	0.94 (0.10)
Burr($\rho = -5$)						
$(4, 2)$	0.23 (0.02)	0.35 (0.10)	0.50 (0.21)	0.23 (0.02)	0.35 (0.10)	0.51 (0.21)
$(2, 4)$	0.70 (0.20)	0.85 (0.10)	0.97 (0.02)	0.69 (0.21)	0.85 (0.10)	0.97 (0.02)
ARSTA($\theta = 0.2$)						
$(1.5, 1.2)$	0.36 (0.21)	0.39 (0.14)	0.53 (0.18)	0.26 (0.11)	0.37 (0.11)	0.48 (0.17)
$(1.2, 1.5)$	0.68 (0.17)	0.81 (0.14)	0.84 (0.21)	0.71 (0.18)	0.83 (0.11)	0.94 (0.10)
SVSTU($\theta = 0.95$)						
$(4, 2)$	0.33 (0.19)	0.40 (0.14)	0.49 (0.19)	0.26 (0.10)	0.37 (0.11)	0.48 (0.17)
$(2, 4)$	0.72 (0.18)	0.79 (0.15)	0.87 (0.18)	0.70 (0.17)	0.83 (0.11)	0.95 (0.08)
ARCH						
$(4, 2)$	0.49 (0.27)	0.49 (0.23)	0.54 (0.21)	0.31 (0.17)	0.39 (0.15)	0.49 (0.19)
(2,4)	0.70 (0.20)	0.70 (0.21)	0.62 (0.27)	0.71 (0.17)	0.78 (0.16)	0.88 (0.19)

Note: The power is reported for different sample sizes (T=500, 2000), different locations of the (true) breakpoints (r=0.25, 0.50, 0.75) and different jump scenarios (α_1, α_2) for the tail index. The power is size-corrected using small sample critical values (see Tables 2-3) and is calculated as the rejection frequency under the null hypothesis of parameter constancy over 20,000 Monte Carlo replications. The optimal number of highest order statistics m is determined by applying the Beirlant et al. (1999) algorithm. The parameters α and $\rho = -\beta/\alpha$ refer to the tail index and the ratio of the second order parameter to the tail index, respectively. First order serial correlation of an autoregressive process with stable innovations (ARSTA), the volatility persistence parameter in GARCH(1,1) models and in stochastic volatility models with student-t innovations (SVSTU) is always denoted by θ.

Table 10: Breakpoint estimates for sequential test

DGP	T=500			T=2,000		
(α_1, α_2)	breakpoints					
	r=0.25	r=0.50	r=0.75	r=0.25	r=0.50	r=0.75
Stable						
$(1.5, 1.2)$	0.81 (0.08)	0.79 (0.09)	0.81 (0.04)	0.76 (0.14)	0.70 (0.13)	0.80 (0.05)
$(1.8, 1.2)$	0.77 (0.13)	0.72 (0.13)	0.81 (0.04)	0.45 (0.24)	0.58 (0.12)	0.79 (0.04)
$(1.2, 1.5)$	0.83 (0.03)	0.83 (0.05)	0.84 (0.01)	0.83 (0.02)	0.82 (0.03)	0.82 (0.03)
$(1.2, 1.8)$	0.84 (0.01)	0.84 (0.01)	0.84 (0.01)	0.79 (0.05)	0.78 (0.06)	0.81 (0.03)
Student						
$(4, 2)$	0.80 (0.09)	0.78 (0.10)	0.81 (0.04)	0.78 (0.11)	0.71 (0.13)	0.80 (0.04)
$(2, 4)$	0.79 (0.12)	0.69 (0.24)	0.55 (0.18)	0.84 (0.02)	0.63 (0.28)	0.55 (0.13)
Frechet						
$(4, 2)$	0.55 (0.27)	0.59 (0.13)	0.79 (0.04)	0.30 (0.16)	0.51 (0.06)	0.76 (0.03)
$(2, 4)$	0.83 (0.03)	0.82 (0.03)	0.82 (0.03)	0.79 (0.06)	0.81 (0.04)	0.82 (0.03)
Burr$(\rho = -0.2)$						
$(4, 2)$	0.73 (0.17)	0.74 (0.12)	0.81 (0.05)	0.63 (0.22)	0.66 (0.15)	0.79 (0.06)
$(2, 4)$	0.82 (0.04)	0.82 (0.04)	0.69 (0.16)	0.67 (0.17)	0.70 (0.14)	0.75 (0.12)
Burr$(\rho = -5)$						
$(4, 2)$	0.51 (0.26)	0.60 (0.13)	0.79 (0.04)	0.29 (0.13)	0.51 (0.05)	0.77 (0.03)
$(2, 4)$	0.81 (0.04)	0.80 (0.05)	0.82 (0.03)	0.78 (0.08)	0.82 (0.04)	0.83 (0.02)
ARSTA$(\theta = 0.2)$						
$(1.5, 1.2)$	0.80 (0.08)	0.79 (0.08)	0.81 (0.04)	0.76 (0.12)	0.72 (0.13)	0.80 (0.05)
$(1.2, 1.5)$	0.83 (0.04)	0.84 (0.01)	0.78 (0.10)	0.83 (0.02)	0.83 (0.02)	0.82 (0.04)
SVSTU$(\theta = 0.95)$						
$(4, 2)$	0.77 (0.12)	0.75 (0.11)	0.80 (0.06)	0.78 (0.10)	0.70 (0.13)	0.79 (0.06)
$(2, 4)$	0.81 (0.04)	0.82 (0.04)	0.75 (0.11)	0.83 (0.02)	0.84 (0.01)	0.82 (0.03)
ARCH						
$(4, 2)$	0.71 (0.19)	0.73 (0.14)	0.78 (0.10)	0.77 (0.13)	0.73 (0.12)	0.81 (0.05)
(2,4)	0.77 (0.17)	0.64 (0.24)	0.63 (0.18)	0.84 (0.007)	0.54 (0.26)	0.61 (0.12)

Note: The power is reported for different sample sizes (T=500, 2000), different locations of the (true) breakpoints (r=0.25, 0.50, 0.75) and different jump scenarios (α_1, α_2) for the tail index. The power is size-corrected using small sample critical values (see Tables 2-3) and is calculated as the rejection frequency under the null hypothesis of parameter constancy over 20,000 Monte Carlo replications. The optimal number of highest order statistics m is determined by applying the Beirlant et al. (1999) algorithm. The parameters α and $\rho = -\beta/\alpha$ refer to the tail index and the ratio of the second order parameter to the tail index, respectively. First order serial correlation of an autoregressive process with stable innovations (ARSTA), the volatility persistence parameter in GARCH(1,1) models and in stochastic volatility models with student-t innovations (SVSTU) is always denoted by θ.

Table 11: Structural change tests (01/01/1988-13/08/2007)

Country	m^*	Recursive test		Bootstrapped critical values		
		Q_F	Q_B	$CV_B(90\%)$	$CV_B(95\%)$	$CV_B(99\%)$
Argentina	122	0.324	0.807	2.170	2.649	3.914
Australia	116	0.199	0.941	2.307	2.866	4.420
Austria	72	0.381	0.649	2.366	2.953	4.569
Britain	57	0.481	0.715	2.449	3.025	4.294
Canada	129	2.581	0.541	2.656	3.242	4.507
Chile	98	0.067	0.898	2.012	2.499	3.904
France	125	0.432	0.362	1.929	2.324	3.250
Germany	180	0.936	0.793	2.323	2.755	3.725
Hong Kong	165	1.583	2.911**	1.951	2.374	3.433
(break date)		(**3-6-98**)				
India	56	1.845	0.229	2.033	2.581	3.975
Indonesia	85	1.484	1.211	2.025	2.631	4.393
Italy	194	0.939	1.591	2.613	3.130	4.068
Japan	74	0.412	0.619	2.122	2.925	5.512
Korea	126	1.853	0.726	1.960	2.362	3.445
Malaysia	127	0.642	1.284	2.971	4.196	7.474
Mexico	92	0.452	0.515	2.440	3.156	5.105
Philippines	66	1.191	1.009	2.834	3.486	4.867
Spain	113	1.864	1.731	2.306	2.774	3.954
Taiwan	50	1.194	6.211***	1.989	2.412	3.424
(breakdate)		(**2-6-1991**)				
Thailand	176	1.564	0.943	5.134	6.066	8.005
US	151	0.714	0.564	3.146	3.786	5.100

Note: The forward and backward version of the recursive test are denoted by Q_F and Q_B, respectively. The optimal number of highest order statistics m is determined by applying the Beirlant et al. (1999) algorithm. Bootstrapped critical values are reported for different levels of significance. Critical values are calculated as averages over 10 estimates of the appropriate quantiles. Each quantile is calculated on a bootstrapped sample. The number of bootstrap replications is set equal to 20,000. Statistically significant rejections of the null hypothesis of tail index constancy at the 10, 5 and 1 percent significance level are denoted by $*$, $**$ and $***$, respectively. The break dates (dd/mm/yy) of corresponding significant breaks are reported in bold.