DEVELOPED AND EMERGING EQUITY MARKET TAIL RISK: IS IT CONSTANT?

STEFAN STRAETMANS[A,*] AND BERTRAND CANDELON[B]

ABSTRACT. The tail of equity returns is typically governed by a power law (i.e. "fat tails"). However, the constancy of the so-called tail index α which dictates the tail decay has been hardly studied. Using the Hill estimator to estimate the tail index, we study the finite sample properties of endogenous stability tests for α. We show that the finite sample critical values strongly depend on the underlying distributional assumptions for the stock returns. We therefore recommend a bootstrap-based version of the stability test as an alternative to the test's asymptotic distribution. Upon applying this stability test to return tails of developed and emerging equity markets, the evidence for structural shifts is found to be rather weak. This is reconforting news for the proponents of Extreme value Theory (EVT). They typically assume stationary tail behavior when applying tail index and extreme quantile estimators to the downside risk of equity portfolios.

1. INTRODUCTION

Financial market turmoil like the 1997 Asian crisis, the LTCM debacle, the Mexican "Tequila" crisis, the Russian debt crisis or the recent subprime credit crunch has increased the awareness of both academics and practitioners on the importance of acurately 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. This constitutes one of the first studies to acknowledge that

---

Key words and phrases. Heavy tails; Tail Index; Extreme Value Analysis; Endogenous stability test;
Finite sample properties; Bootstrap
JEL classification: F31, G15, G19, C49.

[a]Maastricht University (School of Business and Economics), the Netherlands
[b]Maastricht University (School of Business and Economics), the Netherlands
*Corresponding author; Address: Maastricht University (School of Business and Economics), P.O. Box 616, 6200 MD Maastricht, the Netherlands; e-mail: s.straetmans@maastrichtuniversity.nl. 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 Université d'Orléans.

overnight financial market turbulence cannot well be decribed 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 nonconstant tail behavior, i.e., a time varying α.[2] Stochastic Volatility models such as the ARCH- and GARCH-type class reconcile a stationary unconditional df (constant α) with clusters of high and low volatility in the conditional df. However, 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) stays invariant over very long time periods; or, alternatively, whether there may be a trend-wise increase in α and accompanying long-run risk). Otherwise stated, can highly volatile periods like the Asian crisis or the Mexican Tequila crisis and periods of market quiescence be explained by the same stationary distributional model or do crisis periods induce structural breaks in tail behavior, i.e., shifts in α?

Testing for structural change in the tail behavior of the unconditional distribution is important from both a statistical and policy perspective. First, whether Extreme Value Theory (EVT) or the cited ARCH and GARCH models of the conditional df are applicable or not depends on the stationarity assumption for the unconditional tail. 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 to calculating the unconditional Value-at-Risk

[1]Jansen and de Vries (1991), Longin (1996), Lux (1996) and Hartmann et al. (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 et al. (1990, 1992), Hols and de Vries (1991) and Hartmann et al. (2003). Bond extremes have been rather neglected in the empirical literature; De Haan et al. (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.

(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.[3] To the best of our knowledge, Quintos et al. (2001) constitutes the only stability study on detecting single breakpoints in α that also proposes estimators for break *dates* in α.[4] Our study extends and refines the previous breakpoint analyses in several directions. First, we select the number of extreme returns to estimate α by minimizing its Asymptotic Mean Squared Error (AMSE) instead of conditioning on a fixed fraction of the total sample. The former approach constitutes common practice in EVT whereas taking a fixed percentage of extremes leads to a degenerate asymptotic limiting df for the α-estimator and accompanying stability tests. Second, our simulation study of the stability tests' finite sample properties is much more general than previous studies because we also use data generating processes (DGP's) that consider higher order tail behavior or empirical stylized facts like e.g. volatility clustering in returns. Last but not least, we apply stability tests to the tails of a large cross section of developed and emerging equity market index returns. Emerging stock markets may be seen as a "control" sample because breaks in the tail index α - if present - may be expected to occur more often in a context of high political risk, recurring exchange rate regime shifts, or underdeveloped governance and supervision.

Anticipating our results, we find that size, (size-corrected) power and the ability to detect breaks in finite samples vary considerably with the assumed DGP. That is the reason why we propose to bootstrap the

[3]The breakpoint literature includes Koedijk et al. (1990, 1992), Jansen and de Vries (1991), Pagan and Schwert (1990) and Straetmans et al. (2008). One can distinguish tests for structural change in α 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 on stock index tail asymmetries).

[4]Werner and Upper (2002), Galbraith and Zernov (2004) and Candelon and Straetmans (2006) already apply the Quintos et al. (2001) methodology to test for tail stability in bund Future returns, US stock market returns and Asian currency returns, respectively. However, they all use the Quintos et al. (2001) asymptotic critical values. We argue in this paper that these critical values do not take into account the bias in the Hill estimator and lead to overrejection of the null hypothesis of tail index constancy.

critical values in empirical applications for each data set separately. Moreover, the outcomes of our experiments on size-corrected power and the ability to detect breaks 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 our cross section of developed and emerging equity returns, we mainly detect breaks in the tail behavior of emerging currencies.

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

2. Testing structural change in 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 review of the QFP stability tests for the tail index α in section 2.2.

2.1. **Regular variation.**

We start from the empirical stylized fact that sharp fluctuations in financial market prices are poorly described by exponentially declining tails, i.e., financial returns exhibit fat tails, see e.g. Mandelbrot (1963) for early references or the more recent monograph by Embrechts et al. (1997). Without loss of generality (and for sake of notational convenience) we adopt the convention to take the negative of a return such that all presented estimation and testing procedures are expressed in terms of the right tail, i.e., the survivor function $P\{X \geq x\} := 1 - F(x)$. 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 Stadmüller (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 $P\{X \geq x\} \simeq ax^{-\alpha}[1 + b\ln x]$. If $b = 0$, the tail specializes to an exact Pareto.

The regular variation property implies that the upper extremal returns (appropriately scaled) 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. $E\left[X^r\right], r > \alpha$, are unbounded, signifying the "fat tail property". Models of the unconditional distribution like the Student-t, symmetric stable, Burr, and Frechet dfs as well as Stochastic Volatility models of the conditional distribution like the GARCH class all exhibit heavy tails.[5] As for the tail of the standard normal distribution, a popular tail approximation expresses the survivor function $1 - \Phi(\cdot)$ 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 with this type of tail decay are classified as "thin tailed" because the tail probabilities $1 - \Phi(x)$ decline much faster to zero as in (2.1); but these distributions possess all moments, and hence do not capture what is 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

[5]Note that 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, the 2nd order parameter can be freely chosen. The value of β is unknown for the GARCH class. All the mentioned distributional models together with some temporally dependent variants will be put at work in the Monte Carlo section.

sample size n. The main convergence in distribution result for $\widehat{\alpha}$ is summarized in the following theorem:

Theorem 1. *(Convergence in distribution) 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* $\sqrt{m}\,(\widehat{\alpha}-\alpha) \xrightarrow{d} N\left(0, \eta\alpha^2\right)$.

(B) If $m = cn^{2\beta/2\beta+\alpha}$ *then* $\sqrt{m}\,(\widehat{\alpha}-\alpha) \xrightarrow{d} N\left(\varphi\alpha, \eta\alpha^2\right)$ *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* $\varphi = sign(b)\,(2\beta/\alpha)^{-1/2}$.

Asymptotic normality for the Hill statistic was initially established by e.g. Hall (1982) and Haeusler and Teugels (1985) for the i.i.d. case ($\eta = 1$). Convergence in distribution has also been derived in the presence of temporal dependencies ($\eta \neq 1$).[6]

Loosely speaking, Theorem 1 implies that convergence requires m to rise with n at a "sufficiently slow" pace, i.e., $m, n \to \infty$ but $m/n \to 0$. Selecting a fixed fraction of extremes $\kappa = m/n$ (as under Dumouchel's (1983) rule), however, does not guarantee proper convergence because both (A) and (B) in Theorem 1 are violated. 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. Notice that 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). The AMSE minimization principle is applied in virtually all empirical EVT studies and we will therefore use this criterion in the rest of the paper. Notice, however, that Theorem 1 also makes clear that applying the AMSE criterion also 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.

2.2. **Structural change in tail behavior.** QFP propose a recursive, rolling and sequential procedure for detecting single unknown breaks in the Hill statistic $\widehat{\alpha}$. Let t denote the endpoint of a subsample of

[6]Quintos et al. (2001) derives convergence in distribution of the Hill statistic under stationary GARCH processes with conditionally normal innovations. 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.

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 $m_{w^*} = c\,(w^*)^{\frac{2\beta}{(2\beta+\alpha)}}$.[7] 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 $r = t/n$. 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 about the direction of a break, one is interested in testing the null

[7]Notice that the number of upper order extremes is increasing in the subsample size for the recursive estimator but is constant for the rolling estimator.

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] In line with Quandt's (1960) pioneering work on endogenous breakpoint determination in linear time series models, the candidate break date r is selected where the testing sequences (2.5), (2.6) and (2.7) reach their supremum which renders the most likely time point for the constancy hypothesis to be violated.

QPF also establish asymptotic distribution theory for the three stability tests under the two cases (A) and (B) from Theorem 1. We are interested in the asymptotic behavior under outcome (B) as this corresponds with the standard practice of minimizing the AMSE of $\widehat{\alpha}$ with respect to m. Upon denoting Q, Q^* and $Q^\#$ as the recursive, rolling and sequential test statistics, the corresponding limiting distributions boil down to:

$$\begin{aligned}
Q &= \sup \eta_Y^{-1} Y_n^2(r) \to^d \sup \left[\bar{W}(r) + \varphi r^{1/2}\left(1 - r^{1/2}\right)\right]^2, \\
Q^* &= \sup \eta_V^{-1} V_n^2(r) \to^d \sup \left[\bar{W}(r,\gamma_0) + \varphi \gamma_0^{1/2}\left(1 - \gamma_0^{1/2}\right)\right]^2, \\
Q^\# &= \sup \eta_Z^{-1} Z_n^2(r) \to^d \sup \left[\bar{W}^\#(r) + \varphi r^{1/2}\left(1 - \left(r/(1-r)\right)^{1/2}\right)\right]^2, \\
r &\in [0.15, 0.85], \qquad \varphi = sign(b)\,(2\beta/\alpha)^{-1/2},
\end{aligned}$$

with $\bar{W}(r) = W(r) - rW(1)$, $\bar{W}(r,\gamma_0) = W(r,\gamma_0) - (r-s)\,W(1,1)$, $\bar{W}^\#(r) = W(r) - (r/(1-r))\,W(1-r)$ and where $W(r)$ is a standard Wiener process. The scaling factors η_Y^{-1}, η_V^{-1} and η_Z^{-1} correct for possible nonlinear dependence in the data; but estimators for these correction factors are only known for the GARCH class. We will implement these correction factors in the empirical application.

Nonsurprisingly, the parameter φ that governed the Hill statistic's asymptotic bias reappears in the limiting distributions of the stability tests. The bias term suggests that "globally valid" asymptotic critical values that can be applied to all financial return tails do not apply. In the next section we investigate the impact of this bias term on the critical values, power and ability to date breaks in finite samples. Whereas

[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 efficieny one might stop calculating the tests beyond the upper bound of $(1-\varepsilon)\,n < n$. Conform with Andrews (1993), we set $\varepsilon = 0.15$.

the asymptotic distributions across different DGP's can only differ due to differences in $\varphi = sign(b)\left(2\beta/\alpha\right)^{-1/2}$, the small sample outcomes for critical values, power and break datation are also potentially depending on the sample size n, the optimal nuisance parameter m^* and the scaling factor a.

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 (2.1). The number of upper order extremes for the Hill statistic is chosen such as to minimize the Asymptotic Mean Squared Error of the Hill estimator. We calculate (size-corrected) small sample power against a variety of realistic break scenarios 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 3.1. The analytic derivation of the nuisance parameters for these DGP's is discussed in 3.2. The outcomes of our Monte Carlo experiments (small sample size, power, location of break dates) are summarized in 3.3.

3.1. **Data generating processes.** We choose regularly varying data generating processes such as to get a sufficiently large heterogeneity in parameter values (a, b, α, β). Monte Carlo simulations are reported for the symmetric stable df, Student-t, Fréchet (Type-II extreme value), i.e. $P\{X > x\} = \exp\left(-x^{-\alpha}\right)$, Burr, 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 Stochastic Volatility model. Thus, we distinguish between i.i.d. draws and dependent draws.

For sake of comparison with the QFP Monte Carlo results we generate symmetric stable draws using the algorithm proposed by Samorodnitsky and Taqqu (1994):

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

where $0 < \alpha < 2$ represents the tail index. For $\alpha = 2$, the stable df coincides with the normal df which implies that α is not interpretable as the tail index in this case. The parameter γ is drawn uniformly on

$[-\pi/2\ \pi/2]$ whereas W is exponentially distributed with mean 1.[9] The tail index of the Student-t, Frechet and Burr is varied between 2 and 4 which is the range one typically observes for stock market tails, see also Jansen and de Vries (1991) or Hartmann et al. (2004) for previous studies on stock market tail behavior.

We also allow for models that exhibit linear and nonlinear dependence because the i.i.d. assumption is too restrictive for financial return data. First order serial correlation (linear dependence) is generated by means of an AR(1) process with first order autocorrelation $\theta = 0.1$ and with symmetric stable innovations.[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 model from Danielsson et al. (2001):

$$\begin{aligned} Y_t &= U_t \sqrt{\frac{v}{\chi^2(v)}} H_t, \\ H_t &= \beta Q_t + \theta H_{t-1}, \end{aligned} \tag{3.2}$$

[9]Mandelbrot (1963) was arguably the first proponent of using this distribution class for modelling financial returns. He mainly based his choice on the property that the stable distribution is preserved under addition *over the full distributional support* (up to a location and scale adjustment). However, Feller (1971, ch. VIII.8) has proven that the additivity property only holds for the *tail area* of regularly varying dfs (the class of stable dfs constituting a notable exception). This suggests that the stable model might be overly restrictive. Also, one can argue that less heavy tailed models are to be preferred because the class of stable dfs fails to have a finite 2nd moment. Finally, notice that 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 and economically 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 (but economically minor) 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 $X_t \overset{d}{=} \left(1 + \theta^{\phi} + \theta^{2\phi} + \cdots\right) u_t = \left(\frac{1}{1-\theta^{\phi}}\right)^{\frac{1}{\phi}} u_t$. Thus the AR(1) dependent stable draws and the i.i.d. stable innovations exhibit the same distribution upon some scaling constant.

with $P\{U_t = -1\} = P\{U_t = 1\} = 0.5$, $\beta = 0.1$, $\theta = 0.9$ and where Q is standard normally distributed. 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). Finally, we also simulate from 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 equals 4, see Appendix A for details on the relation between the GARCH parameters and the tail index.

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 centre the smaller will be the variance of the estimator but the more bias will be introduced. Goldie and Smith (1987) therefore proposed to select m in (2.2) such as to minimize 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 reponsible for the bias in the Hill statistic, i.e., if either b or β equal 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:

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

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.[12]

Table 1 reports the parameter vectors (a, b, α, β) for the tail expansions of all considered DGP's.

[12]It can be easily shown that for $m = m^*$ the bias part and the variance part of the AMSE vanish at the same rate ($n^{-2\beta/2(\beta+\alpha)}$). If either the bias or the variance part is converging at a higher rate, then the other part converges more slowly than $n^{-2\beta/(2\beta+\alpha)}$, and hence the whole AMSE converges more slowly.

[Insert Table 1]

The optimal nuisance parameter m^* to be used in the Monte Carlo simulation is also included. Further details on the tail expansion derivations are provided in Appendix A. A number of interesting observations can be made from this table. First, the table reveals that the condition $\alpha = \beta$ employed in a number of previous studies (Quintos et al. (2001), Hall (1982)) only holds for the Type II extreme value df (Frechet) and the symmetric stable class but not for the other considered models. 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 further be clarified when discussing the simulation results.

The Monte Carlo study for small sample critical values will be performed using the analytic expressions for m^* in Table 1.[13] However, for the Monte Carlo study of power and ability to date breaks, as well as for the empirical application, we shall minimize a sample equivalent of the Asymptotic Mean Squared Error (AMSE) along the lines of Beirlant et al. (1999) in order to select the optimal m^*.[14] These authors derived an exponential regression model for the log-spacings of upper order statistics from regularly varying tails:

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

with $1 \leq j \leq m$. Moreover, $\gamma = 1/\alpha$, $\rho = -\beta/\alpha$, $(f_1, f_2, \cdots, f_m)$ 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

[13]The tail expansion parameters of a GARCH(1,1) model are unknown and we therefore resort to the Beirlant et al. (1999) criterion in order to determine m.

[14]Subsample bootstrap algorithms (see, e.g., Danielsson et al., 2001) to select m^* by means of AMSE minimization constitute an alternative route; 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.

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., $m^* = \arg\min_m [AMSE(\widehat{\gamma})]$.

For sake of convenience, we make two additional assumptions when implementing the Beirlant et al. (1999) criterion. First, we impose the restriction $\alpha = \beta$ ($\rho = -1$) 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 is notoriously difficult (see, e.g., Gomes et al., 2002a, b) which makes it beneficial to impose the restriction. Moreover, simulations have shown that the Beirlant criterion still performs well under the restriction $\alpha = \beta$ even when this restriction does not hold. Second, we did 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, 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.[15]

3.3. **Monte Carlo results.** As a benchmark for comparison with the rest of the simulations, we start by shortly reconsidering the small sample behavior of the Hill estimator. From Theorem 1 we recapitulate that the Hill statistic - when applied on regularly varying tails and conditioned on a nuisance parameter m that minimizes AMSE - exhibits small sample bias and standard deviation:

$$E(\widehat{\alpha}) - \alpha \simeq m^{-1/2} sign(b) \alpha^{3/2} (2\beta)^{-1/2}, \tag{3.6}$$

with $s.e.(\widehat{\alpha}) \simeq m^{-1/2}\widehat{\alpha}$ and $m = f(a, b, \alpha, \beta, n)$. Clearly the bias depends on the parameters of the tail expansion (2.1) as well as the

[15]When $\alpha = \beta$, it follows from (3.3) that $m^* = 2^{1/3}(a/b)^{1/3} n^{2/3}$. Thus, the temporal constancy requirement for $c = 2^{1/3}(a/b)^{1/3}$ imposes the same requirement on a and b.

sample size n.[16] We would like to know to what extent these bias and variance properties transmit to the small sample critical values and power properties of the considered stability tests for $\widehat{\alpha}$. We also want to establish whether the bias problem erodes the ability of the tests to accurately locate break dates. To that aim, we simulate from the set of models that have been introduced in the previous section. The optimal number of highest order statistics m^* is chosen analytically by minimizing the Asymptotic Mean Squared Error (AMSE) provided the tail expansion (2.1) is known. For GARCH(1,1) models, the parameters a, b and β stay unknown and we resort to choosing m by applying using the Beirlant et al. (1999) algorithm.

Table 2 reports simulated means and standard deviations of the Hill statistic for samples of size $n = 500, 2000$. Averages and standard deviations are calculated over 20,000 replications.

[Insert Table 2]

The table 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 $|E(\widehat{\alpha}) - \alpha|$ decrease with increasing sample size n when m^* is chosen "optimally" in the sense that AMSE is minimized. Indeed, the right hand side (RHS) panel estimates ($m = 0.1n$) 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 still show a large heterogeneity in small sample bias and estimation accuracy across different distributions. As predicted by the theoretical bias expression (3.6), 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

[16]From the optimal path in (3.4a), it can be easily seen that an increase in the scaling constant a - or, alternatively, a decrease in the 2nd order parameter b - reduces the bias and standard deviation of the Hill statistic via an increase in m^*. The effects of varying α and β on bias and estimation risk are less straightforward to determine and will therefore be derived via Monte Carlo simulation.

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 tranfers 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 $n = 500$ and 2000 we generate 20,000 simulation replications to obtain estimates of the 90th, 95th and 99th percentile of the stability tests' small sample distribution.

[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 presistence) 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 as predicted already by Theorem 2. The critical values for $\rho = -5$ actually come close to the asymptotic critical values reported in QFP. This should not surprise given the fact that the Burr tail is very close to a pure Pareto in the latter case (Hill estimators do not exhibit asymptotic bias for pure Pareto data). But the table also convincingly shows that using asymptotically unbiased critical values leads to huge size distortions in tests of parameter constancy when the true critical values are upward biased due to the asymptotic Hill bias.

Next, Tables 5 and 6 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 by means of the small sample critical values in the previous table. The breakpoint estimates are also based on 20,000 replications but the reported average break point estimates are only based on "candidate"-breaks $\widehat{r}$ that are statistically significant according to the 95% small sample critical values from Tables 3-4.[17]

[Insert Tables 5 and 6]

Just as in the QFP analysis, the direction of change in α seems to be the crucial factor. The recursive and rolling tests 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 that occur 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 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

[17] For break scenarios (α_1, α_2) we calculate the power and break estimates using the 95% small sample critical value that correponds with $\min(\alpha_1, \alpha_2)$. Quintos et al. (2001), on the contrary, calculated the power properties by conditioning on the 95% asymptotic critical values. Moreover, they evaluated the ability to date breaks by averaging over all "candidate"-break points (both statistically significant and insignificant ones).

clearly outperforms the other two tests for most considered DGP's provided the break scenario implies an increase in tail fatness ($\alpha_1 > \alpha_2$).[18] However, the recursive test's inability to detect breaks when $\alpha_1 < \alpha_2$ is more apparent than real, see Candelon and Straetmans (2006). 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 will also be our strategy 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 but 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 but 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 in that pareto-type case.

4. Empirical Application

We perform stability tests for a large cross section of 21 developed and emerging stock markets. Daily price indices (excluding dividends) denominated in local currency were obtained from Datastream inc.[19] Returns are calculated as log first differences and our sample ranges from January 1, 1988 until August 13, 2007 which amounts to $5,117$

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

[19]Stock prices decline at ex-dividend date but these price declines rarely enter the return tails. In other words, working with clean price indices or total return indices does not seem to matter much for the analysis of extreme values.

daily prices.[20] If breaks occur in the tail index of equity market returns, one expects more frequent breaks in emerging stock market tails which makes it interesting to distinguish between developed and emerging stock markets. First, emerging stock markets historically exhibit a higher degree of instability and volatility (i.e. a higher incidence of financial crises). Moreover, the institutional framework (government regulation and supervision, corporate governance culture etc.) in which traders of financial assets have to operate is still in development, changing frequently and not always free of corruption. Also, emerging banking and financial systems are typically less developed than their Western counterparts and central banks are far from independent. Finally, most developing countries are characterized by higher political and country risks (probability of regime changes) which also potentially impacts stock markets.

Given the poor properties of the rolling and sequential test in terms of power and ability to detect breaks, we focus on 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 fat tailed model assumed for the tail. This is because the severity of the Hill bias and its resulting influence on the small sample critical values of stability tests 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. 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 are bootstrapped (CV_B) at the 90, 95 and 99 percent levels.[21] As the

[20] In a twin paper, we also employed the same stability testing framework to a variety of asset classes (including stock indices) that encompass the 2007-2009 financial crisis. This does not fundamentally change the testing results for equity markets, i.e. breaks remains scarce, see Straetmans and Candelon (2013).

[21] 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 vast majority of empirical studies on the magnitude of

scaling factor already 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.[22] We run the recursive test both in calendar time (forward test) and in reverse calendar time (backward test) in order to detect potential falls and rises in the tail index, respectively.

Results of the stability tests are contained in Tables 7 for 21 mature and emerging stock market indices. For sake of comparison, we also calculate stability tests for the QFP sample (this subsample is situated around the Asian crisis from January 2, 1995 until October 16, 1998). Testing results for that subsample are summarized in Table 8. The number of upper order extremes m^* (first table column) used in estimating the test statistic and the bootstrap-based critical values are determined using the Beirlant *et al.* (1999) method. The maximum values for the forward and backward version of test (4.1) are included in the columns labelled Q_F and Q_B, respectively. Evidently, bootstrapped critical values are identical for the forward and backward test. The null of parameter constancy is rejected if the sup-value calculated according to (4.1) exceeds the bootstrap-based critical values, e.g. $Q > CV(p)$ with $p = 5\%$ or 1%. Statistically significant break dates are reported between brackets beneath the testing values (dd/mm/yy).

[Insert Table 7, 8]

The tables only provide minor evidence for time variation in the tail behavior. The subsample results for Hong Kong, Italy, Indonesia and South Korea provide evidence for a decrease in α over the short sample period whereas the full sample results for Taiwan and Hong Kong suggest an increase in α over the extended sample. The subsample break dates are suggestive of a relation between the outbreak of the Asian crisis and the decrease in α (an increase in tail probability mass) whereas the full sample breaks suggest an opposite movement in α, maybe due to a stabilization in the market brought by institutional reform and liberalization in the aftermath of the crisis. However, only the full sample result for Taiwan and the subsample result for South Korea are statistically significant at the 1% significance level. That the (scant)

the tail index for stock markets find that $\widehat{\alpha}$ hoovers around 3 and is significantly above 2, see e.g. Jansen and de Vries (1991) or Longin (1996).

[22] 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.

empirical evidence for structural change in α is mainly generated by emerging market data is not too 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, notice that even emerging markets exhibit stationary tail behavior in a majority of cases. Finally, upon comparing our subsample outcomes with those of QFP, we were unable to find a break for Indonesia and Malaysia. The different testing outcomes may be due to the fact that asymptotically unbiased critical values can lead to overrejection of the null of parameter stability.

The empirical application leaves us with the conviction that heavy tails are quite constant over time and that the tail index α constitutes a structural characteristic of financial series useful for the assessment of long run risk, stress testing and financial stability.

5. Conclusions

This paper provided a thorough study of the small sample behavior of some popular tests for detecting time variation in the tail index of equity returns. The tests are "endogeneous" in the sense that they produce an estimate of the breakpoint location upon detection of a statistically significant break. Our Monte Carlo experiments determined 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. Conform to the preceding empirical literature, the number of upper order extremes is selected by minimizing the sample Mean Squared error of the Hill statistic. The DGP's were chosen to mimic 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 persists in large samples and removing it can be shown to be notoriously difficult. We therefore propose a bootstrap-based procedure to determine the critical values of the stability test when using real-life data. Using simulation-based or bootstrap-based small sample critical values, (recursive) stability tests were found to exhibit satisfactory power and were also able to detect breaks quite well. Applying the test on a large set of emerging and developed stock market data, we were 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.

Appendix A. 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

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

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:

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

see e.g. Mikosch and Starica (1998). Empirical evidence suggests that $2 < \alpha < 4$ and we therefore used these boundary values in the Monte Carlo simulations. When $\alpha = 2$, eq. (C.1) implies that $\lambda + \beta_1 = 1$.This still leaves an infinite number of possible parameter combinations.As for the upper bound value $\alpha = 4$, eq. (C.1) boils down to

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

Upon substitute $\beta_1 = c - \lambda$ $(c < 1)$ into (C.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 simulations, see e.g. Danielsson and de Vries (1997).

Appendix B. 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:

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

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 Fréchet 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 Fréchet case, the 2nd order parameter can be freely chosen for the Burr model. This implies that the Burr distribution becomes indistinguishable from a pure Pareto distribution for large β.

The other DGPs do not exhibit explicitc dfs in closed form which may somewhat complicate 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 we can exploit an existing tail expansion (Ibragimov and Linnik (1971, ch. 2)) in order to determine the parameters in (B.1):

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

Only considering the expanion's first two terms renders the second order approximation:

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

Thus the parameter vector that we need for determining m^* boils down to:

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

It follows from the 2nd order term that the restriction $\alpha = \beta$ holds for the symmetric stable class.[23]

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 asymptotic expansion for the class of regularly varying densities easily follows from (B.1):

$$\begin{aligned} G'(p) &= g(p) \simeq a\alpha p^{\alpha+1} + ab(\alpha+\beta)p^{\alpha+\beta+1} \\ &= ax^{-\alpha-1}\left(\alpha + b(\alpha+\beta)x^{-\beta-1}\right), \end{aligned} \qquad \text{(B.2)}$$

with p small (x large). Upon rewriting the 2nd order Taylor expansion of the student-t density for large x (small p) in the above format, we

[23] This restriction also holds for the Fréchet model because it falls within the symmetric stable class of distributions.

obtain:

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

with x large. This 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 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 are also identically distributed to I.I.D. student-t draws 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 $c=\left(\frac{\alpha(\alpha+\beta)^2}{2\beta^3b^2}a^{2\beta/\alpha}\right)$ leaves the value of $m^*=cn^{\frac{2\beta}{2\beta+\alpha}}$ invariant.

Appendix C. 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

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

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:

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

see e.g. Mikosch and Starica (1998). Empirical evidence suggests that $2 < \alpha < 4$ and we therefore used these boundary values in the Monte Carlo simulations. When $\alpha = 2$, eq. (C.1) implies that $\lambda + \beta_1 = 1$.This still leaves 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. (C.1) boils down to

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

Upon substitute $\beta_1 = c - \lambda$ $(c < 1)$ into (C.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 simulations.

———

References

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

[2] Andrews, D. (1993). "Tests for parameter stability and structural change with unknown change point". Econometrica 59, 817-858.

[3] Beirlant, J., Dierckx, G., Goegebeur, Y., Matthys, G. (1999). "Tail Index Estimation and an Exponential Regression Model". Extremes 2, 177-200.

[4] Candelon, B., Lütkepohl, H. (2001). "On the reliability of Chow type tests for parameter constancy in multivariate dynamic models". Economics Letters 73, 155-160.
[5] Candelon, B., Straetmans, S. (2006). "Testing for multiple regimes in the tail behavior of emerging currency returns". Journal of International Money and Finance 25, 1187-1205.
[6] Danielsson, J., de Haan, L., Peng, L., 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] Danielsson, J., de Vries, C.G. (1997). "Tail index and quantile estimation with very high frequency data". Journal of Empirical Finance 4, 241-257.
[8] Davidson, R., Flachaire, E. (2007). "Asymptotic and Bootstrap Inference for Inequality and Poverty Measures". Journal of Econometrics 141, 141-166.
[9] Drees, H. (2003). "Extreme Quantile Estimation for Dependent Data with Applications to Finance". Bernoulli 9, 617-657.
[10] Dumouchel, W.H. (1983). "Estimating the stable index α in Order to Measure Tail Thickness: A critique". Annals of Statistics 11, 1019-1031.
[11] Embrechts, P., Klüppelberg, C, Mikosch, T. (1997). Modelling Extremal Events. Springer, Berlin.
[12] Feller, W. (1971). An Introduction to Probability Theory and its Applications (Volume I). Wiley, New York (3rd edition).
[13] Galbraith, J.W., Zernov, S. (2004). "Circuit breakers and the tail index of equity returns". Journal of Financial Econometrics 2, 109-129.
[14] Goldie, C.M., Smith, R. (1987). "Slow variation with remainder: Theory and applications". Quarterly Journal of Mathematics 38, 45-71.
[15] Gomes, M.I., Martins, M.J. (2002). "Asymptotically unbiased estimators of the tail index based on external estimation of the second order parameter". Extremes 5, 5-31.
[16] Gomes, M.I., Haan, L. de, Peng, L. (2003). "Semi-parametric estimation of the second order parameter in statistics of extremes". Extremes 5, 387-414.
[17] Haan, L. de, Stadtmüller, U. (1996). "Generalized Regular Variation of Second Order". Journal of the Australian Mathematical Society (Series A) 61, 381-395.
[18] Haan, L. de, Jansen, D.W., Koedijk, K.G., 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.
[19] Haeusler, E., Teugels, J. (1985). "On asymptotic normality of Hill's estimator for the exponent of regular variation". Annals of Statistics 13, 743-756.
[20] Hall, P. (1982). "On some simple estimates of an exponent of regular variation". Journal of the Royal Statistical Society (Series B) 42, 37-42.
[21] Hartmann, P., Straetmans, S., 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: Implications for Monetary, Regulatory and International Policies, MIT Press, Cambridge (MA): 361-383.
[22] Hartmann, P., Straetmans, S., de Vries, C.G. (2004). "Asset market linkages in crisis periods". Review of Economics and Statistics 86, 313-326.

[23] Hartmann, P., Straetmans, S., de Vries, C.G. (2006). "Banking System Stability: a Cross-Atlantic Perspective". In: Carey, M., Stulz, R.M. (Eds.), the Risk of Financial Institutions, The University of Chicago Press (Chicago and London): 133-193.

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

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

[26] Ibragimov, I.A., Linnik, Y.V. (1971). Independent and stationary sequences of random variables. Wolters-Noordhof, Groningen.

[27] Jansen, D.W., Koedijk, K.G., de Vries, C.G. (2000). "Portfolio selection with limited downside risk". Journal of Empirical Finance 7, 247-269.

[28] Jansen, D.W., 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.

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

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

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

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

[33] 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.

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

[35] Mikosch, T., Starica, C. (2000). "Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process". Annals of Statistics 28, 1427–1451.

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

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

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

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

[40] Ross, S.A, Westerfield, R.W., Jaffe, J. (2005). Corporate Finance. McGraw-Hill (7th International Edition).

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

[42] Straetmans, S., Candelon, B. (2013). "Long-term asset tail risks in developed and emerging markets". Journal of Banking and Finance 37, 1832-1844.

[43] Straetmans, S., Verschoor, W., Wolff, C. (2008). "Extreme U.S. Stock Market Fluctuations in the Wake of 9/11". Journal of Applied Econometrics 23, 17-42.

[44] Wagner, N. (2005). "Autoregressive conditional tail behavior and results on government bond yield spreads". International Review of Financial Analysis 14, 247-261.

[45] Werner, T., Upper, C. (2002). "Time variation in the tail behaviour of Bund futures returns". Journal of Futures Markets 24, 387-398.

TABLE 1. Tail expansion parameters and corresponding optimal nuisance parameters for selected Data Generating processes

	Tail expansion parameters				opt. m^*
	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	$2^{1/3}\left(\frac{a}{b}n\right)^{2/3}$
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	$-\frac{2}{\alpha}$	$\left(\frac{\alpha(\alpha+2)^2a^{4/\alpha}}{16b^2}\right)^{\frac{\alpha}{4+\alpha}}n^{\frac{4}{4+\alpha}}$
Frechet	1	$-1/2<0$	α	-1	$2n^{2/3}$
Burr	1	$-\rho^{-1}<0$	nonrestricted		$\left(-\frac{(1-\rho)^2}{2\rho}\right)^{\frac{1}{1-2\rho}}n^{-\frac{2\rho}{1-2\rho}}$

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

	$\widehat{\alpha}$ $(s.e.)$			
	$m = m^*$		$m = 0.1n$	
	$n = 500$	$n = 2,000$	$n = 500$	$n = 2,000$
	Panel A: I.I.D. models			
Stable(α)				
1.2	1.23 (0.16)	1.24 (0.10)	1.25 (0.18)	1.23 (0.09)
1.5	1.79 (0.44)	1.66 (0.25)	1.77 (0.27)	1.74 (0.13)
Student(α)				
2	1.79 (0.28)	1.87 (0.19)	1.72 (0.23)	1.70 (0.11)
4	3.16 (0.74)	3.40 (0.56)	2.43 (0.31)	2.41 (0.15)
Frechet(α)				
2	1.88 (0.16)	1.92 (0.10)	1.98 (0.28)	1.96 (0.14)
4	3.75 (0.32)	3.84 (0.21)	3.97 (0.57)	3.91 (0.28)
Burr($\alpha, -\rho$)				
$(2, -0.5)$	1.68 (0.28)	1.77 (0.21)	1.59 (0.21)	1.57 (0.10)
$(2, -5)$	1.97 (0.11)	1.98 (0.06)	2.04 (0.29)	2.01 (0.14)
$(4, -0.5)$	3.37 (0.56)	3.56 (0.42)	3.18 (0.43)	3.14 (0.21)
$(4, -5)$	3.93 (0.22)	3.96 (0.12)	4.08 (0.59)	4.02 (0.29)
	Panel B: time dependence (1st or 2nd moment)			
AR(α, θ)				
$(1.5, 0.2)$	1.82 (0.48)	1.68 (0.27)	1.78 (0.29)	1.74 (0.14)
$(1.5, 0.4)$	1.87 (0.55)	1.70 (0.32)	1.80 (0.34)	1.75 (0.17)
SVSTU(α, θ)				
$(4, 0.85)$	3.21 (0.76)	3.42 (0.57)	2.45 (0.33)	2.41 (0.16)
$(4, 0.95)$	3.27 (0.77)	3.43 (0.56)	2.52 (0.37)	2.43 (0.19)
GARCH(α, θ)				
$(4, 0.85)$	4.38 (1.77)	3.75 (0.82)	2.82 (0.45)	2.66 (0.23)
$(4, 0.95)$	4.74(1.99)	3.88 (0.92)	2.81 (0.50)	2.62 (0.27)

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 tests: i.i.d. models

DGP	n=500			n=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)
Burr(α, ρ)						
$(2, -1)$	1.81 (0.03)	2.43 (0.03)	4.35 (0.19)	1.80 (0.02)	2.29 (0.03)	3.69 (0.12)
$(2, -5)$	1.54 (0.03)	1.95 (0.04)	3.07 (0.09)	1.56 (0.01)	1.93 (0.01)	2.84 (0.07)
			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)
Burr(α, ρ)						
$(2, -1)$	1.95 (0.01)	2.64 (0.03)	4.64 (0.16)	1.73 (0.02)	2.22 (0.04)	3.51 (0.06)
$(2, -5)$	1.66 (0.01)	2.10 (0.03)	3.25 (0.09)	1.53 (0.01)	1.82 (0.02)	2.55 (0.05)
			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)
Burr(α, ρ)						
$(2, -1)$	19.03 (0.33)	27.13 (0.48)	49.78 (1.60)	16.80 (0.31)	23.09 (0.61)	39.84 (1.17)
$(2, -5)$	20.14 (0.24)	27.72 (0.59)	49.08 (1.21)	19.75 (0.21)	26.37 (0.42)	43.87 (0.88)

Note: critical values are reported for varying sample sizes (n), and different levels of statistical significance. Critical values are based on 20,000 Monte Carlo replications. Corresponding standard errors for the critical values are reported between brackets (s.e.). The parameters α and $\rho = -\beta/\alpha$ refer to the tail index and the second order parameter, respectively.

Table 4. Small sample critical values for recursive, rolling and sequential tests: dependent models

DGP	$n = 500$			$n = 2,000$		
	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 varying sample sizes (n), and different levels of significance. Critical values are based on 20,000 Monte Carlo replications. Corresponding standard errors for the critical values are reported between brackets (s.e.). The 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 5. Size-corrected finite sample power for recursive, rolling and sequential tests

DGP($\alpha_1;\alpha_2$)	$n=500$			$n=2,000$		
	$r=0.25$	$r=0.5$	$r=0.75$	$r=0.25$	$r=0.50$	$r=0.75$
Stable$(1.5,1.2)$						
- rec	22	32	25	53	71	55
- rol	7	8	5	32	38	22
- seq	14	28	42	15	45	69
Stable$(1.2,1.5)$						
- rec	1.18	1.36	1.5	1.96	2.66	1.1
- rol	5	9	7	22	37	30
- seq	6	3	1	10	4	2
Student$(4,2)$						
- rec	21	32	24	49	73	62
- rol	6	6	4	27	32	15
- seq	10	21	35	12	41	71
Student$(2,4)$						
- rec	0.5	0.7	2	2.54	0.94	1.18
- rol	4	6	5	15	31	27
- seq	3	1	0.4	0.6	0.1	0.6
Burr$(4,2)$						
$\rho=-1$						
- rec	31	37	26	52	66	53
- rol	19	22	15	35	45	30
- seq	17	41	57	14	45	68
Burr $(2,4)$						
$\rho=-1$						
- rec	1.16	0.5	1.3	0.3	0.22	1.76
- rol	15	21	19	32	46	36
- seq	1	0.2	0.08	0.28	0.02	0.1
SVSTU$(4,2)$						
$\theta=0.95$						
- rec	20	31	23	49	71	59
- rol	12	22	16	43	66	55
- seq	10	22	36	12	43	71
SVSTU$(2,4)$						
$\theta=0.95$						
- rec	0.56	0.70	1.80	3.16	1.84	1.16
- rol	0.2	0.26	0.66	1.28	0.74	0.8
- seq	1.82	1.1	1.98	0.34	0.20	0.64
ARCH(4,2)						
- rec	6.2	16.74	22.18	17.42	31.54	22.56
- rol	2.7	3.62	4.16	7.94	15.86	22.16
- seq	6.72	10	14.88	9.14	20.24	40.6
ARCH(2,4)						
- rec	0.34	2.86	2.7	0.06	0.36	0.94
- rol	2.82	1.8	1.2	5.4	3.44	2.06
- seq	1.30	0.90	1.74	0.32	0.08	0.54

Note: the power is reported for different sample sizes (n=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 finite sample critical values and is calculated as the rejection frequency under the null hypothesis of parameter constancy using 20,000 Monte Carlo replications. The parameters α and $\rho=-\beta/\alpha$ refer to the tail index and the second order parameter, respectively. The volatility persistence parameter in the stochastic volatility models with student-t innovations (SVSTU) is denoted by θ.

Table 6. Breakpoint estimates for recursive, rolling and sequential tests

DGP($\alpha_1;\alpha_2$)	$n = 500$			$n = 2,000$		
	$r = 0.25$	$r = 0.5$	$r = 0.75$	$r = 0.25$	$r = 0.50$	$r = 0.75$
Stable$(1.5, 1.2)$						
- rec	0.42 (0.17)	0.53 (0.13)	0.64 (0.16)	0.33 (0.12)	0.50 (0.10)	0.66 (0.14)
- rol	0.37 (0.22)	0.39 (0.14)	0.51 (0.18)	0.65 (0.10)	0.36 (0.11)	0.48 (0.18)
- seq	0.81 (0.08)	0.79 (0.09)	0.81 (0.04)	0.76 (0.14)	0.70 (0.13)	0.80 (0.05)
Stable$(1.2, 1.5)$						
- rec	0.62 (0.11)	0.55 (0.15)	0.48 (0.17)	0.48 (0.13)	0.48 (0.08)	0.48 (0.16)
- rol	0.68 (0.17)	0.81 (0.14)	0.86 (0.20)	0.72 (0.17)	0.84 (0.11)	0.95 (0.09)
- seq	0.83 (0.03)	0.83 (0.05)	0.84 (0.01)	0.83 (0.02)	0.82 (0.03)	0.82 (0.03)
Student$(4, 2)$						
- rec	0.40 (0.17)	0.53 (0.13)	0.67 (0.15)	0.33 (0.13)	0.51 (0.10)	0.70 (0.11)
- rol	0.37 (0.22)	0.39 (0.14)	0.49 (0.17)	0.26 (0.10)	0.37 (0.11)	0.49 (0.18)
- seq	0.80 (0.09)	0.78 (0.10)	0.81 (0.04)	0.78 (0.11)	0.71 (0.13)	0.80 (0.04)
Student$(2, 4)$						
- rec	0.58 (0.20)	0.42 (0.22)	0.41 (0.17)	0.53 (0.11)	0.51 (0.12)	0.39 (0.15)
- rol	0.71 (0.18)	0.81 (0.14)	0.85 (0.20)	0.72 (0.17)	0.83 (0.11)	0.95 (0.08)
- seq	0.58 (0.20)	0.42 (0.22)	0.41 (0.17)	0.53 (0.11)	0.51 (0.12)	0.39 (0.15)
Burr$(4, 2)$						
$\rho = -1$						
- rec	0.38 (0.16)	0.51 (0.16)	0.61 (0.19)	0.30 (0.09)	0.50 (0.08)	0.70 (0.10)
- rol	0.28 (0.13)	0.37 (0.13)	0.48 (0.18)	0.26 (0.10)	0.37 (0.11)	0.48 (0.18)
- seq	0.73 (0.17)	0.74 (0.12)	0.81 (0.05)	0.63 (0.22)	0.66 (0.15)	0.79 (0.06)
Burr$(2, 4)$						
$\rho = -1$						
- rec	0.62 (0.16)	0.47 (0.22)	0.40 (0.19)	0.65 (0.16)	0.31 (0.19)	0.37 (0.14)
- rol	0.71 (0.18)	0.82 (0.12)	0.92 (0.12)	0.71 (0.17)	0.83 (0.11)	0.94 (0.10)
- seq	0.82 (0.04)	0.82 (0.04)	0.69 (0.16)	0.67 (0.17)	0.70 (0.14)	0.75 (0.12)
SVSTU$(4, 2)$						
$\theta = 0.95$						
- rec	0.40 (0.17)	0.52 (0.14)	0.64 (0.17)	0.35 (0.14)	0.51 (011)	0.56 (0.17)
- rol	0.33 (0.19)	0.40 (0.14)	0.49 (0.19)	0.26 (0.10)	0.37 (0.11)	0.48 (0.17)
- seq	0.77 (0.12)	0.75 (0.11)	0.80 (0.06)	0.78 (0.10)	0.70 (0.13)	0.79 (0.06)
SVSTU$(2, 4)$						
$\theta = 0.95$						
- rec	0.63 (0.16)	0.43 (0.21)	0.46 (0.19)	0.58 (0.14)	0.58 (0.13)	0.58 (0.17)
- rol	0.72 (0.18)	0.79 (0.15)	0.87 (0.18)	0.70 (0.17)	0.83 (0.11)	0.95 (0.08)
- seq	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)						
- rec	0.41 (0.18)	0.52 (0.15)	0.65 (0.19)	0.37 (0.15)	0.60 (0.14)	0.75 (0.09)
- rol	0.49 (0.27)	0.49 (0.23)	0.54 (0.21)	0.31 (0.17)	0.39 (0.15)	0.49 (0.19)
- seq	0.71 (0.19)	0.73 (0.14)	0.78 (0.10)	0.77 (0.13)	0.73 (0.12)	0.81 (0.05)
ARCH(2,4)						
- rec	0.35 (0.23)	0.71 (0.20)	0.67 (0.21)	0.71 (0.15)	0.82 (0.05)	0.82 (0.06)
- rol	0.70 (0.20)	0.70 (0.21)	0.62 (0.27)	0.71 (0.17)	0.78 (0.16)	0.88 (0.19)
- seq	0.77 (0.17)	0.64 (0.24)	0.63 (0.18)	0.84 (0.07)	0.54 (0.26)	0.61 (0.12)

Note: estimated break dates are reported for different sample sizes (n=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" break dates are calculated over 20,000 Monte Carlo replications. Average break date estimates are obtained by averaging over the statistically significant "candidate" breaks using the finite sample critical values. The parameters α and $\rho = -\beta/\alpha$ refer to the tail index and the second order parameter, respectively. The volatility persistence parameter in stochastic volatility models with student-t innovations (SVSTU) is denoted by θ.

TABLE 7. Forward and backward recursive testing outcomes (Full sample: 01/01/1988-13/08/2007)

Country	m^*	Recursive test		Bootstrapped critical values		
		Q_F	Q_B	$CV(90\%)$	$CV(95\%)$	$CV(99\%)$
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. Critical values are based on 20,000 bootstrapped sample replications. Statistically significant rejections of the null hypothesis of tail index constancy at the 5% and 1% significance level are denoted by ** and ***, respectively. The break dates (dd/mm/yy) of corresponding significant breaks are reported in bold.

TABLE 8. Structural change tests (Full sample 02/01/1995-16/10/1998)

Country	m^*	Recursive test		Bootstrapped critical values		
		Q_F	Q_B	$CV(90\%)$	$CV(95\%)$	$CV(99\%)$
Argentina	52	0.418	2.164	2.401	2.908	4.232
Australia	83	0.399	0.378	2.544	3.040	4.166
Austria	39	0.982	0.291	2.601	3.229	5.024
Britain	61	1.283	0.208	2.489	3.026	4.297
Canada	43	1.147	0.109	2.111	2.737	4.880
Chile	118	0.62	0.673	1.922	2.334	3.441
France	50	1.097	0.262	2.138	2.616	3.849
Germany	75	0.399	0.206	2.549	3.037	4.137
Hong-Kong	49	3.691*	0.177	3.181	3.859	5.293
(break date)		(21/1/98)				
India	88	0.699	1.210	2.697	3.263	4.493
Indonesia	31	3.283**	0.574	2.554	3.235	5.648
		(28/1/98)				
Italy	66	2.982**	0.856	1.858	2.366	3.920
(break date)		30/9/97				
Japan	65	2.115	0.198	3.550	4.241	5.588
Korea	94	7.802***	0.261	2.398	2.854	3.763
(break date)		(28/8/97)				
Malaysia	67	0.733	0.408	2.252	2.758	4.093
Mexico	74	3.113	0.221	3.754	4.423	5.848
Philippines	58	1.126	0.149	2.497	3.045	4.218
Spain	52	1.64	0.304	2.113	2.672	4.219
Taiwan	29	0.756	0.835	2.592	3.364	6.326
Thailand	55	1.191	0.269	2.481	3.044	4.311
US	43	0.512	0.39	2.082	2.732	4.348

Note: The forward and backward version of the recursive test are denoted by Q_F and Q_B, respectively. Critical values are based on 20,000 bootstrapped sample replications. Statistically significant rejections of the null hypothesis of tail index constancy at the 5% and 1% significance level are denoted by ** and ***, respectively. The break dates (dd/mm/yy) of corresponding significant breaks are reported in bold.