Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Empirical Finance 10 (2003) 559–581
Journal of EMPIRICAL FINANCE
www.elsevier.com/locate/econbase

Testing for differences in the tails of stock-market returns

Eric Jondeau[a,*], Michael Rockinger[b,1]

[a] Banque de France, Centre de recherche 41-1391, 31 rue Croix des Petits Champs Paris 75049, France and ERUDITE, Université Paris 12 Val-de-Marne, Paris 75049, France
[b] HEC Lausanne, Institute of Banking and Financial Management, CH 1015 Lausanne-Dorigny, Switzerland

Received 30 June 1999; accepted 16 October 2002

Abstract

In this paper, we use a database consisting of daily stock-market returns for 20 countries to test for similarities between the left and right tails of returns, as well as across countries. We estimate and test using the distribution of extreme returns over subsamples approach. Via Monte-Carlo simulations, we show that maximum-likelihood estimators are essentially unbiased, provided the size of subsamples is correctly chosen, and that the likelihood-ratio tests on parameters characterizing the behavior of extremes are correctly sized. For actual returns, we find that left and right tails behave very similarly. Across countries, we find that extremes are located at different levels and that their dispersion varies. The tail index, characterizing large extreme realizations, is found to be constant within each geographical group. We verify that the perception that left tails are heavier than right ones is not due to clustering of extremes. The failure to detect statistical significant differences is likely to be due to the relative infrequency of large extremes.
© 2003 Elsevier Science B.V. All rights reserved.

JEL classification: C13; C22; G15; O16
Keywords: Extreme value theory; Generalized extreme value distribution; Emerging markets; Stock-market return

* Corresponding author. Tel.: +33-1-42-92-49-89.
E-mail addresses: eric.jondeau@banque-france.fr (E. Jondeau), Michael.Rockinger@hec.unil.ch (M. Rockinger).
[1] HEC Lausanne, FAME, CEPR, and scientific consultant at the Banque de France.

0927-5398/03/$ - see front matter © 2003 Elsevier Science B.V. All rights reserved.
doi:10.1016/S0927-5398(03)00005-7

1. Introduction

The study of extreme returns has attracted a strong academic interest since, in particular, extreme drops of stock markets may have important consequences for the portfolios of institutional investors. Several academic contributions either focus on the left tail of the distribution of returns, with the purpose of gaining a better understanding for value-at-risk applications, or present estimates of the left and right tails separately, without formally testing whether the extreme realizations in either tail are similar.

Most investors, when asked whether the left tail of returns is heavier than the right one, would spontaneously answer with an affirmative. A first explanation why the left tail could be heavier than the right tail is that stock markets may be subject to bubbles. As bubbles burst, a strong negative return would lead to a market correction. A second explanation could be obtained by reinterpreting the arguments of Campbell and Hentschel (1992). They argue that returns are moved by news. News generate volatility and cluster, in that news are often followed by additional news. They state that a piece of good news increases stock prices, yet some of this increase gets dampened by the increase in risk premium requested for the higher volatility. On the other hand, a piece of bad news lowers stock prices. This drop gets further amplified by the increase in the risk premium. Because of clustering of news, the left tail of returns could therefore be thicker than the right tail.

In this paper, we use elements drawn from extreme value theory (evt) to formally show, for a large set of mature and emerging stock markets, that the left tail of returns is similar to the right tail. We will also show that volatility clustering cannot be hold responsible for this subjective impression that the left tail is heavier than the right tail.

The behavior of extremes is often summarized by the so-called tail index. This index measures the fat-tailedness of the distribution of a return series. Non-parametric techniques, such as the estimators proposed by Pickands (1975), or Hill (1975), are widely used for computing the tail index. Based on high-order statistics, these estimators are very easy to implement. They have been used in a number of empirical studies, such as Koedijk et al. (1990), Jansen and de Vries (1991), Hols and de Vries (1991), and Koedijk and Kool (1992). These techniques belong to the more general tail approach, which considers the exceedances over a high threshold. In this context, some parametric techniques allow to fit the distribution of the tails using a generalized Pareto distribution (gpd). Applications in finance are by McNeil (1996), and McNeil and Saladin (1997). The tail approach, however, has two important drawbacks. First, the small-sample properties of the estimators are likely to be affected by the choice of the threshold. Several papers have proposed methods to select, in an optimal manner, the high threshold (Danielsson and de Vries, 1997; Huisman et al., 2001). Second, tail-index estimators are likely to be severely biased when the series is not iid. When the return series is a dependent process (for instance, a GARCH process), the asymptotic properties of the non-parametric estimators are not clearly established. In the context of the estimation of a gpd for the distribution of tails, McNeil and Frey (2000) have proposed to focus on the conditional behavior of the tails, once the series has been filtered by a GARCH process.

There exists an alternative approach to the tail index that partially avoids these two drawbacks. This approach does not consider the distribution of the tails but, rather, the distribution of the maximum or minimum returns over given subsamples. This technique has been applied to financial data by Longin (1996) and McNeil (1998). The main advantage of this approach is that it does not assume that returns are independent. In the case of dependent, but strictly stationary, processes, the limit distribution of extremes is still a *generalized extreme value* (gev) distribution, but an additional parameter, namely the extremal index, has to be included in the model. This parameter measures the extent to which dependency of returns affects the extremal behavior of the process. Parameters of the gev may be easily estimated with ML. Therefore, estimates of the tail index are likely to be unbiased and the standard likelihood-ratio test (LRT) applies. In a Monte-Carlo simulation, we will verify that the estimates of the tail index are unbiased and that, for well chosen subsample size, also the rejection frequency of the LRT is correct.

Applications of the evt to stock-market returns are abundant. While a few papers focused on low-frequency data (Longin and Solnik, 2001; Hartmann et al., 2001), most studies considered daily data (e.g. Jansen and de Vries, 1991; Loretan and Phillips, 1994; Longin, 1996). Lux (1998) investigated tick-by-tick data of the German DAX stock index. Very few papers, however, focus on the tail behavior of returns in emerging markets. In that field, Claessens et al. (1995) considered the existence of return anomalies and predictability for a set of 20 emerging markets, while Bekaert and Harvey (1997), using the same data, studied the determinants of volatility in emerging markets.[2] The only paper, to our knowledge, which explicitly investigated the behavior of extreme returns in emerging markets is the paper by Quintos et al. (2001). These authors considered tests of structural change in the behavior of extreme returns on three Asian stock markets.

From an empirical viewpoint, a first goal of this paper is to extend the investigation of the behavior of extreme returns to a large set of stock markets sampled at daily frequency. We apply evt techniques to a set of 20 stock-market indices covering mature markets, Asian markets, Eastern European markets, as well as Latin American ones. Second, we address specific issues concerning the behavior of extreme returns. On one hand, we test whether the left and the right tails have similar characteristic parameters. In particular, in emerging equity markets, it is not clear whether crashes are more likely to occur than booms. On the other hand, we test whether characteristic parameters of tails are identical across markets. More specifically, we address the issue whether extreme returns behave similarly in a same geographical group. Finally, we question whether clustering of extremes explains the impression that left tails are heavier than right ones. This leads us to test whether the extremal index is similar on both sides of the distribution.

The structure of the paper is as follows. In Section 2, we briefly recall elements of extreme value theory. We introduce notations and tools used in the paper, and we focus on the case of dependent series. In Section 3, we describe the data, and provide descriptive statistics on the pattern of the distribution of market returns. Section 4 is devoted to the empirical results. We first present simulation evidence on the small-sample properties of the ML estimates of the gev parameters and on the LRT statistics. Then, we report estimation of the gev distribution, and we present the results of the tests of the aforementioned hypotheses. In Section 5, we discuss our empirical evidence and conclude.

[2] See also Rockinger and Urga (2001), who focused on Eastern European markets.

2. Elements of theory for extremes

Several textbooks deal with evt, as those by Leadbetter et al. (1983), Embrechts et al. (1997), or Reiss and Thomas (1997). Therefore, we focus in this section only on features useful for our study.

2.1. The case of independent and identically distributed series

Let $\{X_1,\cdots,X_N\}$ be a sequence of N iid random variables representing daily observations of the return on a stock-market index, with cumulative distribution function F_X. We let $M_N = \max(X_1,\cdots,X_N)$ denote the maximum over the sample period.[3] The Fisher–Tippett theorem, formally proved by Gnedenko (1943), gives the limit distribution of standardized extremes. Suppose that there exist a *location parameter*, μ_N, and a *scale parameter*, $\psi_N > 0$, such that the limit distribution of standardized extremes, $(M_N - \mu_N)/\psi_N$, is non-degenerate, with:

$$\lim_{N\to\infty} Pr\left[\frac{M_N - \mu_N}{\psi_N} \leq y\right] = \lim_{N\to\infty} F_X(M_N)^N = H(y). \tag{1}$$

Then, the limit distribution, denoted H, can be of three types whose general expression is given by the gev distribution, defined as

$$H\xi(y) = \begin{cases} \exp(-(1+\xi y)^{-1/\xi}) & \text{if } 1+\xi y > 0 \text{ and } \xi \neq 0, \\ \exp(-\exp(-y)) & \text{if } \xi = 0, \end{cases} \tag{2}$$

where the parameter ξ is the *tail index*. Among the three limit distributions, the one obtained for $\xi > 0$, called the Fréchet distribution, corresponds to fat-tailed processes, such as Cauchy, Stable, or Student t random variables, as well as GARCH processes. A large body of the empirical finance literature documents that asset returns are fat-tailed, suggesting that the most relevant distribution for extreme returns is the Fréchet distribution (e.g. Longin, 1996, as well as Loretan and Phillips, 1994). Had the data been generated by a Gaussian distribution, then the value of the tail index would have been zero.

The location parameter, μ_N, indicates where extremes are located on average. The scale parameter, ψ_N, indicates the extent to which extreme realizations are dispersed. The tail index, ξ, focuses on *large* extreme returns. Fig. 1 displays the Fréchet distribution for different values of the tail index. We notice that, as the tail index increases, the probability mass in the right tail increases.

To obtain ML estimates of the parameter vector $(\xi,\mu_N,\psi_N)'$ of the gev distribution, we notice that the gev distribution of a general non-centered, non-reduced, random variable is defined by $H_{\xi,\mu_N,\psi_N}(m) = H_\xi((m - \mu_N)/\psi_N)$. Assume now that we observe a time series of the random variable X_t, forming a sample $\{x_1,\cdots,x_T\}$. We consider N-histories, i.e. non-

[3] Since $-\min(-X_1,\cdots,-X_N) = \max(X_1,\cdots,X_N)$, it is enough to develop theoretical elements for the upper tail of the distribution.

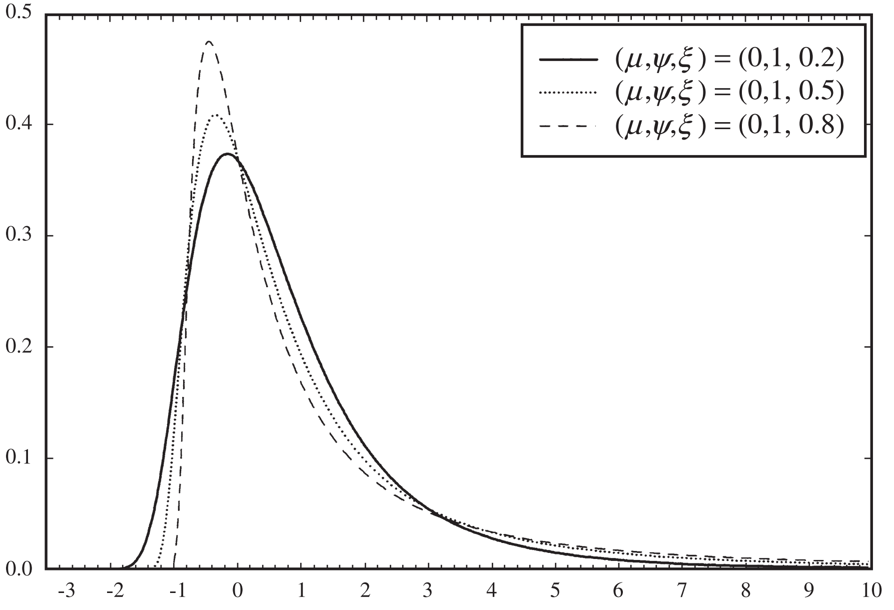

Fig. 1. Plot of various Fréchet distributions.

overlapping subsamples of length N. Picking the maximum over each N-history gives a sample of maxima, denoted $\{m_1,\cdots,m_\tau\}$, with τ being the largest integer smaller or equal to T/N. Parameters can then be easily estimated by applying a maximization routine to the sample likelihood defined as

$$L(m_i, i=1,\ldots,\tau;\xi,\mu_N,\psi_N)=\prod_{i=1}^{\tau} h_{\xi,\mu_N,\psi_N}(m_i)\mathscr{I}_{\{1+\xi(m_i-\mu_N)/\psi_N>0\}}(m_i), \tag{3}$$

where $h_{\xi,\mu_N,\psi_N}(m_i)$ *denotes the density of the gev distribution and* $\mathscr{I}$ *denotes an indicator function. As shown by Smith (1985), ML estimators are distributed asymptotically normal provided* $\xi > -1/2$.

2.2. The case of strictly stationary series

The case of strictly stationary, non-iid, series has been studied by Leadbetter et al. (1983). Embrechts et al. (1997) and McNeil (1998) also provide details and references in the context of the extreme approach applied to financial data.

Assume that X_t is now a strictly stationary, but non-iid, time series, and that $\tilde{X}_t$ is an associated iid series with the same marginal distribution F_X. We also define $M_N=\max(X_1,\cdots,X_N)$ and $\tilde{M}_N=\max(\tilde{X}_1,\cdots,\tilde{X}_N)$. Then, as shown by Leadbetter et al. (1983), the standardized extremes of the two series have the same limit distribution H (.) given in Eq. (2) under the two following conditions: (1) the series X has only weak long-range dependence, and (2) there is no tendency for extremes to cluster. For financial series, the former condition is rather mild, but the latter is not likely to apply. For instance, GARCH processes are weakly dependent, but have clustering of extremes. This is a reason why Kearns and Pagan (1997) obtained that, when returns are distributed as a GARCH process, the tail-index standard error is likely to be severely under-estimated.

When one of the two conditions is not satisfied, the limit distribution of the standardized extremes of X_t is shown to be equal to

$$\lim_{N\to\infty} Pr\left[\frac{M_N-\mu_N}{\psi_N}\leq y\right]=H^{\theta}(y).$$

The parameter θ, called the *extremal index*, measures the relationship between the dependence structure and the extremal behavior of the process (see Hsing et al., 1988; Embrechts et al., 1997, chap. 8). The extremal index verifies $0\leq\theta\leq 1$. The case $\theta=1$ corresponds to weak dependence and independence. Intuitively, the extremes over subsamples of length N from a strictly stationary series with extremal index $\theta<1$ have the same behavior as the extremes over subsamples of length $N\theta$ from the corresponding iid series.

Hsing et al. (1988) and Hsing (1991) indicate how to obtain estimates of the extremal index. Basically, for a given high threshold (say u), the extremal index θ can be estimated by dividing the number of subsamples of size n in which the maximum exceeds the threshold u (K_u) by the total number of exceedances in those subsamples (N_u). This definition suggests that θ can be interpreted as the reciprocal of the mean cluster size. An asymptotically equivalent estimator is given by

$$\tilde{\theta}=\frac{1}{n}\frac{\ln(1-K_u/\tau)}{\ln(1-N_u/(n\tau))}. \tag{4}$$

The asymptotic variance is provided by Hsing (1991). In the empirical section, we will report the estimator $\tilde{\theta}$ only, because it appears to be less sensible to the size of the sample (see the simulations reported in Embrechts et al., 1997).

An important issue is that, when returns are not iid, the behavior of extremes is driven by two components, the behavior of extreme innovations and the dependence structure. To make this statement clearer, assume that returns are indeed drawn from the GARCH model: $X_t=\sigma_t\varepsilon_t$, where ε_t is the innovation, and σ_t the conditional volatility. When ε_t is drawn from a distribution which allows fat tails (say, a Student t distribution), the fat-tailedness of returns comes from the fat-tailedness of innovations, but also from volatility clustering.[4] Since we are interested in measuring the unconditional behavior of extremes, we do not try, in a first step, to disentangle the two components of this behavior. Yet, in Section 4.3.3, we will measure the extent to which the dependence structure is likely to affect the behavior of extremes. More specifically, we will use the extremal index to

[4] In the approach based on the tails of the distribution, McNeil and Frey (2000) focus on the conditional behavior of extreme values. They first fit a GARCH model for returns and then consider the tail behavior of standardized residuals.

compare whether extreme realizations cluster more frequently in the left rather than in the right tail.

3. Preliminary analysis

3.1. The data

Since the aim of this study is to provide cross-country evidence for extreme returns, we consider a large database of global market indices. We consider five indices for mature financial markets (the US, Japan, Germany, France, and the UK). For emerging markets, we consider five Asian indices (Hong Kong, Singapore, South Korea, Taiwan, and Thailand), five Eastern European indices (Hungary, Poland, Russia, the Slovak Republic, and Slovenia), as well as five Latin American indices (Brazil, Chile, Colombia, Mexico, and Peru). This database has been extracted out of Datastream. For some of the series, care had to be taken for the earlier part of the sample, since the reported frequency was not daily.[5] We started our sample when the frequency became daily. For some countries, we had to append two indices.[6] All series end on June 28, 2002.

3.2. Descriptive statistics

Table 1 reports several descriptive statistics on the daily returns of stock-market indices.[7] We display for each index its label, the starting date, as well as the number of observations. The sample size (Nobs) ranges between 2094 for Russia up to 9780 for the US and German markets. The average daily return is positive for all countries.

Skewness (Sk) is a signed measure of the behavior of extreme returns. For mature markets, this statistic is generally found to be negative, suggesting that crashes cause an asymmetric return distribution. For emerging markets, the picture is less clear-cut, since many markets are characterized by a positive skewness. Excess kurtosis (XKu) measures the heaviness of tails relative to the normal one. We uniformly notice that this statistics is too large to be reconciled with a normal distribution. When finite-sample standard deviations are computed with the GMM procedure, we obtain that skewness and kurtosis are often estimated with a very low accuracy (see, for instance, the US market). The Wald statistic for normality tests the null hypothesis that skewness and excess kurtosis are jointly equal to zero. For all stock markets but the US, the hypothesis of normality is rejected at the 5% significance level. The rejection is more clear-cut in Eastern European and Latin American markets.

[5] This is the case for Poland before January 1993. For Brazil, the price figures are too small before January 1992 and do not allow a meaningful computation of returns.

[6] For the French index, we used the CAC Général index before July 9, 1987. For the UK index, we used the FT all-shares before January 1, 1980.

[7] Continuously compounded returns are defined by $X_t = 100 \times \ln(S_t/S_{t-1})$, where S_t is the closing value of an index at time t.

Table 1
Summary statistics on stock-market returns

	US	Japan	Germany	France	UK	Hong Kong	Singapore	South Korea	Taiwan	Thailand
Index	S&P500	Nikkei	DAX	CAC40	FTSE100	Hang Seng	Straits Times	KOSPI	TAIEX	S.E.T.
First Day	01/04/1965	01/02/1969	01/04/1965	01/02/1969	01/02/1969	01/03/1973	01/03/1973	01/02/1975	01/02/1970	05/02/1975
Nobs	9780	8737	9780	8737	8737	7693	7693	7172	8476	7086
Mean	0.024**	0.020	0.023*	0.031**	0.029**	0.033	0.009	0.034	0.042	0.019
Std	0.940**	1.126**	1.093**	1.086**	1.050**	1.962**	1.481**	1.567**	1.851**	1.461**
Sk	−1.552	−0.230	−0.485**	−0.559	−0.309	−1.405	−1.188	−0.258	−0.317	0.166
XKu	40.251	12.053**	8.695**	8.967**	8.723**	33.377	31.572	8.474**	14.377**	8.301**
Wald stat.	4.109	9.867**	15.652**	12.737**	6.762*	6.029*	7.695*	37.114**	6.990*	79.907**
Min	−24.302	−14.346	−12.567	−12.841	−12.429	−20.684	−20.288	−11.106	−15.109	−6.878
$q1$	−2.561	−3.021	−2.649	−2.739	−2.699	−2.924	−2.741	−2.971	−2.848	−2.878
$q5$	−1.566	−1.595	−1.538	−1.566	−1.529	−1.462	−1.434	−1.457	−1.547	−1.502
$q95$	1.529	1.464	1.486	1.529	1.451	1.390	1.424	1.568	1.508	1.472
$q99$	2.586	2.906	2.571	2.471	2.522	2.754	2.651	2.912	2.736	3.135
Max	9.234	11.021	8.099	7.548	8.486	8.776	10.706	6.375	10.734	7.756
LM(1)	130.794**	379.416**	352.484**	85.261**	2040.906**	57.131**	550.791**	190.905**	581.085**	694.868**
QW(10)	21.096*	16.409	26.971**	52.978**	40.876**	21.280*	25.344**	15.706	30.886**	55.476**

	Hungary	Poland	Russia	Slovak Rep.	Slovenia	Brazil	Chile	Colombia	Mexico	Peru
Index	BUX	WIG	Ros	SEVIS 100	SBI	Bovespa	IGPA	IBB	IPC	IGBL
First Day	01/03/1991	01/04/1993	6/21/1994	9/15/1993	01/04/1994	01/02/1992	01/05/1987	01/03/1992	01/05/1988	01/03/1991
Nobs	2997	2475	2094	2293	2214	2737	4040	2500	3779	2997
Mean	0.066	0.106	0.079	0.006	0.047	0.359**	0.073**	0.036	0.110**	0.127**
Std	1.698**	2.295**	3.587**	1.766**	1.311**	3.105**	0.927**	1.236**	1.754**	1.490**
Sk	−0.871	−0.093	0.962	2.387	−0.438	0.423	−0.326	−0.027	0.044	0.391*
XKu	15.370**	4.940	19.236**	39.496**	11.843**	5.983**	12.202	34.587	6.413**	5.130**
Wald stat.	25.570**	41.042**	13.049**	19.961**	13.167**	11.024**	14.242**	13.734**	24.854**	84.061**
Min	−10.660	−4.988	−7.336	−9.631	−8.674	−5.663	−13.346	−15.053	−8.223	−5.988
$q1$	−3.058	−3.128	−2.983	−2.823	−2.853	−2.794	−2.647	−2.522	−2.791	−2.821
$q5$	−1.345	−1.445	−1.408	−1.230	−1.467	−1.539	−1.451	−1.343	−1.514	−1.401
$q95$	1.499	1.574	1.420	1.312	1.460	1.672	1.526	1.512	1.521	1.647
$q99$	2.584	3.090	2.960	2.842	2.909	2.745	2.806	3.034	2.677	3.201
Max	7.980	6.394	12.795	13.600	8.290	9.163	6.900	12.139	6.867	5.894
LM(1)	298.451**	124.830**	27.840**	34.684**	407.328**	95.678**	72.449**	378.504**	247.697**	356.938**
QW(10)	18.374*	39.607**	26.103**	10.951	35.255**	31.865**	140.598**	50.446**	60.904**	135.846**

The first row indicates the name of the index used in the study, while the second row reports the date when a series start. Nobs, Mean, Std, Sk, and XKu denote the number of observations, the mean, the standard deviation, the skewness, and the excess kurtosis of returns, respectively. Standard errors are computed with GMM. Min and Max represent the minimum and maximum of standardized returns, while $q1$, $q5$, $q95$, and $q99$ represent the 1, 5, 95, and 99 percentiles. The 1%, 5%, 95%, and 99% percentiles for a normal distribution are −2.3263, −1.6449, 1.6449, and 2.3263. The LM(1) statistic for heteroskedasticity is obtained by regressing squared returns on one lag. QW(10) is the Box–Ljung statistic for serial correlation, corrected for heteroskedasticity, computed with 10 lags. Significance is denoted by superscripts at the 1% (**) and 5% (*) levels.

To glean further insight in the behavior of extreme returns, we center (by subtracting the mean) and reduce (by dividing by the standard deviation) all return series and consider the standardized minima and maxima as well as various percentiles. We define qx as the xth percentile. Comparison of the absolute value of $q1$ and $q99$, for standardized returns, with the value that should hold, under normality, 2.326, reveals for all markets, that the extreme 1 percentiles are too large to be compatible with a normal distribution. When comparing $q5$ and $q95$ with the associated normal critical values, 1.654, we find that there are not enough realizations compatible with a normal distribution. This confirms that distributions of returns are fat-tailed. When we compare the absolute values of $q1$ with $q99$, we notice very similar coefficients. This is not the case for the most extreme realizations where large differences may be observed. It is this type of observation that motivates our investigation. The size of the extreme return realizations shows that the study of the distribution followed by extreme returns is important.

Next, we test for heteroskedasticity by regressing squared centered returns on one lag. The TR^2 statistic (LM(1)), where R^2 is the coefficient of determination for a sample of size T, is distributed as a χ_1^2 under the null hypothesis of homoskedasticity. This statistic takes very large values, confirming that there is a fair amount of heteroskedasticity in the data.

Last, we investigate whether returns are serially correlated, using a version of the Box–Ljung test which corrects for heteroskedasticity. The test statistic, QW(10), is distributed as a χ_{10}^2 under the null hypothesis of no serial correlation. If we consider the statistic for 10 lags, we find that the statistic is significant for most indices investigated. At the 10% level, we do not reject the null hypothesis of serial dependency for Japan, South Korea, and the Slovak Republic only. Particularly high rejection values are found for Chile and Peru. These results suggest that great care should be taken when using evt techniques that rely on the assumption of independence.

4. Empirical results

4.1. Monte-Carlo simulations

An abundant literature discusses the small-sample properties of alternative tail-index estimators: Dekkers and de Haan (1989), Koedijk and Kool (1992), Dacorogna et al. (1995), Pictet et al. (1998), Danielsson et al. (1998), and Huisman et al. (2001). Most of these papers focus on the Hill estimator and some of its extensions. In this section, we perform a Monte-Carlo simulation to explore the small-sample properties of the tail-index ML estimator.[8] Since we consider several tests concerning characteristic parameters of the extreme distribution, we also investigate the small-sample properties of the LRT statistic, which will be extensively used in the following section. More specifically, we focus on the properties of the LRT statistics for the null hypothesis that $(\xi^-,\mu_N^-,\psi_N^-)=(\xi^+,\mu_N^+,\psi_N^+)$.

[8] We do not consider, specifically, the small-sample properties of the other parameter estimators (μ_N and ψ_N), because the theoretical value of these parameters is not known.

We explore the behavior of the gev when the data is either iid and generated by a number of heavy-tailed distributions, or by a dependent stochastic process. As in most papers cited above, we simulate series from the Cauchy, Stable, and Student *t* distributions, for which the theoretical tail index is well known. We also perform a similar exercise for two ARCH processes. As already discussed, ARCH processes generate fat-tailedness under Gaussianity and a fortiori if the innovations are fat-tailed.

Table 2 reports results for simulations based on 1000 samples with sample sizes $T=10{,}000$ (the upper bound for the sample size of mature markets) and $T=2000$ (the lower bound for the sample size of Eastern European markets). We control for the size of subsamples, reporting results for monthly, $N=20$, quarterly, $N=60$, and semi-annual, $N=120$, subsamples. For each sample and subsample size, we estimate the gev parameters by ML and we compute standard errors based on the standard deviation of the parameter distribution.[9] The table presents the average parameter estimate, the standard error, as well as the empirical *p*-value associated with the 90, 95, and 99 percentiles of the distribution of the LRT statistics for the null hypothesis $(\xi^-,\mu_N^-,\psi_N^-)=(\xi^+,\mu_N^+,\psi_N^+)$. This corresponds to the frequency, out of the 1000 simulated samples, for which we would have rejected the null hypothesis at the 10%, 5%, and 1% levels, respectively. Under the null, the LRT statistic is distributed as a χ_3^2. Thus, a correctly sized test statistic would yield empirical *p*-values equal to 10%, 5%, and 1%.

We are now going to discuss the results of our simulation focusing on samples of size $T=10{,}000$ corresponding to the left part of Table 2. Later on, we will emphasize the differences when the sample size shrinks down to $T=2000$. We start with discussing the quality of our tail index estimations.

In our first experiment, we generate iid data from a Cauchy distribution with theoretical tail index equal to $\xi=1$. The gev estimator performs very well, since the average estimator lies between 0.977 and 1.019, depending on the size of subsamples. The standard error of parameter estimates increases with the horizon, from 0.060 to 0.152. Therefore, on average, the true tail index is clearly within one standard error of the empirical index.

For the Stable distribution, we consider the characteristic exponent $\alpha=1.5$, which corresponds to a theoretical tail index of 0.667. There is no bias for quarterly and semi-annual subsamples with an average estimate of 0.675 and 0.688. However, for monthly subsamples, the average estimate is only 0.554.

For the Student *t* distribution, we simulate time series with degree of freedom equal to $v=2$, 3, and 6.[10] The theoretical tail index is then equal to $1/v$, i.e. 0.5, 0.33, and 0.167, respectively. For a low degree-of-freedom parameter ($v=2$), the small-sample bias is insignificant for large subsamples (quarterly and semi-annually). When v increases,

[9] We also computed, for each sample, the standard error of the ML parameter estimates. The average of these standard errors over all simulated samples can be compared with the standard deviation of the parameter distribution, to check whether problems may occur in the computation of the Hessian matrix. Results (not reported to save space, but available upon request from the authors) indicate that virtually no discrepancy appears between the two definitions of the standard error of parameter estimator.

[10] We also simulated samples with degree of freedom equal to 4 and 5. The choice of 2 and 6 is dictated by extreme cases that occur in actual applications. The choice of $v=3$ is dictated by the average value of estimates that will be reported later.

Table 2
Results of Monte-Carlo experiments

Subsample size	Panel A: Sample size T = 10,000					Panel B: Sample size T = 2000				
	Parameter estimate	Standard error	Empirical p-values			Parameter estimate	Standard error	Empirical p-values		
			10%	5%	1%			10%	5%	1%
Cauchy (ξ = 1)										
Monthly	0.977	0.060	0.143	0.070	0.015	0.995	0.132	0.127	0.072	0.017
Quarterly	1.009	0.104	0.097	0.055	0.006	1.041	0.264	0.140	0.080	0.023
Semi-annually	1.019	0.152	0.133	0.074	0.016	1.121	0.494	0.168	0.103	0.022
Stable(1.5) (ξ = 0.667)										
Monthly	0.554	0.050	0.229	0.132	0.062	0.579	0.112	0.206	0.125	0.050
Quarterly	0.675	0.086	0.134	0.080	0.015	0.707	0.216	0.143	0.079	0.016
Semi-annually	0.688	0.129	0.120	0.066	0.016	0.750	0.390	0.170	0.096	0.023
Student(2) (ξ = 0.5)										
Monthly	0.446	0.043	0.139	0.087	0.020	0.455	0.102	0.138	0.073	0.021
Quarterly	0.484	0.078	0.116	0.068	0.010	0.504	0.200	0.141	0.085	0.021
Semi-annually	0.493	0.112	0.106	0.064	0.016	0.554	0.356	0.169	0.094	0.016
Student(3) (ξ = 0.333)										
Monthly	0.258	0.039	0.070	0.037	0.005	0.259	0.095	0.065	0.033	0.008
Quarterly	0.299	0.067	0.100	0.044	0.010	0.299	0.175	0.077	0.038	0.008
Semi-annually	0.316	0.102	0.077	0.043	0.011	0.341	0.287	0.066	0.030	0.006
Student(6) (ξ = 0.167)										
Monthly	0.065	0.033	0.145	0.090	0.027	0.069	0.068	0.117	0.056	0.012
Quarterly	0.103	0.058	0.120	0.059	0.014	0.118	0.127	0.105	0.053	0.010
Semi-annually	0.121	0.086	0.105	0.053	0.011	0.178	0.240	0.100	0.055	0.017
ARCH(0.05,0.90) (ξ = 0.434)										
Monthly	0.372	0.043	0.080	0.042	0.017	0.375	0.105	0.058	0.026	0.001
Quarterly	0.424	0.078	0.037	0.017	0.003	0.442	0.189	0.040	0.016	0.003
Semi-annually	0.429	0.113	0.028	0.014	0.002	0.494	0.317	0.062	0.027	0.006
ARCH(0.05,0.70) (ξ = 0.315)										
Monthly	0.232	0.040	0.076	0.033	0.007	0.241	0.095	0.081	0.048	0.011
Quarterly	0.287	0.069	0.049	0.022	0.005	0.307	0.174	0.065	0.032	0.005
Semi-annually	0.298	0.109	0.029	0.012	0.001	0.348	0.281	0.058	0.029	0.006

For each sample size, we generate 1000 samples and compute the average and standard deviation of parameter estimates of ξ. The empirical p-values correspond to 10%, 5%, and 1% percent of the distribution of the LRT statistics for the null hypothesis $(\xi^-,\mu_N^-,\psi_N^-)=(\xi^+,\mu_N^+,\psi_N^+)$.

however, we obtain a negative bias for all subsample sizes. Yet, for quarterly and semi-annual subsamples, we find that the bias is non-significant.

These simulations show that the thicker the tail, the better the ML estimation performs. Such a result has already been put forward by Danielsson and de Vries (1997) or Pictet et al. (1998). This reflects the fact that, when the tails become thinner, the frequency of truly extreme returns is insufficient to allow the distribution of extremes to be estimated consistently.

We now turn to the ARCH process: $x_t = \sigma_t \varepsilon_t$, with $\sigma_t^2 = \omega + \alpha x_{t-1}^2$, where ε_t is an iid Gaussian innovation. For this exercise, we use as parameter set (ω,α)=(0.05,0.90) and (0.05,0.70). These parameters correspond to $\xi = 0.315$ and $\xi = 0.434$, as shown by de Haan et al. (1989). We observe that dependence is responsible for a relatively small, and non-significant, bias that vanishes as subsamples become larger. For instance, for the case where $\xi = 0.434$, for monthly subsamples, we obtain an estimation of 0.372 that increases up to 0.429 for semi-annual subsamples.

We presently wish to discuss the size of our LR test. We focus on the null hypothesis that the three characteristic parameters of the left tail are equal to the ones of the right tail. The top row of the table indicates the theoretical rejection frequency of our LRT. The chosen levels are 10%, 5%, and 1%. Inspection of the rejection rates obtained for independent data shows that, in general, the LRT rejects slightly too often the null hypothesis of equality of the tails. For instance, for the Student t distribution with two degrees of freedom, we obtain, for quarterly subsamples, a rejection rate of 11.6%, and 6.8%, rather than 10%, and 5%. The size of the 1% test is correct.

This picture is somewhat modified when we turn to dependent data, simulated in the ARCH. We find that the LRT does not reject often enough, whatever the rejection level, and whatever the tail index of the ARCH. The bias is smallest, however, for small subsamples. For instance, for the ARCH obtained for (ω,α)=(0.05,0.9), we obtain a rejection frequency of 8.0%, 4.2%, and 1.7% rather than the theoretical values of 10%, 5%, and 1%, a rather small bias though. As we will see in the later part of this study, the tail index of our data tends to be around 0.3. This implies that, if our data is generated by some ARCH process, then, for small subsamples, our LRT will be correctly sized.

Presently, we consider the results of the simulation associated with the samples of size $T = 2000$ presented in Panel B, in the right part of Table 2. Our first observation is that the standard error of the tail index increases. In most cases, the standard error doubles. Once this increase of standard errors is accounted for, we notice for all tail-index estimates, obtained for iid fat-tailed distributions, that the 95% confidence intervals always contain the true tail index. For the case of ARCH-dependent data, we notice that, for the quarterly and semi-annual subsamples, the confidence interval of the tail index contains the true parameter, especially for the case $\xi = 0.315$. This tail index is close to the values we find for actual data.

When we turn to the size of the LRT for independent data, we find that, as for the large sample size, the equality of the tails tends to be rejected too easily. When we turn to dependent data and the situation where $\xi = 0.315$, we find also that our test rejects with approximately the correct size for data sampled at a monthly frequency. For instance, for monthly subsamples, we obtain rejection rates of 8.1%, 4.8%, and 1.1%, rather than 10%, 5%, and 1%. On the other hand, we recognize that the bias of the rejection frequency can be quite substantial as the tail index is large, i.e. $\xi = 0.434$. These results tend to suggest that dependency may affect the LRT. However, for actual tail-index values, the bias is rather small, especially for large samples and for small subsamples.

4.2. Estimation of the generalized extreme value distribution

We now consider the distribution of extreme returns, obtained over N-histories for the stock markets under study. For the first two groups of markets (mature and Asian markets), we select quarterly subsamples, while for the last two groups (Eastern European and Latin American markets), we select monthly subsamples. In both cases, this corresponds to an average of 140 extreme observations.

We start our discussion with the parameter estimates obtained for the gev distribution fitted to subsample minima and maxima, separately. Results of this estimation are presented in Table 3. The left part of the table is devoted to shortfalls, whereas the right part is devoted to windfalls.

Presently, we wish to discuss whether there are obvious similarities between the left and right parameter estimates. Later on, we will show that the adequacy of our estimations is good. Inspection of the estimates of ξ^-, ξ^+, and associated standard errors reveals that the value of 0 is excluded for all markets. This implies that, for neither market, the tails behave in a way compatible with the normal distribution, corroborating our earlier finding reported in Table 1, concerning excess kurtosis. This result also indicates that models such as finite mixtures of normals, known to generate tail indices of $\xi = 0$, do not

Table 3
Maximum-likelihood parameter estimates of the gev distribution

	Panel A: Left tail						Panel B: Right tail					
	μ^-	S.E.	ψ^-	S.E.	ξ^-	S.E.	μ^+	S.E.	ψ^+	S.E.	ξ^+	S.E.
Mature markets (N = 60)												
G1 US	1.598	0.056	0.658	0.047	0.282	0.071	1.710	0.060	0.690	0.044	0.132	0.053
G1 Japan	1.882	0.114	1.007	0.079	0.265	0.087	1.946	0.111	1.036	0.082	0.268	0.072
G1 Germany	1.771	0.071	0.758	0.052	0.292	0.073	1.903	0.062	0.741	0.053	0.232	0.051
G1 France	1.928	0.086	0.928	0.062	0.182	0.067	1.979	0.076	0.824	0.056	0.153	0.055
G1 UK	1.794	0.059	0.630	0.050	0.273	0.071	1.731	0.057	0.561	0.047	0.330	0.076
Asia (N = 60)												
G2 Hong Kong	3.009	0.174	1.669	0.144	0.374	0.079	3.184	0.144	1.461	0.125	0.294	0.065
G2 Singapore	2.403	0.115	1.114	0.101	0.407	0.077	2.612	0.118	1.176	0.103	0.265	0.073
G2 South Korea	2.337	0.165	1.461	0.125	0.290	0.074	2.741	0.135	1.302	0.100	0.137	0.056
G2 Taiwan	3.230	0.131	1.434	0.110	0.234	0.070	3.294	0.142	1.417	0.106	0.244	0.069
G2 Thailand	2.168	0.203	1.441	0.135	0.200	0.100	2.269	0.202	1.507	0.140	0.237	0.093
Eastern Europe (N = 20)												
G3 Hungary	1.666	0.095	1.030	0.088	0.413	0.070	1.882	0.111	1.097	0.080	0.267	0.073
G3 Poland	2.509	0.155	1.483	0.129	0.264	0.060	2.774	0.175	1.520	0.122	0.193	0.068
G3 Russia	3.793	0.269	2.569	0.240	0.209	0.065	4.369	0.313	2.761	0.247	0.300	0.078
G3 Slovak Rep.	2.288	0.143	1.297	0.111	0.313	0.079	2.062	0.144	1.280	0.107	0.348	0.091
G3 Slovenia	1.279	0.105	0.912	0.081	0.298	0.084	1.435	0.108	0.911	0.086	0.293	0.080
Latin America (N = 20)												
G4 Brazil	3.309	0.175	1.774	0.138	0.225	0.059	3.879	0.199	2.050	0.172	0.203	0.074
G4 Chile	0.930	0.046	0.559	0.037	0.280	0.061	1.173	0.056	0.665	0.043	0.179	0.054
G4 Colombia	1.273	0.083	0.842	0.073	0.254	0.077	1.493	0.103	1.064	0.083	0.198	0.066
G4 Mexico	1.985	0.097	1.104	0.075	0.259	0.059	2.418	0.090	1.050	0.071	0.240	0.059
G4 Peru	1.497	0.084	0.933	0.084	0.271	0.065	1.809	0.112	1.096	0.085	0.229	0.061

The first column G1,. . ., G4 indicates to which group a given country belongs. Superscripts − and + correspond to the left and right tails. The label S.E. represents the robust standard errors.

adequately capture the tails of market returns.[11] Only models where it is known that the tails belong to the domain of attraction of the Fréchet distribution are potentially successful.

Estimates of the tail index allow to address the issue of the existence of moments of the distribution. Indeed, one has the property that, for all integers r such that $r<1/\xi$, the rth moment exists. We, thus, obtain that, for most markets, the behavior of negative extreme returns is not compatible with the existence of kurtosis. Furthermore, for three markets in Asia and Eastern Europe, even the skewness does not appear to exist. The behavior of positive extreme returns is more in agreement with the existence of high moments. For the US, our findings are consistent with the empirical evidence reported by Longin (1996).

For the countries located in the first group (G1), we notice that the location and scale parameters appear to be similar from country to country. Similarly, the parameters are close in group 2 (G2), but at a higher level. This may be explained by the fact that G2 countries have a higher level of volatility than G1 countries (see also Table 1). When we turn to countries in groups 3 and 4 (G3, G4), we find lots of heterogeneity among the estimates. A simple inspection of the estimates does not lead to any obvious finding. To establish results concerning similarity or disparity in the behavior of extreme returns, we need to perform formal tests.

To assess the accuracy of the estimation of the extreme value distribution, we perform a test of goodness of fit of the gev distribution. We follow Longin (1996) and perform a Sherman (1957) test of goodness of fit. This test, presented in Table 4 and described in the caption thereof, compares the probability given by the estimated gev distribution and the observed frequency, used as a proxy of the true probability. The null hypothesis that the gev distribution is a good approximation of the true probability is rejected, at the 5% (1%) level, in 7 (respectively 1) out of the 40 cases only. Only in the case of Russia, the goodness of fit is rejected, for both tails, at the 5% significance level. This is also the only country where the adequacy of a tail estimation is rejected at the 1% level. The generally good fit corroborates, however, that our model is satisfactory.

4.3. Tests on the behavior of extremes

The empirical literature on the evt has essentially focused on the estimation of the tail index. One reason is that many authors are mainly interested in computing VaR and/or very high quantiles. From this viewpoint, only the tail index is needed and more demanding estimation methods (such as ML estimation of the gev or the gpd) can be seen as too expensive.

4.3.1. Do extreme negative returns behave like extreme positive returns?

We have seen in Table 3, in a comparison between the left and right tail indices, that for most countries the left tail index takes larger values than the right one. However, the standard errors of parameter estimates are rather large. As already mentioned, a formal test

[11] See also Jansen and de Vries (1991, p. 22).

Table 4
Goodness of fit

	Panel A: Left tail			Panel B: Right tail		
	Log-lik.	Goodness-of-fit statistic	*p*-value	Log-lik.	Goodness-of-fit statistic	*p*-value
Mature markets (N = 60)						
G1 US	− 215.19	1.281	0.100	− 209.011	0.268	0.395
G1 Japan	− 251.64	0.312	0.378	− 256.092	− 0.875	0.191
G1 Germany	− 239.01	0.929	0.176	− 230.340	− 0.908	0.182
G1 France	− 232.66	0.447	0.327	− 213.410	− 1.314	0.094
G1 UK	− 184.53	− 0.517	0.303	− 172.361	− 2.038	0.021*
Asia (N = 60)						
G2 Hong Kong	− 294.82	− 1.494	0.068	− 272.359	− 0.589	0.278
G2 Singapore	− 245.53	− 0.873	0.191	− 242.309	1.870	0.031*
G2 South Korea	− 252.64	− 2.188	0.014*	− 228.749	− 0.173	0.431
G2 Taiwan	− 292.08	− 0.869	0.192	− 291.341	− 0.286	0.387
G2 Thailand	− 242.99	1.929	0.027*	− 250.625	0.646	0.259
Eastern Europe (N = 20)						
G3 Hungary	− 274.95	− 0.439	0.331	− 271.628	− 0.022	0.491
G3 Poland	− 261.38	1.081	0.140	− 259.388	− 1.018	0.154
G3 Russia	− 274.91	2.208	0.014*	− 287.625	− 2.599	0.005**
G3 Slovak Rep.	− 229.98	− 0.365	0.357	− 230.523	− 0.065	0.474
G3 Slovenia	− 182.09	− 0.950	0.171	− 181.767	0.279	0.390
Latin America (N = 20)						
G4 Brazil	− 310.26	− 0.056	0.478	− 328.054	0.608	0.272
G4 Chile	− 233.70	0.103	0.459	− 257.455	− 0.046	0.482
G4 Colombia	− 193.97	− 0.354	0.362	− 219.081	0.273	0.393
G4 Mexico	− 343.23	− 0.922	0.178	− 331.844	− 1.145	0.126
G4 Peru	− 248.46	− 0.413	0.340	− 268.598	− 1.737	0.041*

Log-lik. is the sample log-likelihood of the ML estimation. The goodness-of-fit test statistic is computed as $\Omega_\tau = 1/2\sum_{i=0}^{\tau} |H_{\hat{\xi},\hat{\mu}_T,\hat{\psi}_T}(m_i) - H_{\hat{\xi},\hat{\mu}_T,\hat{\psi}_T}(m_{i+1}) - (1/(\tau+1))|$ with $H_{\hat{\xi},\hat{\mu}_T,\hat{\psi}_T}(m_0)=0$ and $H_{\hat{\xi},\hat{\mu}_T,\hat{\psi}_T}(m_{\tau+1})=1$. Ω_τ is asymptotically normal, with mean $(\tau/(\tau+1))^{\tau+1}$ and variance $(2e-5)/(e^2\tau)$. See Sherman (1957) for further details. Significance is denoted by superscripts at the 1% (**) and 5% (*) levels.

is therefore necessary to investigate whether negative extreme returns behave as positive extreme returns.

In Table 5, we present various LR tests of equality of μ, ψ, and ξ for the left and right tails. In the first pair of columns, we present the LRT statistic that tests whether the left and right tails have a same location parameter, i.e. $\mu^- = \mu^+$. We reject the null at the 5% level for most Latin American markets. At this level of significance, in other geographical areas, we are not able to reject the null. Since μ indicates where extreme realizations are anchored, these results suggest that most negative extremes have an absolute value similar to the positive extremes.

When we turn to columns 3 and 4, corresponding to a test of the null hypothesis that $\psi^- = \psi^+$, we find no rejection, except for Colombia. This finding indicates that minima and maxima display the same dispersion.

Table 5
LR tests for equality of parameters in the left and right tails

	H_0: $\mu^- = \mu^+$		H_0: $\psi^- = \psi^+$		H_0: $\xi^- = \xi^+$	
	LRT	*p*-value	LRT	*p*-value	LRT	*p*-value
Mature markets (N= 60)						
G1 US	1.768	0.184	0.232	0.630	3.056	0.080
G1 Japan	0.201	0.654	0.057	0.811	0.001	0.980
G1 Germany	1.950	0.163	0.046	0.831	0.418	0.518
G1 France	0.188	0.664	1.332	0.248	0.100	0.752
G1 UK	0.633	0.426	1.037	0.309	0.296	0.586
Asia (N= 60)						
G2 Hong Kong	0.606	0.436	1.159	0.282	0.495	0.482
G2 Singapore	1.628	0.202	0.190	0.663	1.581	0.209
G2 South Korea	3.698	0.054	0.918	0.338	1.715	0.190
G2 Taiwan	0.109	0.741	0.011	0.916	0.009	0.924
G2 Thailand	0.185	0.667	0.117	0.733	0.056	0.814
Eastern Europe (N= 20)						
G3 Hungary	2.325	0.127	0.303	0.582	1.762	0.184
G3 Poland	1.420	0.233	0.042	0.838	0.336	0.562
G3 Russia	1.903	0.168	0.314	0.575	0.622	0.430
G3 Slovak Rep.	1.315	0.251	0.011	0.916	0.076	0.783
G3 Slovenia	1.199	0.274	0.000	1.000	0.001	0.970
Latin America (N= 20)						
G4 Brazil	4.687	0.030*	1.722	0.189	0.049	0.825
G4 Chile	11.937	0.001**	3.500	0.061	1.324	0.250
G4 Colombia	2.622	0.105	4.170	0.041*	0.325	0.569
G4 Mexico	11.211	0.001**	0.264	0.608	0.041	0.839
G4 Peru	5.364	0.021*	2.167	0.141	0.159	0.690

LRT is the test statistics for the null hypotheses that $\mu^- = \mu^+$, $\psi^- = \psi^+$, and $\xi^- = \xi^+$, respectively. Significance is denoted by superscripts at the 1% (**) and 5% (*) levels.

Finally, we never reject the null hypothesis that the left tail index is equal to the right tail index, i.e. $\xi^- = \xi^+$. A Wald test of equality of the two tail indices confirms this result. This result is, to some extent, surprising, because in many cases the left tail index is estimated to be much larger than the right tail index. This is the case, for instance, for the US, Hong Kong, and Singapore. However, since standard errors are large, we do not reject the null. This contrasts, to some extent, with results obtained by Hartmann et al. (2001), who found that, in the French and US stock markets, the left tail index is significantly larger than the right tail index. However, since they use the Hill estimator on weekly data over a rather short period of time (1987–1999), our results are not directly comparable.

Considered together, our results extend the one by Longin (1996) who showed, for a long series of daily realizations of the S&P, that the gev parameters of the left tail tend to be equal to the ones of the right tail. We find that the exception to this general rule is the set of Latin American markets. Even though the equality of the left and right tail indices and scale parameters cannot be rejected, we find, at the 1% significance level, that for

Chile and Mexico extremes are located further out in the right tail rather than in the left tail.

Similarity between the left and right tails of returns has also been reported for other datasets. For instance, Koedijk and Kool (1992) cannot reject the equality between the left and right tails of Eastern European exchange rates. Danielsson and de Vries (1997, p. 253), using high frequency exchange-rate data, also conclude that the tail shapes appear to be very similar, suggesting tail symmetry. The observation that the behavior of the left tail is similar to the right one appears to be of rather general nature.

4.3.2. Is the behavior of extreme returns similar across markets?

Now, we test whether the behavior of extremes differs between the various markets in a geographical group. In Table 6, we present the LRT statistic which tests, for each tail separately, the null hypothesis that the parameters are constant within each country group. Results for the left tail of the distribution are found in Panel A, whereas results for the right tail are found in Panel B. Considering the left tail of group G1, we obtain a total log-likelihood of -1123.04. When we estimate the model with ψ and ξ free, yet μ set constant for all countries, we obtain $\mu = 1.828$ and a log-likelihood of -1129.49. This yields a LRT statistic, also reported in Table 6, of 12.902. Since the statistic is distributed as a χ^2_4, under the null hypothesis that $\mu_1 = \ldots = \mu_5 = \mu$, the p-value is 1.2%, so that the null is strongly rejected. Inspection of the p-values of Table 6 shows that the constancy of μ across countries can be rejected in all geographical groups.

We proceed similarly for the scale parameter, ψ. We also obtain that constancy of ψ is rejected within all groups, but the Asian one. For the Asian markets, we cannot reject that the dispersion of extremes is similar.

Table 6
LR tests for equality of parameter within a given group

	H$_0$: $\mu_i = \mu_j$		H$_0$: $\psi_i = \psi_j$		H$_0$: $\xi_i = \xi_j$		ξ^-	S.E.	Confidence interval
	LRT	p-value	LRT	p-value	LRT	p-value			
Panel A: Left tail									
G1 Mature markets	12.902	0.012	27.963	0.000	1.727	0.786	0.261	0.031	[0.20,0.32]
G2 Asia	40.587	0.000	10.047	0.040	4.214	0.378	0.309	0.037	[0.23,0.38]
G3 Eastern Europe	107.236	0.000	91.028	0.000	4.023	0.403	0.308	0.036	[0.24,0.38]
G4 Latin America	230.729	0.000	128.082	0.000	0.358	0.986	0.260	0.030	[0.20,0.32]
	H$_0$: $\mu_i = \mu_j$		H$_0$: $\psi_i = \psi_j$		H$_0$: $\xi_i = \xi_j$		ξ^+	S.E.	Confidence interval
	LRT	p-value	LRT	p-value	LRT	p-value			
Panel B: Right tail									
G1 Mature markets	13.342	0.010	30.690	0.000	5.300	0.258	0.217	0.031	[0.15,0.28]
G2 Asia	31.450	0.000	5.499	0.240	2.330	0.675	0.240	0.036	[0.17,0.31]
G3 Eastern Europe	116.334	0.000	86.046	0.000	1.584	0.812	0.284	0.039	[0.21,0.36]
G4 Latin America	257.022	0.000	135.702	0.000	0.570	0.966	0.209	0.030	[0.15,0.27]

LRT is the test statistics for the null hypotheses that the various parameters are constant within each geographical group. Under the null, LRT is distributed as a χ^2_4. The last column contains the 95% confidence interval of the various tail indices.

Turning to ξ, finally, shows that equality of the tail index cannot be rejected in either group. We therefore report in the table the estimates of ξ^- and ξ^+ for each group. As far as the left tail is concerned, tail indices are very close, ranging from 0.16 up to 0.30, and in all cases the fourth moment does not appear to exist. In contrast, right tail indices are twice as dispersed, ranging from 0.21 to 0.28, and the fourth moment is found to be infinite in the case of Eastern European markets only.

When we compare the left and right tail indices, we notice again that the point estimates associated with the left tail tend to be larger. When we turn to the confidence intervals, we notice that the right tail is always included in the 95% confidence interval. It should be noticed that presently, the parameters are estimated with much larger samples leading to smaller standard errors. Yet, we still cannot reject, from a statistical point of view, that the left and right tail indices are equal.

To sum up, these findings suggest that, across countries, the area where extreme returns are located differs. For instance, a -10% crash may seem natural for an emerging market, whereas it would be a large extreme in some more mature market. Moreover, the dispersion of extremes differs across markets. Indices, belonging to mature markets, have in general little variability across extremes, whereas there is more variability in Eastern European ones. Yet, the behavior of returns is found to be rather similar far out in the tails. One possibility for this finding is that the number of extreme events is just not large enough as to allow the statistical detection of asymmetries in extreme tail events.

4.3.3. How dependent are the extremes?

Finally, we address the issue of how dependent the extreme returns are. This is an important issue, since dependency of extremes directly contributes to the fat-tailedness of returns. As mentioned in the Introduction, the perception that left tails are heavier than the right tails could be due to a clustering of events in the left tail. Table 7 reports, for each market, the estimate of the extremal index θ, computed using Eq. (4), and the associated standard error. For the first two groups of markets, i.e. mature and Asian markets, we consider a threshold u such that 200 observations are above u. For the last two groups, the threshold is such that 50 observations are above u. Since we are interested in clustering of news, we focus on blocks of weekly length ($n=5$).

We recall that a value of $\theta=1$ corresponds to no clustering. From a statistical point of view, we find that some clustering occurs for all series. Given that our estimation method is independent of the fact that data clusters, our interest is, therefore, on the relative magnitude of the measure of clustering. This leads us to consider the difference between the cluster intensity of the left and right tails.[12] We find that, for all countries in groups 1 and 2, the left tail clusters more than the right tail. One exception to this rule is Thailand. On the other hand, countries located in Eastern Europe or Latin America tend to have

[12] We used a Wald statistic for testing the null hypothesis that $\theta^-=\theta^+$, assuming that the covariance between both parameter estimators is zero. We verified, using Monte-Carlo simulations of an ARCH process, that this assumption is costless for such processes.

Table 7
Estimates of the extremal index (obtained for weekly horizons)

	Left tail		Right tail		Difference	
	θ^-	S.E.	θ^+	S.E.	$\theta^- - \theta^+$	t-stat
Mature markets						
G1 US	0.875	0.024	0.907	0.025	– 0.032	– 0.923
G1 Japan	0.851	0.023	0.911	0.021	– 0.060	– 1.926
G1 Germany	0.858	0.025	0.924	0.021	– 0.066	– 2.021*
G1 France	0.862	0.025	0.922	0.021	– 0.060	– 1.838
G1 UK	0.824	0.024	0.868	0.027	– 0.044	– 1.218
Asia						
G2 Hong Kong	0.779	0.030	0.867	0.026	– 0.088	– 2.217*
G2 Singapore	0.773	0.026	0.834	0.025	– 0.061	– 1.691
G2 South Korea	0.820	0.024	0.853	0.025	– 0.033	– 0.952
G2 Taiwan	0.787	0.030	0.869	0.025	– 0.082	– 2.100
G2 Thailand	0.826	0.028	0.815	0.025	0.011	0.293
Eastern Europe						
G3 Hungary	0.821	0.054	0.779	0.072	0.042	0.467
G3 Poland	0.719	0.050	0.740	0.055	– 0.021	– 0.283
G3 Russia	0.875	0.049	0.875	0.043	0.000	0.000
G3 Slovak Rep.	0.959	0.035	0.893	0.048	0.066	1.111
G3 Slovenia	0.829	0.049	0.764	0.052	0.065	0.910
Latin America						
G4 Brazil	0.909	0.047	0.887	0.042	0.022	0.349
G4 Chile	0.899	0.053	0.857	0.043	0.042	0.615
G4 Colombia	0.934	0.038	0.890	0.048	0.044	0.719
G4 Mexico	0.880	0.054	0.859	0.049	0.021	0.288
G4 Peru	0.885	0.042	0.736	0.055	0.149	2.153*

The way the extremal index is computed is described in Section 2.2. Weekly horizons ($n=5$) are used to assess the clustering of volatility. The t-stat corresponds to the Wald test of the null hypothesis $\theta^- = \theta^+$.

positive clusters. The t-statistics of a bi-lateral test is, however, significant in only three out of the 20 estimations.

These results show that, when we consider the point estimates, events in the left tail cluster more than those in the right tail. However, this finding is not statistically significant. As a consequence, clustering of news cannot be the reason why investors consider the left tail to be heavier than the right tail.

5. Discussion

Most investors would affirm that the left tail of the return distribution of stocks is heavier than the right one, according to the idea that crashes are more likely to occur than booms, and based on the observation that there have been drops in the stock index of a magnitude never exceeded by increases, in absolute value. For instance, the S&P dropped on a single day in October 1987 by – 24.03%, whereas the largest increase was 9.234%.

Similarly, the French CAC index had as worse return, a drop of − 12.48%, whereas its best performance since 1969 was 7.54%.

In this study, we formally investigate whether there is a statistically significant difference between the left and right tails of returns. To lead this investigation, we consider tests based on the gev distribution. The gev has three parameters allowing us to measure the location, the dispersion, and the asymptotic behavior of extremes. To assess the quality of the estimates obtained for the gev, and to validate the rejection rate of the likelihood ratio test, we perform a simulation exercise. We find that the tail-index estimations are better, the thicker the tail index. In general, the true tail index is located within the confidence interval of our estimated index, and the simulation indicates the size of subsamples to choose, so that the LRT has the right rejection rate. When data is dependent, simulated according to an ARCH process, with a tail index similar to the one of actual data, we find that rather small subsamples should be chosen.

Having confirmed the quality of our estimation method, we cannot reject, for actual data, the hypothesis that the characteristic parameters of the left and right tails are equal. An exception to this rule is Latin America, where we find that extreme values are located further out in the right tail than in the left tail. In all cases, the left tail index is not significantly different from the right one. As such, we are able to extend the study of Longin (1996) who finds equality of the tails for the US market.

In a subsequent test, we explore whether the tail parameters are identical within geographic areas. For none of the geographic areas can we reject the hypothesis that the tail index is the same for all countries. We find, however, differences in the location and the dispersion of extremes within the various areas. This finding suggests that data should not be pooled to yield larger samples.

This leaves us with the question why do investors have the perception that the left tail is heavier than the right tail? Can it be because extremes cluster more in the left tail rather than in the right tail? To investigate this issue, we consider the extremal index, a natural measure of clustering. In a formal test, we cannot reject the assumption that extremes cluster equally in both tails even though point estimates are slightly higher for the left tail. As a consequence, clustering does not seem to be the reason why left tails are deemed heavier than right tails. We are, therefore, left with the possibility that the samples that we are using, even though they may consist of up to 10,000 observations, are still too small to yield statistically significant tests. The perception of investors concerning the tail thickness could be due to the size of the very few left outliers.

Acknowledgements

Our data has been extracted from Datastream. Rockinger (formerly at HEC-Paris) acknowledges help from the TMR grant on *Financial market's efficiency* as well as from the Fondation HEC. We also wish to thank the anonymous referee, the editor (F. Palm), Sanvi Avouyi-Dovi, Jean Cordier, Henri Pagès, Patrice Poncet, Pierre Sicsic, as well as Seminar participants at the Banque de France, for their comments. The usual disclaimer applies. This paper does not necessarily reflect the view of the Banque de France.

References

Bekaert, G., Harvey, C.R., 1997. Emerging equity market volatility. Journal of Financial Economics 43, 29–77.

Campbell, J.Y., Hentschel, L., 1992. No news is good news: an asymmetric model of changing volatility in stock returns. Journal of Financial Economics 31, 281–318.

Claessens, S., Dasgupta, S., Glen, J., 1995. Return behavior in emerging stock markets. World Bank Economic Review 9, 131–151.

Dacorogna, M.M., Müller, U.A., Pictet, O.V., de Vries, C.G., 1995. The Distribution of Extremal Foreign Exchange Rate Returns in Extremely Large Data Sets, mimeo, Olsen and Associates. Zurich and Erasmus University, Rotterdam.

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., 1998. Using a Bootstrap Method to Choose the Sample Fraction in Tail Index Estimation. Working Paper, Erasmus University Rotterdam and Tinbergen Institute.

de Haan, L., Resnick, S.I., Rootzén, H., de Vries, C.G., 1989. Extremal behaviour of solutions to a stochastic difference equation with applications to ARCH processes. Stochastic Processes and their Applications 32, 213–224.

Dekkers, A.L., de Haan, L., 1989. On the estimation of the extreme value index and large quantile estimation. Annals of Statistics 17, 1795–1832.

Embrechts, P., Klüppelberg, C., Mikosch, T., 1997. Modelling Extremal Events. Springer Verlag, Berlin.

Gnedenko, B.V., 1943. Sur la distribution limite du terme maximum d'une série aléatoire. Annals of Mathematics 44, 423–453.

Hartmann, P., Straetmans, S., de Vries, C.G., 2001. Asset Market Linkages in Crisis Period, ECB Working 71.

Hill, B.M., 1975. A simple general approach to inference about the tail of a distribution. 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.

Hsing, T., 1991. Estimating the parameters of rare events. Stochastic Processes and their Applications 37, 117–139.

Hsing, T., Hüsler, J., Leadbetter, M.R., 1988. On the exceedance point process for a stationary sequence. Probability Theory and Related Fields 78, 97–112.

Huisman, R., Koedijk, K.G., Kool, C.J.M., Palm, F., 2001. Tail-index estimates in small samples. Journal of Business and Economic Statistics 19, 208–216.

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, 18–24.

Kearns, P., Pagan, A., 1997. Estimating the density tail index for financial time series. Review of Economics and Statistics 79, 171–175.

Koedijk, K.G., Kool, C.J.M., 1992. Tail estimates of East European exchange rates. Journal of Business and Economic Statistics 10, 83–96.

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.

Leadbetter, M.R., Lindgren, G., Rootzén, H., 1983. Extremes and Related Properties of Random Sequences and Processes. Springer Verlag, New York.

Longin, F.M., 1996. The asymptotic distribution of extreme stock market returns. Journal of Business 69, 383–408.

Longin, F.M., Solnik, B., 2001. Extreme correlation of international equity markets. Journal of Finance 56, 649–676.

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

Lux, T., 1998. The Limiting Extremal Behavior of Speculative Returns: An Analysis of Intra-Daily Data from the Frankfurt Stock Exchange. University of Bonn Discussion Paper B/436.

McNeil, A.J., 1996. Estimating the Tails of Loss Severity Distributions Using Extreme Value Theory. Working Paper, ETH Zentrum.

McNeil, A.J., 1998. Calculating Quantile Risk Measures for Financial Return Series Using Extreme Value Theory. Working Paper, ETH Zentrum.

McNeil, A.J., Frey, R., 2000. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. Journal of Empirical Finance 7, 271–300.

McNeil, A.J., Saladin, T., 1997. The Peaks over Thresholds Method for Estimating High Quantiles of Loss Distributions. Working Paper, ETH Zentrum.

Pickands, J., 1975. Statistical inference using extreme order statistics. Annals of Statistics 3, 119–131.

Pictet, O.V., Dacorogna, M.D., Müller, U.A., 1998. Hill, bootstrap and jackknife estimators for heavy tails. In: Adler, R.J., Feldman, R.E., Taqqu, M.S. (Eds.), A Practical Guide to Heavy Tails: Statistical Techniques and Applications. Birkhäuser, Boston, pp. 283–310.

Quintos, C., Fan, Z., Phillips, P.C.B., 2001. Structural change tests in tail behaviour and the Asian crisis. Review of Economic Studies 68, 633–663.

Reiss, R.D., Thomas, M., 1997. Statistical Analysis of Extreme Values. Birkhäuser, Basel.

Rockinger, M., Urga, G., 2001. A time-varying parameter model to test for predictability and integration in the stock markets of transition economies. Journal of Business and Economic Statistics 19, 73–84.

Sherman, L.K., 1957. Percentiles of the Ω_n statistic. Annals of Mathematical Statistics 28, 259–268.

Smith, R.L., 1985. Maximum likelihood estimation in a class of non-regular cases. Biometrika 72, 67–90.