Journal of Banking & Finance 37 (2013) 1832–1844

Contents lists available at SciVerse ScienceDirect

Journal of Banking & Finance

journal homepage: www.elsevier.com/locate/jbf

Long-term asset tail risks in developed and emerging markets

Stefan Straetmans [a,*], Bertrand Candelon [b]

[a] Maastricht University, School of Business and Economics, Department of Finance, P.O. Box 616, 6200 MD Maastricht, The Netherlands
[b] Maastricht University, School of Business and Economics, Department of Economics, P.O. Box 616, 6200 MD Maastricht, The Netherlands

ARTICLE INFO

Article history:
Available online 9 October 2012

JEL classification:
F31
G15
G19
C49

Keywords:
Tail index
Extreme value analysis
Endogenous stability test
Finite sample properties
Bootstrap

ABSTRACT

A power law typically governs the tail decay of financial returns but the constancy of the so-called tail index which dictates the tail decay remains relatively unexplored. We study the finite sample properties of some recently proposed endogenous tests for structural change in the tail index. Given that the finite sample critical values strongly depend on the tail parameters of the return distribution we propose a bootstrap-based version of the structural change test. Our empirical application spans developed and emerging financial asset returns. Somewhat surprisingly, emerging stock market tails are not more inclined to structural change than their developed counterparts. Emerging currency tails, on the contrary, do exhibit structural shifts in contrast to developed currencies. Our results suggest that extreme value theory (EVT) applications in hedging tail risks can assume stationary tail behavior over long time spans provided one considers portfolios that solely consist of stocks or bonds.

1. Introduction

The 1997 Asian crisis, the LTCM debacle or the recent subprime credit crunch have increased the awareness of both academics and practitioners on the importance of accurately assessing the likelihoods of so-called extreme events. Stated otherwise, fluctuations in financial markets whose occurrence is relatively rare can drive banks or institutional investors into overnight financial distress when they strike. However, the academic interest into large tail events is far from new (for an early reference see e.g. Mandelbrot, 1963). He was one of the first to acknowledge that overnight financial market turbulence cannot be captured by the normal distribution function (df). More specifically, tail probabilities show a polynomial tail decay ("heavy" tails) in contrast to the exponential tail decays of so-called "thin tailed" models like the normal df and most financial asset classes exhibit this "heavy tail" characteristic. Numerous empirical studies focus on identifying the degree of probability mass in the tail by estimating the so-called tail index α.[1] The integer part of this parameter reflects the number of bounded statistical moments of the corresponding unconditional df.

The causes and consequences of changes in the tail index (provided changes occur) remain relatively unexplored. Conditional volatility models like the GARCH-type class reconcile a stationary unconditional df (constant tail index) 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-term risk like unconditional quantiles) remains invariant over long time periods. In other words: can highly volatile periods like the 2007–2010 financial turmoil and periods of market quiescence both be explained by a single unconditional df? Potential causes of tail index changes include structural shifts like e.g. changing trading systems, financial regulatory reform and financial liberalization or changes in the political environment. Moreover, economists seem to agree that these structural changes are more frequently happening in emerging economies. Our empirical application therefore distinguishes between developed and emerging return tails in order to evaluate whether emerging return tails are relatively more prone to structural shifts in the tail index.

Testing for structural change in the tail behavior of the unconditional distribution is relevant from both a statistical and economic perspective. First, whether extreme value theory (EVT) or e.g. the cited GARCH models are applicable depends on the stationarity assumption for the unconditional tail. Also, a

* Corresponding author. Tel.: +31 43 388 36 79; fax: +31 43 388 48 75.
E-mail addresses: s.straetmans@maastrichtuniversity.nl (S. Straetmans), b.candelon@maastrichtuniversity.nl (B. Candelon).

[1] Jansen and de Vries (1991), Longin (1996) and Hartmann et al. (2004) investigate the probability mass in the tails of stock market returns; whereas Koedijk et al. (1990, 1992), Hols and de Vries (1991) and Hartmann et al. (2003) consider fat tails in foreign exchange rate returns. Bond extremes remain relatively unexplored except for de Haan et al. (1994) and Hartmann et al. (2004).

0378-4266/$ - see front matter
http://dx.doi.org/10.1016/j.jbankfin.2012.09.022

non-constant tail index implies a violation of covariance stationarity which complicates standard statistical inference based on regression analysis. From an economic perspective, quantifying the correct level of the tail index is relevant for risk managers as it constitutes a necessary ingredient for calculating the unconditional Value-at-Risk (VaR) very far into the distributional tail, i.e. so-called "tail risk". Indeed, whereas regulatory instances require the financial industry to report and backtest 5% and 1% VaR, these events hardly represent extreme events that can trigger financial companies into overnight financial distress. Instead, evaluating downside risk much further into the tail represents useful additional information to e.g. traditional stress testing approaches. Other EVT applications in portfolio selection and risk management include safety first portfolio selection for pension funds (Jansen et al., 2000) or the assessment of trading limits for unhedged forex positions in commercial banks (see Danielsson and de Vries, 1997). If one incorrectly assesses the actual tail index value in these exercises due to e.g. the presence of structural breaks, unconditional VaR quantiles are most probably biased which erodes the effectiveness of financial risk management and the proper monitoring of overall financial stability (e.g. wrong allocation of risky investments in pension fund portfolios, wrong trading limits for forex traders within banks, etc.).

The scant empirical literature on the constancy issue mainly focuses on testing for a single known (i.e. exogenously selected) breakpoint in the tail index.[2] To the best of our knowledge, Quintos et al. (2001) constitutes the only stability study on detecting (single) breakpoints as well as corresponding break dates in the tail index.[3] Our study extends and refines the previous breakpoint analyses in several directions. First, we select the number of extreme returns to estimate the tail index 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 tail index 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 a large cross section of assets and asset classes whereas previous studies typically only focus on a limited number of assets within the same asset class. We also distinguish between developed market financial assets and emerging market financial assets in order to judge whether the latter are more prone to shifts in the tail behavior.

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 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 (at least 2000 observations). Upon applying a bootstrap-based version of this test to a large cross section of assets and asset classes, we mainly detect breaks in the tail behavior of emerging currencies.

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 extensive empirical investigation on the tail stability of a variety of developed and emerging asset tails. Section 5 contains concluding remarks.

2. Testing structural change in tail behavior: theory

We provide a short digression on the theory and estimation of the tail index α followed by a discussion of some temporal stability tests for this parameter. 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 et al. (1997). Without loss of generality, we express estimation and testing procedures in terms of the right tail, i.e. the survivor function $P\{X \geqslant x\} := 1 - F(x)$. Our empirical investigation focuses on sharp drops in the prices of risky securities. This requires taking the negative of a return series prior to applying the sketched framework. Under fairly general conditions, we can approximate the survivor function of heavy tailed (or "regularly varying") distributions by the second order Taylor expansion for large x:

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

with $a > 0,\ \alpha > 0,\ b \in \Re,\ \beta > 0$, see e.g. de Haan and Stadtmüller (1996). The parameters β and b that govern the second order behavior in (1) reflect the deviation from pure Pareto behavior in the tail. Notice that if we talk about the "second order parameter" of a fat tailed or regularly varying process later on in the paper, we always refer to the ratio $\rho = -\beta/\alpha$. The case $\beta = \rho = 0$ corresponds to the expansion $P\{X \geqslant x\} \approx ax^{-\alpha}[1 + b\ln x]$. The tail specializes to an exact Pareto when $b = 0$.

The regular variation property implies that the (appropriately scaled) upper extremal 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 (1) decays if x is increased. A lower tail index implies a slower probability decay and higher probability mass in the tail of X, *ceteris paribus* the level of x. The regular variation property, *inter alia*, implies that distributional moments $E(X^r)$ with $r > \alpha$, are unbounded, signifying "fat tails". Regularly varying probability distributions include the Student-*t*, symmetric stable, Burr, and Frechet df as well as the GARCH class of conditional volatility models.[4] 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 $\phi(x)$:

$$1 - \Phi(x) \approx \frac{\phi(x)}{x}, \quad x \quad \text{large} = (2\pi x)^{-1}\exp\left(-\frac{1}{2}x^2\right),$$

which clearly describes an exponentially declining tail, see Feller (1971a, p. 175). We classify distributions with this type of tail decay as "thin tailed" because the tail probability $1 - \Phi(x)$ declines much

[2] 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 the tail index 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.

[3] 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 for the tail index and lead to overrejection of the null hypothesis of tail index constancy.

[4] 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 (1) holds, though, with α equal to the degrees of freedom parameter and $\beta = 2$. As for the Burr df, the 2nd order parameter can be freely chosen. The value of β is unknown for the GARCH class.

faster to zero as in (1); but these distributions possess all moments, and hence do not capture what is typically observed in financial data.

The paper's focus is on the finite sample properties of (single break) temporal stability tests for tail index estimators. The scrutinized test statistics use Hill's (1975) estimator as an input. Let $X_{1,n} \leqslant X_{2,n} \leqslant \cdots \leqslant X_{n,n}$ represent the ascending order statistics that correspond with the returns series X for a sample of size n.

Hill's estimator boils down to:

$$\hat{\alpha} = \left(\frac{1}{m} \sum_{j=0}^{m-1} \ln \left(\frac{X_{n-j,n}}{X_{n-m,n}} \right) \right)^{-1}, \quad (2)$$

such that $\hat{\alpha} > 0$, with m the number of highest order statistics used in 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 following theorem summarizes the main convergence in distribution result for the Hill estimator:

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

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

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

see e.g. Hall (1982) and Haeusler and Teugels (1985) for the i.i.d. case ($\eta = 1$). Quintos et al. (2001) generalize this result to stationary GARCH processes with conditionally normal innovations.

Loosely speaking, Theorem 1 implies that proper convergence in distribution 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 $\kappa = m/n$. Previous studies argue that this simple rule-of-thumb performs well in finite samples but its lack of asymptotic justification constitutes a fundamental problem (see e.g. Dumouchel, 1983). We will therefore abstain from using this criterion.

Condition (B) in Theorem 1 provides a natural alternative towards selecting the nuisance parameter because one can easily show that the expression for the nuisance parameter m under (B) minimizes the Asymptotic Mean Squared Error (AMSE) for the Hill estimator (see e.g. Danielsson and de Vries, 1997). Virtually all empirical EVT studies exploit the AMSE minimization principle and we therefore use this criterion in the rest of the paper.[5] Theorem 1 also shows that the AMSE criterion induces an asymptotic bias in the Hill statistic, i.e. $E(\hat{\alpha} - \alpha) \approx m^{-1/2}\varphi\alpha$. We will thoroughly document the finite sample consequences of this bias effect on the accompanying stability tests in the Monte Carlo simulation section (Section 3). For more elaborate expositions on extreme value theory and estimation, see e.g. the monographs by Leadbetter et al. (1983) or Embrechts et al. (1997).

The main goals of the paper are to investigate the finite sample properties of a trio of (single break) stability tests for the Hill statistic introduced earlier by Quintos et al. (2001) and to apply it to detect single breaks in a large set of assets and asset classes from developed and emerging markets. The stability tests differ in the way subsamples are constructed for the Hill estimates. We define the recursive estimator on subsamples $[1; t] \subset [1; n]$ as follows:

$$\hat{\alpha}_t = \left(\frac{1}{m_t} \sum_{j=0}^{m_t-1} \ln \left(\frac{X_{t-j,t}}{X_{t-m_t,t}} \right) \right)^{-1}, \quad (3)$$

with $m_t = ct^{\frac{2\beta}{(2\beta+\alpha)}}$. We condition the rolling estimator on a fixed subsample size $w < n$. Rolling over the subsample requires shifting the subsample through the full sample by eliminating past observations and adding future observations whilst keeping the subsample size constant at w:

$$\hat{\alpha}^*_t = \left(\frac{1}{m_w} \sum_{j=0}^{m_w-1} \ln \left(\frac{X_{w-j,w}}{X_{w-m_w,w}} \right) \right)^{-1}, \quad (4)$$

with $m_w = cw^{\frac{2\beta}{(2\beta+\alpha)}}$. Finally, in order to calculate the sequential test statistic, we partition the total sample in recursive subsamples $[1; t]$ and $[t + 1; n]$ and shift t (reflecting the potential break) through the full sample. We calculate subsample Hill statistics (recursive estimators) for both subsamples using (3). The recursive Hill estimators for the first and the second recursive subsamples correspond with $\hat{\alpha}_{1t}$ and $\hat{\alpha}_{2t}$, respectively. One also often refers to the latter estimator as the "reverse" recursive estimator because it requires (3) to be calculated in reverse calendar time.[6]

We construct the three (recursive, rolling and sequential) tests using the sequences:

$$Y_n^2(r) = \left(\frac{tm_t}{n} \right) \left(\frac{\hat{\alpha}_t}{\hat{\alpha}_n} - 1 \right)^2, \quad (5)$$

$$V_n^2(r) = \left(\frac{wm_w}{n} \right) \left(\frac{\hat{\alpha}^*_t}{\hat{\alpha}_n} - 1 \right)^2, \quad (6)$$

$$Z_n^2(r) = \left(\frac{tm_t}{n} \right) \left(\frac{\hat{\alpha}_t}{\hat{\alpha}_{2t}} - 1 \right)^2, \quad (7)$$

with $r = t/n$ representing a fraction of the full sample. Expressions (5) and (6) reflect the fluctuations in the Hill statistic's recursive and rolling values relative to the full sample Hill statistic whereas the sequential test uses (7) to compare the fluctuations of the recursive estimator with the reverse recursive estimator.

The null hypothesis of a time invariant tail index α boils down to:

$$H_0 : \alpha_{[nr]} = \alpha, \quad \forall r \in R_\varepsilon = [\varepsilon; 1 - \varepsilon] \subset [0; 1], \quad (8)$$

with $[nr]$ representing the integer value of nr. One would like to test this null hypothesis against the two-sided alternative hypothesis H_A: $\alpha_{[nr]} \neq \alpha$. For sake of convenience we calculate the above test 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 common in the construction of parameter constancy tests (see e.g. Andrews, 1993).[7] Conform with Quandt's (1960) seminal work on endogenous breakpoint determination in linear time series models, we select the candidate break date r where the sequences (5)–(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 finite 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

[5] In principle, one can also use condition (A) as a selection criterion. For example, choose a strictly positive δ in $m^* = cn^{\frac{2\beta}{(2\beta+\alpha)}-\delta}$. However, although this criterion guarantees asymptotic unbiasedness, finite sample bias still exists. 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 typically penalize bias and variance equally and prefer to trade-off bias and variance such as under condition (B). Finally, it is unclear how to choose δ.

[6] We calculate the recursive Hill statistic $\hat{\alpha}_{2t}$ using Eq. (3) on the subsample $[n; t + 1]$ which implies an inversion of the observations' chronology, i.e. we put the most recent observations at the beginning of the subsample.

[7] The restricted choice of r implies that $\varepsilon n \leqslant t \leqslant (1 - \varepsilon)n$. Conform with Andrews (1993), we set $\varepsilon = 0.15$.

financial time series. Each model exhibits regularly varying tails and obeys the asymptotic second order expansion (1). The number of upper order extremes for the Hill statistic minimizes the Asymptotic Mean Squared Error of the Hill estimator. We calculate finite sample critical values and size-corrected finite 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% significance level. Section 3.1 provides a short description of the main data generating processes (DGP's). Section 3.2 contains the analytic derivation of the nuisance parameters for these DGP's. Section 3.3 reports finite sample critical values, size-corrected power properties and the ability to date breaks.

3.1. Data generating processes

We choose a variety of heavy tailed DGP's and accompanying parameter values (a, b, α, β) that all obey the asymptotic expansion (1). We base our Monte Carlo simulations on the symmetric stable df, Student-t, Burr, i.e. $P\{X > x\} = (1 + x^{\beta})^{1/\rho}$ with $\rho = -\beta/\alpha$, AR(1) with stable innovations, GARCH(1,1) with conditionally normal errors and a Stochastic Volatility model.[8] Thus, we distinguish between i.i.d. draws and dependent draws. The tail index of the Student-*t* and Burr distribution functions varies between 2 and 4 which is in line with all previous empirical research on heavy tails in financial markets.[9]

Previous studies, including the Quintos et al. (2001) paper, only study the finite sample behavior of stability tests for the tail index under the class of stable distribution functions (dfs). However, the symmetric stable model has some severe drawbacks as a device for modelling financial returns. First, the property that sums of stable dfs remain stable distributed after appropriate scaling (additivity property) seems overly restrictive for real-life data. Indeed, Feller (1971b, p. 278) shows that the class of regularly varying or "heavy tailed" dfs exhibits additivity in the tail area but not over the full distributional support. Also, the stable class fails to have a finite variance when their tail index is lower than 2. Finally, the normal df is a "local alternative" for the stable model which implies that stable processes with a tail index only slightly smaller than 2 can hardly be distinguished from a normal df on the basis of Hill estimates in very large samples. The other models we use for simulation do not suffer from these drawbacks.

In order to further mimic the time series properties of real-life financial return data, we also use models that exhibit dependence in returns and volatility. We use an AR(1) process with nonzero first order serial correlation θ and with symmetric stable innovations to generate serially dependent data due to market microstructure effects in high frequency data, see e.g. Andersen and Bollerslev (1997). In order to generate persistence in volatility, we employ two distinct models. First, we implement the following model to generate returns X proposed by Danielsson et al. (2001):

$$X_t = U_t \sqrt{\frac{v}{\chi^2(v)}} H_t, \quad P\{U_t = -1\} = P\{U_t = 1\} = 0.5$$

$$H_t = \beta Q_t + \theta H_{t-1},$$

with $\beta = 0.1$, $\theta = 0.9$ and where Q is drawn from a standard normal df. As usual, $\chi^2(v)$ stands for the chi-square distribution with v degrees of freedom. The unconditional df of X is Student-*t* distributed with $\alpha = v$ degrees of freedom. The multiplicative factor U guarantees the fair game property $E_{t-1}X_t = 0$ but preserves the volatility clustering feature. Second, we also simulate from a GARCH(1,1) model with conditionally normal innovations. We choose the sum of the GARCH volatility parameters $\theta = \beta_0 + \beta_1$ such that the tail index of the corresponding unconditional df equals 4. The GARCH model class enables one to change the volatility persistence θ *ceteris paribus* the tail index.[10]

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 one uses from the distributional centre the smaller is the variance of the estimator at the cost of an increase in bias. Goldie and Smith (1987) therefore propose to select the number of upper order extremes m used in estimating (2) by minimizing the Asymptotic Mean Squared Error (AMSE) of the Hill statistic. Using the second order expansion (1) for regularly varying tails, Danielsson and de Vries (1997) derive an expression of the AMSE for the Hill estimator:

$$\text{AMSE}(\hat{\alpha}) = a^{-2\beta/\alpha} \frac{1}{\alpha^2} \frac{\beta^2 b^2}{(\alpha+\beta)^2} \left(\frac{m}{n}\right)^{\frac{2\beta}{\alpha}} + \frac{1}{\alpha^2 m}, \tag{9}$$

where the first part is the squared bias and the second part is the asymptotic 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 at least one of these parameters equals zero, the bias term disappears and the distributional tail (1) specializes to an exact Pareto. Minimizing (9) w.r.t. m renders the optimal number of highest order statistics:

$$m^* = cn^{2\beta/(2\beta+\alpha)}, \; c = \left(\frac{\alpha(\alpha+\beta)^2}{2\beta^3 b^2} a^{2\beta/\alpha}\right)^{\frac{\alpha}{(2\beta+\alpha)}}, \tag{10}$$

which is the same expression as under condition (B) of Theorem 1.

We obtain the parameter set (a, b, α, β) – and thus the value of m^* – for distinct distributional models by developing the tail expansion (1).[11] However, for stochastic processes with unknown tail expansion parameters and for the real-life data in the empirical section, the closed-form expression (9) for AMSE does not exist. Instead, we implement the Beirlant et al. (1999) algorithm that minimizes a sample equivalent of the AMSE.[12] In order to save computation time, we do not determine the optimal nuisance parameter m^* for each recursive, rolling or sequential subsample in (5)–(7) separately. Instead, we determine the full sample estimate for m^* which automatically identifies the full sample scaling constant c in (10) by $\hat{c} = \hat{m}/n^{2/3}$.[13] Extrapolating the optimal path

[8] We use the algorithm by Samorodnitsky and Taqqu (1994) to generate symmetric stable draws.

[9] The Burr distribution is admittedly also not a realistic model for financial return modelling but it enables one to investigate the effects of changing the second order parameter ρ while keeping the tail index constant.

[10] Mikosch and Starica (2000) show that the unconditional distribution of a GARCH(1,1) process with conditionally normal standardized residuals exhibits a heavy tail. They also derive a closed-form relation between the tail index and the parameters of the conditional variance equation. For $\alpha = 4$ (a representative value for the tail index of financial assets in the empirical literature), the closed-form relation specializes to a quadratic equation in the GARCH parameters (β_0, β_1) governing the conditional variance equation. Exact parameter values can be calculated by restricting the parameter sum $\theta = \beta_0 + \beta_1$ to values below 1. All technical details on this closed-form expression are provided in Appendix C ("Calibration of GARCH(1,1) parameters") of the corresponding working paper, see Straetmans and Candelon (2012).

[11] This is the case for the stable, student-*t*, Burr and stochastic volatility models. We include details on the accompanying tail expansion derivations in Appendix B ("Derivation of 2nd order expansion parameters") of the corresponding working paper, see Straetmans and Candelon (2012).

[12] Subsample bootstrap algorithms to select m by means of AMSE minimization constitute an alternative route (see e.g. Danielsson et al., 2001); but these subsample strategies typically require much larger samples than the ones we use and are therefore unsuited for the present analysis.

[13] The exponent 2/3 follows from imposing the restriction $\alpha = \beta$ on the tail expansion parameters. This circumvents the need for estimating β separately. Moreover, simulations convincingly show that the Beirlant criterion still performs well under this restriction even when the true values of β and α differ, see e.g. Beirlant et al. (1999).

for m to the subsamples defined by the stability tests (and using the notation from Section 2), we obtain $\hat{m}_t = \hat{c}t^{2/3}$ for the recursive and sequential tests and $\hat{m}_w = \hat{c}w^{2/3}$ for the rolling test, respectively. 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.

3.3. Monte Carlo results

We first investigate the impact of breaks in the tail index on the finite sample performance of tail index and extreme downside risk measures, i.e. should we care about the detection and presence of breaks when applying EVT techniques that assume stationary tail behavior? Next, we evaluate the finite sample critical values and power of the considered stability tests for the tail index. We also investigate the ability of the tests to locate break dates. To this aim, we simulate from the set of models introduced in the previous section.

Prior to investigating the finite sample performance of the stability tests for the tail index, we consider the finite sample performance of the Hill statistic and a popular quantile estimator that uses the Hill statistic as input. More specifically, we employ the semi-parametric quantile estimator introduced by de Haan et al. (1994):

$$\hat{q}_p = X_{n-m,n}\left(\frac{m}{pn}\right)^{\frac{1}{\hat{\alpha}}}, \qquad (11)$$

and where the "tail cut-off point" $X_{n-m,n}$ is the $(n-m)$th ascending order statistic (or loosely speaking the mth smallest return) from a sample of size n such that $q > X_{n-m,n}$. The quantile $\hat{q}_p$ is interpretable as the daily Value-at-Risk (VaR) at the p% significance level. Financial extremes by definition do not strike often but if they occur they can drive financial institutions into overnight distress and jeopardize overall financial stability. Thus, looking at VaR numbers further into the tail than usual (i.e. corresponding with very low levels of p) is potentially relevant for both risk managers and regulators. As an illustration, consider the problem of allocating upper limits on open positions to foreign currency dealers by the treasurers of the forex dealing room of an international bank.[14] The trading limits depend on the probability p on a single large negative currency return that can bring the bank's solvency in jeopardy. In this example, the level p is interpretable as the insolvency risk the management considers "acceptable". Suppose the management chooses a critical loss level $s < 0$ which stands for the maximum loss that can be incurred without running into solvency problems. A simple way to determine the maximum allowable investment I is to set $I = s/\hat{q}_p$ with $\hat{q}_p$ the extreme quantile estimator as defined in (11).

Turning to the finite sample performance of the above tail index and quantile estimators, we know from Theorem 1 that both estimators ($\hat{\alpha}$ and $\hat{q}_p$) are asymptotically biased under condition (B). A Monte Carlo investigation can clarify to what extent the asymptotic bias transfers into finite sample bias and estimation risk for the Hill statistic and the quantile estimator. Table 1 contains averages, standard errors and Root Mean Squared Errors (RMSEs) for $\hat{\alpha}$ and $\hat{q}_p$. We perform the Monte Carlo experiment for sample sizes of 8000 observations and for 10,000 replications. The VaR's significance level p equals the inverse of the sample size.

Table 1 distinguishes between models that either generate dependent or independent draws (lower and upper panel, respectively). In the upper panel, we let the tail index α and the second order parameter β (or ρ) vary; in the lower panel, we manipulate the degree of serial correlation or volatility clustering (parameter θ) *ceteris paribus* α and β. The outcomes show a large heterogeneity in finite sample bias and estimation accuracy across different distributions. This reflects the differences in the second order behavior of the considered tail models. Notice that in case of pure Pareto-type tail behavior (no second order behavior), the Hill statistics and the corresponding quantile estimates are unbiased.

Biases in the Hill statistic and the estimated quantiles necessarily exhibit opposite signs, see the quantile formula (11). Consistent with Theorem 1, the sign of the bias in the Hill estimator corresponds with the sign of b. Notice also that bias and standard error of the Hill statistic are smaller for heavier tails. This is because lighter tails are closer to a thin tailed local alternative like the normal distribution that does not satisfy (1). This decreases the accuracy of tail estimation techniques that assume regular variation as a starting point. It is also worth noticing what happens when the second order parameter ρ changes for given values of α. The Burr outcomes reveal that the bias and standard error of the Hill estimator decrease for higher values of ρ, i.e. the more the tail expansion (1) approximates a pure Pareto tail the smaller will be the bias and estimation risk. The lower panel of the table illustrates the impact of temporal dependence on bias and variance properties of tail index and quantile estimators. Both higher serial correlation in the AR(1) processes as well as a higher persistence in volatility clustering (Stochastic Volatility and GARCH model class) increase bias and standard error.

Next we add breaks in the tail index to the simulation setup in order to see how this alters the finite sample performance of tail index and extreme quantile estimation as compared to Table 1. Table 2 reports the corresponding outcomes.

For sake of convenience we limit ourselves to those i.i.d. cases where the true quantile is known such that we can calculate the Root Mean Squared Error (RMSE) for the latter estimator. We either assume that tails become fatter (a decrease in α or "thin-to-fat" scenario) or thinner (an increase in α or "fat-to-thin" scenario). Moreover, we vary the within-sample location r of the break date ($r = 0.25$, 0.50 and 0.75). For sake of comparison, we also report the previous table's tail index and quantile estimates without tail breaks. Evidently, the "true" quantile relevant for current risk management and stability assessments is the quantile based on the post-break tail index. This e.g. implies that we have to compare the outcomes for the break scenario $(\alpha_1, \alpha_2) = (4, 2)$ with the no break case of $\alpha = 2$.

The main message of Table 2 is that the finite sample performance deteriorates in the presence of breaks as compared to the situation without breaks. How big the impact of breaks on finite sample performance is crucially depends on the location of the break and the direction of the tail index shifts. Nonsurprisingly, a fat-to-thin shift (thin-to-fat shift) leads to an overestimation (underestimation) of the true tail risk – as measured by the true tail quantile q_p. Moreover, the erosion in finite sample performance due to the presence of breaks is more severe when the break occurs relatively late in the sample. Understandably, this is due to the fact that the bulk of the data used in estimating the full sample tail indices and quantiles do not exhibit the currently relevant tail index. Finally, notice that the impact of tail index breaks on finite sample performance is more severe under the fat-to-thin regime shift. The intuition behind this result is that the Hill statistic and quantile estimation is conditional upon the m largest observations such that outlier behavior from the fat tailed initial sample remains in the selection of the m largest observations in the latter part of the sample. This is not the case under the thin-to-fat regime shift.

[14] See Danielsson and de Vries (1997) for a more elaborate discussion and for other applications of extreme quantile estimation for e.g. institutional investors.

Table 1
Tail index and quantile estimation in the absence of breaks.

DGP	Tail index est.			Quantile est.			True q
	$\hat{\alpha}$	s.e. $(\hat{\alpha})$	RMSE$(\hat{\alpha})$	$\hat{q}$	s.e.$(\hat{q})$	RMSE$(\hat{q})$	
Panel A: i.i.d. models							
Stable(α)							
1.2	1.23	0.06		593.57	155.43		
1.5	1.60	0.15		128.23	38.92		
Student(α)							
2	1.91	0.12	0.15	70.84	13.30	15.32	63.23
4	3.60	0.41	0.60	13.16	1.88	2.06	12.31
Burr(α,$-\rho$)							
(2, −1)	1.94	0.08	0.10	97.22	14.05	16.07	89.44
(2, −5)	1.99	0.03	0.03	91.28	6.04	6.31	89.44
(4, −1)	3.88	0.17	0.21	9.81	0.70	0.78	9.46
(4, −5)	3.98	0.06	0.06	9.54	0.31	0.32	9.46
Panel B: dependence in the first or second moment							
AR(α, θ)							
(1.5, 0.2)	1.61	0.16	0.20	135.11	49.57		
(1.5, 0.4)	1.62	0.19	0.23	146.75	70.23		
SVSTU(α, θ)							
(4, 0.85)	3.56	0.41	0.60	13.19	1.89	2.09	12.31
(4, 0.95)	3.57	0.42	0.60	13.16	1.90	2.08	12.31
GARCH(α,θ)							
(4, 0.85)	3.57	0.41	0.59	7.28	1.46		
(4, 0.95)	3.58	0.50	0.65	7.54	4.08		

Notes: We simulate averages, standard errors and Root Mean Squared Errors (RMSEs) for the tail index and the tail quantile for samples of 8000 draws and for 10,000 replications. The corresponding significance level for the VaR estimation equals the inverse of the sample size. We generate symmetric stable draws using the simulation method of Samorodnitsky and Taqqu (1994): $X_{stable} = \frac{\sin \alpha\gamma}{(\cos\gamma)^{1/\alpha}}\left(\frac{\cos(1-\alpha)\gamma}{W}\right)$, where $0 < \alpha < 2$ represents the tail index. We draw the parameter γ uniformly on $\left[-\frac{\pi}{2};\frac{\pi}{2}\right]$ whereas W is exponentially distributed with mean 1. We obtain Student-*t* draws by using independent standard normal draws N_i ($i = 1,\ldots,v$) in the following way: $X_{student} = N_1(N_2^2 + N_3^2 + \cdots N_v^2)^{-1/2}$ with v the degrees of freedom parameter. We define the Burr df as $F(x) = 1 - (1 + x^{\beta})^{1/\rho}$ where α is the tail index and $\rho = -\beta/\alpha$ the second order parameter. We obtain Burr draws by equating uniform (0;1) draws to $F(x)$ and by solving the expression for x. We use an AR(1) process with nonzero first order serial correlation θ and with symmetric stable innovations (with tail index α) to generate serially correlated data. We simulate a stochastic volatility model based on Student-*t* draws (SVSTU) using the expression $X_{SVSTU} = U_t\sqrt{\frac{v}{\chi^2(v)}}H_t$ with $\chi^2(v)$ the chi-square df with v degrees of freedom, $U_t = \pm 1$ a discrete random variable with state probability of 1/2 and $H_t = \beta Q_t + \theta H_{t-1}$ a persistent AR(1) process where the innovations Q_t originate from a standard normal df. Finally, we also simulate from a GARCH(1,1) model with conditionally normal innovations and where the sum of the GARCH parameters $\theta = \beta_0 + \beta_1$ is chosen such that the tail index of the corresponding unconditional df equals 4.

Table 2
Tail index and quantile estimation when breaks are present.

DGP	Break	Tail Index est.			Quantile est.			True q
	(α_1, α_2)	$\hat{\alpha}$	s.e.$(\hat{\alpha})$	RMSE$(\hat{\alpha})$	$\hat{q}$	s.e.$(\hat{q})$	RMSE$(\hat{q})$	
Panel A: Student								
No break	(2,2)	1.91	0.12	0.15	70.89	13.27	15.32	63.23
$r = 0.25$	(4,2)	1.98	0.12	0.13	57.32	10.24	11.82	63.23
$r = 0.50$	(4,2)	2.14	0.14	0.19	42.37	7.32	22.11	63.23
$r = 0.75$	(4,2)	2.44	0.16	0.47	27.49	4.17	35.98	63.23
No break	(4,4)	3.57	0.42	0.60	13.14	1.89	2.07	12.31
$r = 0.25$	(2,4)	2.45	0.16	1.56	27.46	4.16	15.71	12.31
$r = 0.50$	(2,4)	2.13	0.14	1.87	42.41	7.37	30.99	12.31
$r = 0.75$	(2,4)	1.98	0.13	2.02	57.56	10.41	46.43	12.31
Panel B: Burr ($\rho = -1$)								
No break	(2,2)	1.94	0.08	0.10	96.76	13.81	15.63	89.44
$r = 0.25$	(4,2)	1.95	0.09	0.09	82.82	11.89	13.61	89.35
$r = 0.50$	(4,2)	2.02	0.09	0.09	62.79	8.92	28.09	89.35
$r = 0.75$	(4,2)	2.32	0.11	0.34	34.58	4.50	55.04	89.35
No break	(4,4)	3.88	0.17	0.21	9.83	0.69	0.78	9.46
$r = 0.25$	(2,4)	2.32	0.11	1.68	34.54	4.49	25.49	9.46
$r = 0.50$	(2,4)	2.02	0.09	1.98	62.71	8.83	53.98	9.46
$r = 0.75$	(2,4)	1.95	0.09	2.05	82.66	11.85	74.15	9.46
Panel C: Burr ($\rho = -5$)								
No break	(2,2)	1.990	0.03	0.03	91.24	5.99	6.25	89.44
$r = 0.25$	(4,2)	2.155	0.03	0.16	62.57	3.92	27.15	89.44
$r = 0.50$	(4,2)	2.451	0.04	0.45	37.45	2.14	52.04	89.44
$r = 0.75$	(4,2)	2.977	0.05	0.98	19.84	0.95	69.61	89.44
No break	(4,4)	3.980	0.06	0.07	9.55	0.31	0.33	9.46
$r = 0.25$	(2,4)	2.978	0.05	1.02	19.81	0.94	10.40	9.46
$r = 0.50$	(2,4)	2.451	0.04	1.54	37.41	2.12	28.03	9.46
$r = 0.75$	(2,4)	2.156	0.03	1.84	62.46	3.90	53.14	9.46

Notes: We simulate averages, standard errors and Root Mean Squared Errors (RMSEs) for the tail index and the tail quantile for samples of 8000 draws and for 10,000 replications. The corresponding significance level for the VaR estimation equals the inverse of the sample size. We represent the location of the breaks by r.

The asymptotic distribution of the considered stability tests crucially depends on the asymptotic behavior of the underlying Hill statistic described in Theorem 1. Whereas condition (A) renders one set of critical values that only applies under pure Pareto-type tail behavior, condition (B) implies that the parameter $\varphi = sign(b)(2\beta/\alpha)^{-1/2}$ determining the asymptotic bias of the Hill estimator also enters the asymptotic critical values. In this section we investigate the impact of this asymptotic bias term on the critical values, power and ability to date breaks in finite samples. In theory, it is possible to calculate the biased as well as the bias-corrected asymptotic critical values with great precision provided one knows the parameter φ. As to date, however, there are no estimators for φ that exhibit a satisfactory finite sample performance that is robust across the main types of regularly varying tail models, see e.g. Danielsson and de Vries (1997) or Gomes et al. (2003).[15] Moreover, whereas the asymptotic distributions across different DGP's only differ due to differences in φ, the finite sample critical values, power and break date estimates may also depend on the sample size n and the optimal value for m. Otherwise stated, we would like to know to what extent the bias of the Hill estimator influences the finite sample size and power properties of the stability tests as well as their ability to accurately identify break dates.

Tables 3 and 4 report simulated critical values for i.i.d. models and models that exhibit temporal dependence, respectively. We split each table in three panels for the recursive, rolling and sequential tests in (5)–(7). We calculate the quantiles of the test statistics for two different sample sizes and generate 20,000 Monte Carlo replications for the considered DGP's. We use these Monte Carlo replications to obtain estimates of the 90th, 95th and 99th percentile of the stability tests' finite sample distribution.

The heterogeneity in the finite sample critical values across different DGP's is comparable with the preceding tables on bias and estimation risk for the Hill and quantile estimators. This illustrates the fact that the DGP's under consideration deviate from pure Pareto tail behavior and also exhibit very different 2nd order tail behavior. Critical values and their estimation risk are higher for those cases that exhibit a stronger 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/estimation risk whereas higher values of the second order parameter ρ (cf. Burr df) decrease the asymptotically unbiased critical values reported in Quintos et al. (2001). This should not surprise given the fact that the Burr tail comes close to a Pareto tail for $\rho = -5$ and that the Hill statistic is asymptotically unbiased for pure Pareto data. But Tables 3 and 4 also illustrate that using asymptotically unbiased critical values lead to a huge overrejection of the null of parameter constancy.

Next, Tables 5 and 6 report finite sample power and estimates of the breakpoints for the recursive, rolling and sequential stability test, respectively. We consider sudden upwards and downwards jumps in α of different magnitudes and at different points in time r. We perform finite sample power calculations and breakpoint estimates for 20,000 replications and conditional on the finite sample critical values from Tables 3 and 4.

The direction of change in α seems to be crucial for the finite sample power and ability to date breaks. 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. One can understand the latter outcome by observing that (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 increases. This initial extremes dominance when $\alpha_1 < \alpha_2$ does not occur for the rolling test since the impact on the Hill estimator 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 clearly outperforms the other two tests for most considered DGP's provided the break scenario implies an increase in tail fatness ($\alpha_1 > \alpha_2$).[16] However, we can easily resolve the lack of power for one type of α-jump by performing the test both in calendar time ("forward" recursive test) as well as by inverting the sample ("backward" recursive test). The forward (backward) version of the recursive test then signals falls (rises) in α. This is the strategy implemented in the empirical application.

Sofar the general discussion on power and break date ability. Notice that there are also large differences in power results and break point detection across different DGP's. The determinants of the bias in the Hill estimator may again be held responsible for this heterogeneity. More specifically, higher values of the persistence parameter θ (either standing for serial correlation or volatility persistence) increase the bias in the Hill estimator and the bias in the estimated break dates but decrease the power. On the other hand, higher values of the second order parameter ρ (cf. Burr df) decrease the bias in the Hill estimator and the bias in the estimated break dates but increase the power. Thus, Tables 5 and 6 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 relatively small samples whereas bias and estimation risk for the break date estimates are negligibly small.

4. Empirical results

We want to assess whether the propensity towards financial extremes changes over time for different asset classes in developed and emerging markets. For that purpose we use a bootstrap-based recursive version of the Quintos et al. (2001) stability test. The recursive test outperforms the rolling and sequential tests in terms of finite sample power and ability to date breaks. We therefore limit ourselves to using the recursive test in the empirical application.

It is well-known that standard regression-based risk proxies like standard errors, CAPM-βs or factor model loadings are not constant over time, see e.g. Ross et al. (2005). We like to know whether and to what extent this instability in traditional risk measures transfers to unconditional tail risk measures like e.g. the tail index or tail quantiles evaluated far into the distributional tail. The riskiness of assets may also differ considerably across asset classes and/or regions (e.g. developed vs. emerging markets). Our empirical investigation therefore encompasses a large cross section of different asset types (stocks, bonds, commodities, foreign exchange, gold, silver and oil). We extract data from Thomson Datastream and express financial returns as log price differences between daily

[15] Quintos et al. (2001, p. 639) also propose a bias correction procedure under the restriction that $sign(b) = 1$. Although this sign restriction holds for the class of stable dfs, b is negative for a majority of regularly varying models. The parameter b can be positive or negative for real data sets which implies that the bias correction procedure of Quintos et al. (2001) increases the bias if $sign(b) = -1$.

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

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

DGP	$n = 500$			$n = 2000$		
	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 (γ = 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)

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

closes. We express developed and emerging stock and bond indices in local currency. The exchange rate data also consist of a developed and emerging currency block. We express all currencies against the US dollar. Finally, prices of oil (Brent Crude), gold and silver are in US$ per barrel or per troy ounce, respectively. We let all series end on 31 December 2009 but they exhibit different starting points depending on data availability.[17]

Large institutional investors like pension funds in search of fresh diversification opportunities invest in a growing variety of asset classes and geographic regions and would like to know whether and to what extent these different asset tail risks change over time. Moreover, we distinguish between stock indices and foreign exchange rates of developed and emerging markets. Given the fact that large institutional investors increasingly invest in emerging markets, it is important to assess whether emerging market assets exhibit relatively more frequent shifts in tail behavior. One may expect this to be the case due to a less stable political and institutional environment as compared to more developed countries.[18]

The simulation section illustrates that several forms of temporal dependence bias the recursive test's finite sample critical values. Upon assuming GARCH-type volatility clustering as the main source of temporal dependence, we implement a GARCH-corrected version of the recursive test:

$$Q = \sup_{r \in R_\tau} \hat{\eta}_t^{-1} Y_n^2(t), \tag{12}$$

where $\hat{\eta}_t$ is the estimate of the time varying scaling factor, see Quintos et al. (2001, p. 643). The extensive Monte Carlo simulations in the previous section convincingly show that the test's finite sample distribution depends on the parameters of the regularly varying tail. In other words, there is no single set of critical values that applies to all return series simultaneously. As a solution to this problem, we use bootstrap-based critical values at the 95% and 99% levels to

[17] We abbreviate developed and emerging stock and bond indices as follows: France (FR), Germany (GE), United Kingdom (UK), United States (US), Japan (JP), Indonesia (INDO), Malaysia (MAL), Thailand (THAI), Mexico (MEX), Chile (CHIL). We exclude emerging bond index data because of insufficient data availability. We consider dividend-adjusted stock indices and 10-year benchmark government bonds. The industrial currency block covers the euro (EUR, and before January 1999 the Deutsche mark), the Japanese yen (JPY), the Pound sterling (GBP), the Swiss franc (CHF) and the Canadian dollar (CND). The emerging currency block includes the Indonesian rupiah (IRD), the Malaysian ringgit (MYR), the Thai baht (THB), the Chilean peso (CLP) and the Mexican peso (MXN). Developed/emerging stock market series and developed/emerging currencies start from 1 January 1973, 2 April 1990, 3 January 1972 and 3 January 1994, respectively. Bond series start in the first half of the 80s: United States, United Kingdom and Germany (1 January 1980), Japan (2 January 1984), France (2 January 1985). Finally, commodities start on 3 January 1972 (gold and silver) and 4 January 1982 (oil), respectively.

[18] A logical multivariate extension would be to investigate whether and to what extent the cross-asset tail dependence shifts over time. Tail dependence can be identified either by means of copulae evaluated in the tail area or by a tail index of an auxiliary variable that summarizes the dependence structure in the tails, see e.g. Hartmann et al. (2006) for an earlier application on the multivariate dependence structure of bank stocks. Straetmans et al. (2008) investigate whether the 9/11 terrorist attacks had a significant impact on the tail dependence between US sectoral indices and the market as a whole.

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

DGP	$n = 500$			$n = 2000$		
	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)						
(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)
(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)

Notes: We report critical values for varying sample sizes n, and different levels of significance. We base critical values on 20,000 Monte Carlo replications. We report corresponding standard errors for the critical values between brackets (s.e.). We denote the first order serial correlation of an autoregressive process with stable innovations (ARSTA), the volatility persistence parameter in GARCH(1,1) models and stochastic volatility models with Student-t innovations (SVSTU) by the parameter θ.

determine finite sample critical values for each financial series separately. As the scaling factor in (12) already corrects the test for any temporal dependence, the bootstrap no longer has to take care of any temporal dependence and we can resort to a "wild" version of the bootstrap instead of a block bootstrap. 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.

Table 7 reports stability test results for a large variety of stock indices, bond indices, exchange rates and commodities. Furthermore, we distinguish between mature and emerging stock markets and currencies. Table 7 also contains tail index and extreme quantile estimates using the estimators earlier introduced in (2) and (11).[19] We calculate the VaR quantiles for a marginal significance level p of 0.015% which implies that we expect the corresponding extreme events to happen once every 6500 days (this amounts to once every $6500/260 \approx 25$ years).[20] We use the Beirlant et al. (1999) method to determine the optimal number of upper order extremes m^* used in estimating the test statistic, the bootstrap-based critical values and the extreme quantiles. As for the maximum values for the forward and backward version of test (12) we include these in the columns labelled Q_F and Q_B, respectively. Evidently, bootstrapped critical values are identical for the forward and backward test. We reject the null of parameter constancy if the sup-value calculated according to (12) exceeds the bootstrap-based critical value, e.g. $Q > CV_B(p)$ with p = 5% or 1%. We report statistically significant break dates between brackets beneath the testing values (dd/mm/yy).

First and foremost, we observe that the overall number of tail index breaks remains limited. We only reject the null hypothesis of a constant tail index at the 1% level for 6 out of 28 assets. We detect the majority of breaks in emerging currency markets. The emerging currency tail break dates confirm earlier research by Candelon and Straetmans (2006) for a shorter sample of Asian currencies. Both the forward breaks (tail index drops) and backward breaks (tail index rises) fall within the time window of the Asian financial crisis. Moreover, the forward breaks precede the backward breaks which suggests a U-shaped pattern for the tail index (and an inverted U-shape pattern for the quantile). As a result of unsustainable speculative pressures all considered Asian currency regimes were abolished during the second half of 1997. The Central Banks of Thailand, Malaysia and Indonesia all announced a managed float in the first half of July. For Thailand, the estimated forward break nearly perfectly coincides with that regime shift but the forward breaks for the other countries seem to lag behind the managed float announcement for approximately 6 months.[21]

[19] In case of two breaks $t_1 < t_2$ we condition the pre-break and post-break estimates on the subsamples $[1, t_1]$ and $[t_2, n]$ respectively.

[20] We assume that there are 260 trading days in a calendar year.

[21] This break date delay may be due to the bias in the Hill estimator that may be more severe than for other assets.

We only find a backward break (tail index rise) for the Mexican Peso which can be interpreted as evidence for the success of the rescue and stabilization policies set up by the Mexican government and the IMF in the aftermath of the December 1994 devaluation of the Mexican Peso against the US$.[22] Finally, the US stock market (stock market crash of 1987) and the UK bond market (expansive monetary policy after the 1987 stock market crash) exhibit a significant tail index shift. All other stock and bond market tails seem to exhibit stationary tail behavior.

We also report pre-break and post-break tail index and quantile values for the statistically significant breaks in the tail index at the 1% level. The bulk of these extreme quantile shifts is situated in emerging currency markets. Notice that the post-break emerging tail risk drops dramatically compared to its pre-break value (except for the Thai currency). This suggests that the backward breaks (rise in the tail index) more than offset the preceding forward breaks (drops in the tail index). The stronger statistical significance (higher sup-values) for the backward breaks point in that direction indeed.

Turning to the extreme VaR estimates in the right part of Table 7 one first of all observes the huge cross asset differences in extreme downside risk. Exchange rate regimes in emerging markets seem completely ineffective in dampening exchange rate volatility as the VaR estimates are much larger for emerging currency regimes as compared to floating (developed) currency markets. Thus attempts towards exchange rate stabilization in emerging markets seem counterproductive over the considered sample period, see also Koedijk et al. (1992) for earlier evidence. Legal and institutional restrictions in emerging stock markets also do not help to curb emerging equity volatility although it does not seem to have such adverse effects as in forex markets. As a matter of fact, emerging and developed stock market tails seem to exhibit extreme downside risk of comparable magnitudes. Finally, mature stock markets exhibit more extreme downside risk than mature bond markets, see also de Haan et al. (1994) or Hartmann et al. (2004).

In order to better grasp the implications of non-constant downside risk for risk management, reconsider the earlier discussed EVT application of allocating upper limits on open positions in the forex dealing room of an international bank. We argue that the maximum allowable investment for forex traders in an open position equals $I = s/\hat{q}_p$ for a given critical loss level $s < 0$ and with $\hat{q}_p$ the extreme quantile estimator as defined in (11). Given the break results for emerging currencies, full sample trading limits are set too conservatively because they do not take into account the "thinning" of the tails due to e.g. subsequent liberalizations and abolishment of exchange rate regimes. Thus, given the unstable tail behavior of emerging currency returns, it is advisable to use shorter data windows that started more recently (ideally after a break has occurred in the tail behavior) for determining trading limits on currency positions. For example, consider a US bank trader that wants to build up a position in Mexican Peso. The superior management determines that the maximum loss the bank can suffer without running into solvency problems is 10,000,000US$. If the trader determines his trading limit using the full sample of Mexican Peso/US$ quotes, the relevant full sample 0.015% VaR quantile in Table 7 reads 30.28%. This, in turn, induces a trading limit of I_{full} = 10,000,000US$/0.3028 = 33,025,099US$. However, using the post-break 0.015% VaR quantile renders a trading limit of I_{post} = 10,000,000US$/0.1109 = 90,171,326US$ which is nearly three times as high. In other words, taking into account that the extreme risk of the currency has diminished over the sample period due e.g. to regime changes also renders the maximum allowable investment less conservative. One can easily generalize this stylized example for one currency to a portfolio of currencies but the principle remains the same.

The results on the currency tails are also interesting from a policy perspective. Although many countries exhibit a fear of floating, fixing the exchange rate in one way or another by e.g. capital

Table 5
Size-corrected finite sample power for recursive, rolling and sequential tests.

DGP(α_1, α_2)	n = 500			n = 2000		
	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)						
θ = 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)						
θ = 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

Notes: We report the power 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. We denote the volatility persistence parameter in the stochastic volatility models with Student-t innovations (SVSTU) by θ.

[22] The stability test does not detect this latter devaluation because the test's interior region $R_\varepsilon = [0.15n, 0.85n]$ does not contain December 1994. The same applies for potential breaks due to the 2007–2010 credit crunch. If present, they only become identifiable when the sample gets longer and the candidate break dates fall within the interior region. However, even if breaks are inside the interior region, they are more difficult to detect when they lie close to the interior region boundaries. This is because the recursive stability test's finite sample power decreases for breaks that lie close to the interior sample boundaries, see our own simulation section results or Candelon and Lütkepohl (2001).

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

DGP (α_1, α_2)	n = 500			n = 2000		
	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)						
θ = 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)						
θ = 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)

Notes: We report estimated break dates for different sample sizes (n = 500 or 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. We calculate "candidate" break dates over 20,000 Monte Carlo replications. We obtain average break date estimates 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. We denote the volatility persistence parameter in stochastic volatility models with Student-t innovations (SVSTU) by θ.

controls or other restrictions on current account and capital account convertibility seems counterproductive. The downside risk estimates, as measured by the extreme quantile q_p, are nearly always higher in the presence of regimes. The economic interpretation is that a float lets exchange rates adjust more smoothly than any other regime that involves some fixity. Monetary history indeed shows that it is extremely difficult for monetary authorities to establish and sustain perfectly credible and time-consistent forex regimes. Imperfectly credible capital controls, however, are to speculators like a red rag to a bull. The inverse relation between extreme depreciation risk and the abolishment of capital controls seems to support Friedman's old plea for flexible exchange rates, see e.g. Friedman (1953).

With an eye toward some sensitivity analysis, we also apply the recursive testing procedure on a few economically meaningful subsamples around crisis episodes like the dotcom bubble or the 1987 stock market crash. The series that do not exhibit full sample breaks are not characterized by subsample breaks either as one would expect. Evidently, subsample breaks should not necessarily be identical to the full sample breaks if the latter fall outside the selected subsample. However, when the full sample breaks do fall in the selected subsample, break results are robust and the location of the break is only marginally altered by the change in sample size.

Overall, the empirical results suggest that heavy tails and corresponding extreme quantiles are remarkably stable over long

Table 7
Forward and backward recursive testing outcomes.

Assets	m^*	Recursive test		Crit.values		Tail index and VaR (%)		
		Q_F	Q_B	0.95	0.99	$\hat{q}(\hat{\alpha})$	$\hat{q}_1(\hat{\alpha}_1)$	$\hat{q}_2(\hat{\alpha}_2)$
Panel A: stock markets (local currency)								
US	150	5.22*** **(7/9/87)**	0.63	2.78	3.81	11.38 (3.00)	5.25 (4.61)	15.10 (2.72)
UK	166	0.52	0.73	2.45	3.43	10.98 (3.19)		
FR	142	1.21	1.27	3.44	4.71	10.45 (3.65)		
GE	100	1.19	0.46	2.38	3.75	10.16 (3.49)		
JP	106	1.28	0.22	2.49	4.13	10.81 (3.43)		
INDO	128	1.06	0.96	3.10	4.49	23.49 (2.71)		
MAL	105	0.41	0.66	2.87	3.95	17.16 (2.74)		
THAI	193	1.33	0.49	2.84	3.93	19.19 (3.06)		
MEX	80	1.69	2.12	4.03	5.36	9.24 (4.57)		
CHIL	96	0.17	0.36	2.13	3.28	8.56 (3.40)		
Panel B: bond markets (local currency)								
US	129	0.48	0.60	3.19	4.15	3.60 (4.33)		
UK	350	1.46	7.44*** **(20/10/87)**	2.96	4.06	6.44 (2.57)	17.52 (1.89)	4.23 (3.03)
FR	110	0.52	0.39	2.94	4.02	2.82 (3.94)		
GE	81	1.34	0.95	2.81	4.26	2.25 (4.70)		
JP	166	1.81	2.56	3.20	4.33	4.79 (2.58)		
Panel C: currencies (w.r.t. US$)								
GBP	119	1.20	1.18	2.08	3.20	5.19 (3.44)		
EUR	148	1.65	0.41	2.41	3.43	4.27 (4.25)		
CAN	174	0.61	0.14	3.29	4.47	3.98 (3.33)		
JPY	155	1.43	0.37	3.02	4.08	4.27 (4.38)		
CHF	311	2.86	1.70	3.80	5.01	5.61 (3.72)		
MXN	250	2.53	6.24*** **(9/10/98)**	2.57	4.010	30.28 (1.71)	280.9 (1.07)	11.09 (2.31)
CLP	147	0.75	0.85	3.72	5.08	6.70 (2.94)		
IDR	267	6.89*** **(24/4/98)**	8.2*** **(3/2/99)**	3.29	4.38	89.41 (1.40)	216.96 (0.58)	26.60 (1.84)
MYR	51	5.97** **(2/1/98)**	11.16*** **(9/9/99)**	4.31	6.21	12.09 (2.16)	5.16 (4.51)	2.21 (3.73)
THB	100	6.48*** **(15/5/97)**	7.09*** **(19/5/98)**	2.25	3.96	19.55 (1.72)	0.88 (3.42)	6.07 (2.58)
Panel D: commodities								
Gold	203	1.92	2.01	2.44	3.53	17.36 (2.74)		
Silver	307	2.47	2.20	2.61	3.81	36.69 (2.35)		
Oil	150	1.23	0.96	3.45	4.56	23.41 (3.17)		

Notes: Country and currency abbreviations stand for: US (United States), UK (United Kingdom), FR (France), GE (Germany), JP (Japan), INDO (Indonesia), MAL (Malaysia), THAI (Thailand), MEX (Mexico), CHIL (Chile), GBP (British pound), EUR (euro), CAN (Canadian dollar), JPY (Japanese yen), CHF (Swiss franc), MXN (Mexican peso), CLP (Chilean peso), IDR (Indonesian rupiah), MYR (Malaysian ringgit), THB (Thai baht). We denote the forward and backward version of the recursive test by Q_F and Q_B, respectively. We base critical values on 10,000 bootstrapped sample replications.
We report the break dates (dd/mm/yy) of corresponding significant breaks in bold. In case of significant breaks in the tail index, we report Value-at-Risk (VaR) estimates for the full sample and the subsamples determined by the break.
** We denote statistically significant rejections of the null hypothesis of tail index constancy at the 5% significance level.
*** We denote statistically significant rejections of the null hypothesis of tail index constancy at the 1% significance level.

periods of time for most of the considered assets and asset classes. Tail index and extreme quantile estimation seem to be useful tools for assessing long-term tail risk, stress testing and financial stability but one has to apply these techniques with care in the presence of breaks in the tail behavior. Hedging tail risk of portfolios containing large positions in emerging currencies constitutes an example.

5. Conclusions

This paper provides a thorough study of the finite 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. Conform to the bulk of the empirical literature, we select the number of upper order extremes by minimizing the sample Mean Squared Error of the Hill statistic. We choose the DGP's such as to mimic some popular empirical stylized facts of financial data. The finite sample critical values, the (size-corrected) power and the ability to date breaks differ a lot across different distributional models and sample sizes. Nonsurprisingly, our simulation experiments show that critical values increase and the power and break date ability decrease when the bias in the Hill estimator becomes more severe. As there are no satisfactory bias reduction methods available for the finite sample critical values, we propose a bootstrap-based procedure for the critical values of the stability test when working with real-life data. We implement a recursive version of the stability test in the empirical application as this version outperforms its rolling and sequential counterparts in a simulation environment. Upon applying the stability test to a large set of asset classes, both in developed as well as emerging markets, we hardly detect any breaks at all, except for emerging currency tails. For those series with breaks in the tail behavior, it is advisable to base tail risk measures like VaR or expected shortfall on the post-break sample.

Acknowledgements

We are indebted to the editor of the Journal of Banking and Finance, Ike Mathur, and an anonymous referee for their constructive comments and suggestions. We also received constructive feedback from seminar participants at the Université d'Orléans, Bath University, Erasmus University Rotterdam and from conference participants at INFINITY Dublin 2011.

References

Andersen, T.G., Bollerslev, T., 1997. Intraday periodicity and volatility persistence in financial markets. Journal of Empirical Finance 4, 115–158.
Andrews, D., 1993. Tests for parameter stability and structural change with unknown change point. Econometrica 59, 817–858.
Beirlant, J., Dierckx, G., Goegebeur, Y., Matthys, G., 1999. Tail index estimation and an exponential regression model. Extremes 2, 177–200.
Candelon, B., Lütkepohl, H., 2001. On the reliability of Chow type tests for parameter constancy in multivariate dynamic models. Economics Letters 73, 155–160.
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.
Danielsson, J., de Vries, C.G., 1997. Tail index and quantile estimation with very high frequency data. Journal of Empirical Finance 4, 241–257.
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.
de Haan, L., Stadtmüller, U., 1996. Generalized regular variation of second order. Journal of the Australian Mathematical Society (Series A) 61, 381–395.
de Haan, L., 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., Lechner, J., Simiu, E. (Eds.), Proceedings from a Conference on Extreme Value Theory and Applications. Kluwer Academic Press, Dordrecht.
Dumouchel, W.H., 1983. Estimating the stable index α in order to measure tail thickness: a critique. Annals of Statistics 11, 1019–1031.
Embrechts, P., Klüppelberg, C., Mikosch, T., 1997. Modelling Extremal Events. Springer-Verlag, Berlin.
Feller, W., 1971a. An Introduction to Probability Theory and its Applications, vol. I. John Wiley and Sons, Chicago.
Feller, W., 1971b. An Introduction to Probability Theory and its Applications, vol. II. John Wiley and Sons, Chicago.
Friedman, M., 1953. Essays in Positive Economics. University of Chicago Press, Chicago.
Galbraith, J.W., Zernov, S., 2004. Circuit breakers and the tail index of equity returns. Journal of Financial Econometrics 2, 109–129.
Goldie, C.M., Smith, R., 1987. Slow variation with remainder: theory and applications. Quarterly Journal of Mathematics 38, 45–71.
Gomes, M.I., de Haan, L., Peng, L., 2003. Semi-parametric estimation of the second order parameter in statistics of extremes. Extremes 5, 387–414.
Haeusler, E., Teugels, J., 1985. On asymptotic normality of Hill's estimator for the exponent of regular variation. Annals of Statistics 13, 743–756.
Hall, P., 1982. On some simple estimates of an exponent of regular variation. Journal of the Royal Statistical Society (Series B) 42, 37–42.
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.
Hartmann, P., Straetmans, S., de Vries, C.G., 2004. Asset market linkages in crisis periods. Review of Economics and Statistics 86, 313–326.
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.
Hill, B.M., 1975. A simple general approach to inference about the tail of a distribution. The Annals of Statistics 3, 1163–1173.
Hols, M.C.A.B., de Vries, C.G., 1991. The limiting distribution of extremal exchange rate returns. Journal of Applied Econometrics 6, 287–302.
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.
Jansen, D.W., Koedijk, K.G., de Vries, C.G., 2000. Portfolio selection with limited downside risk. Journal of Empirical Finance 7, 247–269.
Jondeau, E., Rockinger, M., 2003. Testing for differences in the tails of stock market returns. Journal of Empirical Finance 10, 559–581.
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.
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.
Leadbetter, M.R., Lindgren, G., Rootzén, H., 1983. Extremes and Related Properties of Random Sequences and Processes. Springer-Verlag, Berlin.
Longin, F.M., 1996. The asymptotic distribution of extreme stock market returns. Journal of Business 69, 383–408.
Mandelbrot, B., 1963. The variation of certain speculative prices. Journal of Business 36, 394–419.
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.
Pagan, A.R., Schwert, G.W., 1990. Testing for covariance stationarity in stock market data. Economics Letters 33, 165–170.
Quandt, R., 1960. Test of the hypothesis that a linear regression obeys two separate regimes. Journal of the American Statistical Association 55, 324–330.
Quintos, C., Fan, Z., Phillips, P., 2001. Structural change tests in tail behaviour and the Asian crisis. Review of Economic Studies 68, 633–663.
Ross, S.A., Westerfield, R.W., Jaffe, J., 2005. Corporate Finance. McGraw-Hill, New York.
Samorodnitsky, G., Taqqu, M., 1994. Stable Non-Gaussian Random Processes. Chapman and Hall, New York.
Straetmans, S., Candelon, B., 2012. Fats Tails in Small Samples. Working Paper, Maastricht University.
Straetmans, S., Verschoor, W., Wolff, C., 2008. Extreme US stock market fluctuations in the wake of 9/11. Journal of Applied Econometrics 23, 17–42.
Werner, T., Upper, C., 2002. Time Variation in the Tail Behaviour of Bund Futures Returns. Working Paper, European Central Bank.