TECHNICAL APPENDIX TO "EXTREME U.S. STOCK MARKET FLUCTUATIONS IN THE WAKE OF 9/11" (JOURNAL OF APPLIED ECONOMETRICS, FORTHCOMING)

S.T.M. STRAETMANS[A], W.F.C. VERSCHOOR[B], AND C.C.P. WOLFF[C,*]

1. Introduction

The results below are not included in the main body of the paper for sake of space considerations. We refer to most of these additional calculations/findings via footnotes in the paper or short remarks in the text.

2. Additional asymmetry tests

We also performed structural change and asymmetry tests for conventional CAPM-type βs as compared to the tail-βs in the main body of our paper. We therefore calculated CAPM-βs like in Ang and Chen (2002). These results can be compared with our 1st set of results on tail-βs with respect to the NYSE Composite.[1] The latter authors calculate correlations and βs for sets of different portfolios conditional on the tail area. More specifically, the two-sided truncated CAPM-βs boil down to:

$$\beta_u = \frac{cov\left(X_1, X_2 | X_1 > x_1, X_2 > x_2\right)}{var\left(X_2 | X_1 > x_1, X_2 > x_2\right)}$$

$$\beta_l = \frac{cov\left(X_1, X_2 | X_1 < x_3, X_2 < x_4\right)}{var\left(X_2 | X_1 < x_3, X_2 < x_4\right)} \quad (2.1)$$

[A]Limburg Institute of Financial Economics (LIFE), Maastricht University, the Netherlands
[B]LIFE, Maastricht University and Radboud University Nijmegen, the Netherlands.
[C]LIFE, Maastricht University, the Netherlands and CEPR
[*]Corresponding author; Address: Limburg Institute of Financial Economics (LIFE), Economics Faculty, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands; e-mail: c.wolff@finance.unimaas.nl.

[1]Linear dependence measures of sectoral portfolios w.r.t. oil index or other factors can also be considered but we believe that the results for one conditioning factor already sufficiently illustrate similarities and differences between both dependence measures.

with β_u and β_l referring to the CAPM-βs conditioned on the upper and lower bivariate tails, respectively. These βs can simply be obtained by running the CAPM regression with OLS over the defined subsample. We calculated these βs for two tail subsamples. First, $(x_1, x_2) = (\mu_1, \mu_2)$ with the latter pair referring to the sample means of the return data. Second, the marginal exceedance probability $p = 5\%$ implies that the lower quantiles $(x_3, x_4) < (0, 0)$ are the 5% left tail quantiles of the historical return distribution whereas the upper quantiles $(x_1, x_2) > (0, 0)$ correspond with the 95% upper quantiles of the historical returns. Lower and upper quantiles can differ because of asymmetry.

We also tested for structural change and asymmetry in β_l and β_u. Analogous to the tail-β tests in the paper, we consider two structural change tests $\left(H_0 : \beta_{l_1} = \beta_{l2} \text{ and } H_0 : \beta_{u_1} = \beta_{u2}\right)$and two asymmetry tests $\left(H_0 : \beta_{l_1} = \beta_{u_1} \text{ and } H_0 : \beta_{l_2} = \beta_{u_2}\right)$. The structural change tests take 9/11 as the sample midpoint and test for time variation in the lower and upper CAPM-β separately; the asymmetry tests test for asymmetry between β_l and β_u for the pre-9/11 and post-9/11 period separately. The equality tests are calculated using Newey-West standard errors to correct for heteroskedasticity and autocorrelation. Notice that our approach differs from the asymmetry test of Ang and Chen (2002) because we wanted to stay as close as possible to the 9/11 paper.[2]

Prior to discussing the estimation and testing results of the truncated CAPM-βs, one should realize that comparing CAPM-βs with tail-βs should be done with great care because these are two very different measures. First, the tail area on which the truncated CAPM-βs are calculated is necessarily less extreme than the tail areas we typically consider when using extreme value analysis. Whereas EVT enables one to evaluate co-exceedance probabilities beyond the historical sample

[2]Ang and Chen propose to test for asymmetry relative to some symmetric benchmark model such as the bivariate normal. For symmetric benchmark models the lower and upper CAPM-βs are equal. Thus, if e.g. the lower beta β_l significantly differs from the normality-implied value of β_l and the upper beta β_u does not significantly differ from the theoretical value under the normal, one can conclude that there is asymmetry. However, this criterion for asymmetry is always relative to some benchmark model. As such one could interpret the Ang and Chen approach as a goodness-of-fit test of the data to the tails of a (symmetric) benchmark model.

boundaries, β_l and β_u are only defined in-sample.[3] Second, CAPM-βs are linear dependence measures whereas semi-parametric estimates of co-exceedance probabilities can both capture linear and nonlinear dependence in the data. As such, the Ang-Chen framework may be ill-suited to measuring dependence for extreme tail areas because dependence in crisis periods may well be nonlinear.

The above discussion makes clear that an absolute comparison of a regression-based dependence measure and an EVT-based dependence measure (the co-exceedance probability) is relatively meaningless. However, it is interesting to know whether the ranking of the return pairs on the basis of the two systematic risk measures differs a lot. Moreover, is structural change /asymmetry more or less pronounced for the regression-based βs as compared to the tail-βs? Let us now turn to the estimation and testing results for the CAPM-βs and consider these issues.

Table 1 contains estimation results for full sample ("unconditional") βs and truncated ("conditional") CAPM-βs as defined in (2.1). The first two columns contain unconditional CAPM-βs (i.e. using all data) for the pre-9/11 and post-9/11 period whereas the other columns distinguish between lower and upper tails and the pre-9/11 and post-9/11 period. Moreover, we either truncate on the basis of the sample means (columns with header "mean-truncated") or go further into the bivariate tail (columns with header "tail-truncated"). Notice that the tail-βs for the NYSE Composite in Figure 2 of the paper are defined much further in the bivariate tail, i.e., for p=0.02%.[4] The market portfolio is chosen to be the NYSE Composite. We previously argued that a direct comparison of CAPM-βs and the tail-βs is relatively meaningless. One can, however, look at how a ranking of the sectors according to their systematic risk - as measured by CAPM-β or tail-β - differs. By visual inspection one sees that high CAPM-βs often do not correspond with high tail-βs and vice versa. Thus, the two measures often seem to tell a different story. Thus, linear dependence measures applied to the tail area tell different stories than more general dependence measures

[3]However, EVT becomes more inaccurate if one uses more data from the centre of the bivariate distribution. The Pareto tail for univariate tails and the Ledford-Tawn model for joint probabilities break down for more moderate in-sample quantiles. In general, EVT is built on limit laws for minimum/maximum returns which implies it is only suited for modelling the distributional tail.

[4]However, there are virtually no joint exceedances in this case for the historical data which implies that CAPM-βs based on the corresponding tail area would become extremely noisy and inaccurate. We therefore opted for a significance level p=5%.

such as our tail-β. As earlier suggested, the discrepancy between the two dependence measures might be due to nonlinearities in extreme dependence that cannot be captured by linear CAPM-βs.

[Insert Table 1, 2]

Table 2 contains test results for CAPM-β equality tests. Structural change and asymmetry test statistics are contained in the left and right panels, respectively. Tests are performed for the above discussed truncations of the original data. First, we see that there is abundant evidence of structural change in the full sample β ($\beta_1 = \beta_2$ column). Structural change results for truncated lower and upper βs is less significant but constancy of β_l and β_u can still be rejected in the majority of the cases (see left structural change panel). Notice that we also detected structural change in the NYSE tail-βs, albeit to a lesser extent. However, the sign of the changes differs for both dependence measures. Upon considering the statistically significant changes only, tail-βs all tend to rise after 9/11 while CAPM-βs only rise in roughly half of the cases. Second, asymmetries seem to be less common than structural change. This also holds for the asymmetry tests in our EVT framework. However, the sign of the significant asymmetries is usually reverse for both dependence measures. Whereas the significant CAPM-β asymmetries mostly indicate a dominance of upward CAPM-βs, we find the reverse for significant asymmetries between co-crashes and co-booms in an EVT framework.

3. Robustness of the heavy tail assumption

One may be concerned that the assumption of heavy or Pareto tail declines (regular variation at infinity) might not necessarily always hold. Indeed, the generalized extreme value (GEV) distribution for scaled minima or maxima nests three subclasses of models: (fat-tailed) Pareto tails, (fat-tailed) Weibull tails and (thin-tailed) Gumbel tails (see e.g. Embrechts *et al.* (1997)). Assuming that extremal financial returns are *always* regularly varying at infinity bears some misspecification risk indeed (although many predecessors made this assumption in academically published work). In order to check to what extent tails are governed by a Pareto-type decline, one can estimate the tail shape parameter γ in the GEV distribution

$$G_\gamma (x) = \exp\left(- (1 + \gamma x)^{-1/\gamma} \right),$$

and check the sign of $\widehat{\gamma}$. In case of Pareto tail decline, we know that $\gamma = 1/\alpha > 0$. The cases $\gamma = 0$ and $\gamma < 0$ correspond with the Gumbel and Weibull limit laws, respectively. The Hill estimator is unsuited for estimating the more general parameter γ because it only works under Pareto tails and always renders positive estimates *by construction*. Semi-parametric estimators based on order statistics have been developed for the general tail shape parameter γ, see e.g. Pickands (1975) or Dekkers *et al.* (1989). The sign of the estimates indicates the type of tail decline. Moreover, upon knowing the asymptotic distribution of $\widehat{\gamma}$, one can test hypotheses of the type $\gamma = 0$, $\gamma \leq 0$ or $\gamma \geq 0$. In order to get some feeling of the type of tail decline (regularly varying or not) we calculate the DEDH estimator for γ (but the analysis below does not change much if it is performed with the Pickands estimator). The DEDH estimator reads:

$$\widehat{\gamma} = M_n^{(1)} + 1 - \frac{1}{2}\left\{1 - \frac{\left(M_n^{(1)}\right)^2}{M_n^2}\right\}^{-1},$$

with

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

and

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

Notice that $X_{1,n} \leq \cdots \leq X_{j,n} \leq \cdots \leq X_{n,n}$ represent the ascending order statistics of the financial return series. One can see that $M_n^{(1)}$ is the inverse of the Hill statistic. The cutoff point m again determines the number of extremes used in estimation. Dekkers et al. established consistency and asymptotic normality for $\widehat{\gamma}$.

This estimator can now be put at work in order to check which limit law governs the extremes of financial returns. A positive $\widehat{\gamma}$ suggests fat tailed Pareto behavior whereas a negative $\widehat{\gamma}$ suggest a fat tailed Weibull. To stay consistent with the Hill calculations in the paper, we conditioned the DEDH estimator on the same value of the nuisance parameter m.[5] Table 3 contains estimation and testing results for the DEDH estimator.

[Insert Table 3]

The point estimates $\widehat{\gamma}$ (left panel) are conditioned on the same optimal m* as the Hill statistics in the paper. In the right panel we test the null hypothesis of thin tails $H_0 : \gamma = 0$ against the (one-sided) heavy tailed alternative $H_1 : \gamma > 0$. One could perform an asymptotic test using the asymptotic variance expression given in Dekkers *et al.* for the case $\gamma \geq 0$ but this test only holds for i.i.d. data. In order to stay consistent with the main body of the paper, we therefore block bootstrapped the standard deviation of $\widehat{\gamma}$ using block sizes of 50.[6]

The table distinguishes between upper and lower tails and pre- and post-9/11 tails. Negative point estimates were only detected twice; moreover, DEDH estimates consistently stay positive over a wide range of m (plots available upon request). As for the null hypothesis of thin tails it can be rejected at the 99% significance level in nearly 80% of the considered tail cases. In addition to the DEDH estimation and testing outcomes we also performed a simple Jarque-Bera normality test; normality was strongly rejected in all cases. The remaining non-rejections for the DEDH test might be due to the high asymptotic variance of the DEDH estimator and the resulting poor power properties, even in medium-sized samples. Concluding, we believe that the regular variation assumption is a defendable modelling assumption on the basis of the above evidence. If we would wrongly assume that tails are heavy when they are actually thin tailed, capital adequacy policies for the long run based on extreme quantile estimation would be too conservative.

As concerns the tail of the auxiliary variable $Z_{\min}$ (see eq. (2.7) in the paper) that summarizes the dependence structure of return pairs (X, Y), it is regularly varying *by construction* with a tail index α hovering between 1 (asymptotic dependence) and 2 (complete asymptotic

[5] At first sight, one could argue that the more general estimators like Pickands and DEDH estimator are preferable above the Hill statistic estimator because they eliminate tail model risk. However, this goes at the expense of estimation risk, i.e., their asymptotic variances are much higher than the asympotic variance of the Hill estimator. The intuition behind this result is that the Pareto model is nested into the GEV distribution. Hence, an estimator starting from the latter distributional assumption will be more imprecise because it has to encompass the three allowed limit laws.

[6] Just as in the main body of the paper, the block bootstrap is meant to take account of the temporal dependence structure in financial returns.

independence). Thus, a pretest for the tail of $Z_{\min}$ is not necessary prior to calculating the tail-βs.[7]

4. Estimation risk and asymptotic normality under EVT

In the previous section we argued that the regular variation assumption is an acceptable working hypothesis for our data. The reader might also be worried that the semi-parametric quantile estimator may perform poorly (high bias and variance) and that the normal approximation for the equality tests for structural change and asymmetry may not hold for the sample size we employ. First, the small sample performance and convergence of estimates and test statistics may be hampered by linear (1st order autocorrelation) and nonlinear temporal dependencies (stochastic volatility) in the data. From Hsing (1991), Resnick and Starica (1998) and Embrechts *et al.* (1997) we know that the Hill estimator is a consistent estimator for dependent data like the ARMA processes and ARCH-type processes but temporal dependence may still harm in small(er) samples. The fact that the quantile estimator is defined as a 1st order approximation constitutes a 2nd potential source of poor small sample performance.

4.0.1. *Linear dependence.* Embrechts et al. (1997, p. 270) argue that serial correlation of the order of 0.2 or higher leads to an upward bias in the Hill statistic (and thus a downward bias in estimated quantiles, see Figure 5.5.4). However, 1st order serial correlations in financial data are usually much weaker than the serial correlation parameters used in the Embrechts simulation ($\theta = 0.2, 0.5, 0.9$). Moreover, at lower return frequencies (daily, weekly) there is virtually no statistically and economically significant autocorrelation. At higher frequencies such as our data set, market microstructure effects might induce statistically significant but economically minor 1rst order serial correlations; but their point estimates are too small to cause large biases in the Hill estimator (the bulk of 1st order serial correlations in our data do not exceed 0.1 in absolute value - and most of them are positive).

4.0.2. *1st order approximation.* If financial returns are fat tailed - and focusing on the upper tails by taking $X^* = -X$ - the class of heavy

[7] QQ plots constitute an alternative graphical device in order to get some feeling of the nature of tail decline. All QQ plots we made become approximately linear in the tails which is indicative of Pareto tail decline. Nonsurprisingly, the linearity (and thus the start of the tail) starts much earlier for the bivariate auxiliary variables $Z_{\min}$ compared to the original raw returns.

tailed distributions exhibits the regular variation at infinity property

$$P\{X > x\} = l(x)\, x^{-\alpha},$$

with x large and where $l(x)$ is a slowly varying function.Under some mild extra conditions the class of regularly varying distributions obeys the following 2nd order expansion for large x:

$$P\{X > x\} \simeq a x^{-\alpha}\left(1 + b x^{-\beta} + o\left(x^{-\beta}\right)\right), \tag{4.1}$$

as $x \rightharpoonup \infty$ and $\alpha, \beta, a > 0$.The higher the 2nd order parameter β the quicker $l(tx)/l(x)$ converges to 1 for large x. An intuitive derivation of the quantile estimator goes as follows (Danielsson and de Vries (1997)). Let $p > 1/n$ but close to $1/n$, and $q < 1/n$, where n is the sample size and p, q stand for exceedance probabilities. We want to estimate the out of sample quantile x_q by using the empirical counterpart of the in sample quantile x_p. By using the above expansion, we get:

$$p \simeq a(x_p)^{-\alpha}\left[1 + b x_p^{-\beta}\right] \quad \text{and } q \simeq a(x_q)^{-\alpha}\left[1 + b x_q^{-\beta}\right].$$

Division of p and q and rearranging renders

$$x_q \simeq x_p \left(\frac{p}{q}\right)^{1/\alpha}\left(\frac{1 + b x_q^{-\beta}}{1 + b x_p^{-\beta}}\right)^{1/\alpha}.$$

Ignore the second term and replace x_p by its empirical counterpart $X_{n-m,n}$ $(X_{1,n} \leq \cdots \leq X_{n-m,n} \leq \cdots \leq X_{n,n})$ for which m/n is closest to p. The extreme quantile estimator then becomes:

$$x_q \simeq x_p \left(\frac{m}{np}\right)^{1/\alpha},$$

where α is estimated by the Hill statistic:

$$\widehat{\alpha} = \left(\frac{1}{m}\sum_{j=0}^{m-1} \ln\left(\frac{X_{n-j,n}}{X_{n-m,n}}\right)\right)^{-1}, \tag{4.2}$$

and with m the number of highest order statistics used in the estimation. Clearly, whether the factor $\left(\frac{1+bx_q^{-\beta}}{1+bx_p^{-\beta}}\right)^{1/\alpha}$ is negligible or not will depend on the magnitudes of the parameter values α, β, and b.

4.0.3. *Optimal threshold selection.* Tail index estimators like the Hill statistic are characterized by a bias/variance trade-off. The more data used from the distributional centre the smaller will be the variance of the estimator but the more bias will be introduced. Goldie and Smith

(1987) therefore proposed to select m such as to minimize the asymptotic Mean Squared Error (AMSE). Upon assuming the 2nd order expansion (4.1), Danielsson and de Vries (1997) derive an expression of the AMSE in terms of the 2nd order expansion parameters:

$$AMSE(\widehat{\alpha}, m) = a^{-2\beta/\alpha} \frac{1}{\alpha^2} \frac{\beta^2 b^2}{(\alpha+\beta)^2} (\frac{m}{n})^{\frac{2\beta}{\alpha}} + \frac{1}{\alpha^2 m}, \tag{4.3}$$

where the first part is the squared bias and the second part is the variance. Minimizing (4.3) w.r.t. m renders the optimal number $m^*$of highest order statistics

$$m^* = cn^{2\beta/2\beta+\alpha}, \qquad c = \left(\frac{\alpha(\alpha+\beta)^2}{2\beta^3 b^2} a^{2\beta/\alpha} \right)^{\alpha/2\beta+\alpha} \tag{4.4a}$$

For each of the data generating processes below (except GARCH(1,1)) we derived the parameter vector (α, β, a, b) and the accompanying analytic selection criterion (4.4a). This highly simplifies the Monte Carlo simulation. As for the GARCH(1,1) process the parameters of the 2nd order expansion are unknown and we have to resort to the Beirlant *et al.* (1999) aproach to estimating the empirical AMSE.

4.0.4. *Data generating processes and results.* Monte Carlo simulations are carried out for a large set of regularly varying distributions. We choose the data generating processes such as to get a large cross section of different parameter values (a, b, α, β) in (4.1). The distributional models chosen for the Monte Carlo study were student-t, Frechet, i.e. $P\{X > x\} = \exp(-x^{-\alpha})$, Burr, i.e. $P\{X > x\} = \left(1 + x^\beta\right)^{-\alpha/\beta}$, symmetric stable df, AR(1) with stable innovations, GARCH(1,1) and a Stochastic Volatility model. So, we clearly distinguished between models that generate i.i.d. data and dependent models. The tail index equals the degrees of freedom parameter v $(\alpha = v)$ in case of the Student-t. The tail index of the Student-t, Frechet, Burr and the Stochastic Volatility model is varied between 2 and 4, the range one typically observes in empirical applications. For generating symmetric stable draws we used the algorithm proposed by Samorodnitsky and Taqqu (1994):

$$X_{stable} = \frac{\sin \phi\gamma}{(\cos \gamma)^{1/\phi}} \left(\frac{\cos(1-\phi)\gamma}{W} \right)^{(1-\phi)/\phi} \tag{4.5}$$

where $0 < \phi \leq 2$ represents the characteristic exponent. The parameter γ is drawn uniformly on $[-\pi/2 \ \pi/2]$ whereas W is exponentially distributed with mean 1. One has to be careful when linking the value of the characteristic exponent to the amount of probability mass in the tails of the stable df. Let α stand for the tail index as defined earlier. For $\phi < 2$, $\alpha = \phi$; but when $\phi = 2$, the distribution becomes normal and all moments exist which implies $\alpha = \infty$.

In addition to i.i.d. data, we also investigate the performance of our estimator for dependent data. We focus on three stochastic processes that are characterized by either dependence in the 1st or the 2nd distributional moment: the first one exhibits 1st order serial correlation and the other two exhibit conditional heteroscedasticity. Serially correlated data are generated using an AR(1) process with $\rho = 0.1$ and symmetric stable innovations. The additivity property for the innovations implies that the serially correlated draws will also be stable and hence one can select the nuisance parameter m optimally using the parameters of the 2nd order expansion (4.1) for the symmetric stable df. A second stochastic process that we use as a simulation vehicle for dependent data is taken from Danielsson *et al.* (2001). It constitutes a rudimentary stochastic volatility model characterized by volatility persistence:

$$\begin{aligned} Y_t &= U_t W_t H_t, & P\{U_t = -1\} = P\{U_t = 1\} = 0.5 \\ H_t &= \beta Q_t + \gamma H_{t-1}, & Q_t \sim N(0,1), \ \ \beta = 0.1, \ \ \gamma = 0.9 \\ W_t &= \sqrt{\frac{1-\gamma^2}{\beta^2}} \frac{\sqrt{v}}{\sqrt{Z_t}}, & Z_t \sim \chi(v). \end{aligned}$$

The multiplicative factor U_t guarantees the fair game property (without this factor, the model both exhibits dependence in the 1rst and the second moment). This process generates volatility clusters in Y_t and is designed such that Y_t is Student-t distributed with v degrees of freedom. Hence, we can determine the theoretical second order parameters for the expansion of regularly varying tails and use these to calculate the optimal nuisance parameter m for the Hill statistic and the quantile estimator. Finally, we simulated form a GARCH(1,1) process, by using standard normal innovations, an intercept of 10^{-6} in the variance equation and coefficients for the lagged squared return and variance term of 0.31 and 0.59, respectively. These parameters were chosen such that they sum up to 0.9 (reflecting the stylized fact of persistence). These values also correspond to a theoretical tail index of $\alpha = 4$.

The second order limit expansion (4.1) is unknown for the GARCH process but it is known for all the other dfs. The theoretical β value for the Student-t is $\beta = 2$ and for the Frechet and symmetric stable it is $\beta = \alpha$. As for the Burr, β can be chosen independent from α. Thus the latter distribution is perfectly suited to study what happens to the small sample properties of the estimates and test statistics if one changes the 2nd order parameter while leaving the 1st order parameter

α fixed. We simulated from the Burr distribution for values of β equal to 2 and 4. Upon knowing the parameters of the 2nd order expansion, the optimal m^* can now be analytically calculated by eq. (4.4a). Notice that we limit ourselves to univariate simulation experiments because the whole estimation and testing framework for the bivariate case is mapped back into a univariate framework anyway in the Ledford-Tawn approach.[8]

Table 4 reports simulation results for the average and standard deviation of the Hill statistic and accompanying quantile estimator.

[Insert table 4]

The table reports true values of the quantile q_p (for those cases for which they can be determined analytically) as well as the average and standard deviation of α and q_p for 5,000 Monte Carlo replications and for samples of 8,000 draws; this reflects the length of the pre-9/11 and post-9/11 empirical financial return series (n=8,380). The quantiles are calculated for two values of the marginal exceedance probability p. Not surprisingly, the bias in the Hill estimator and corresponding quantiles increases with α. Moreover, the standard errors also increase when tails get thinner. The intuition behind this result is that lighter tails are closer to a thin tailed local alternative like the normal distribution. This decreases the accuracy of tail estimation techniques that assume regular variation as a starting point. Also, and conform to our a priori intuition, the quantile estimates become less precise the further one looks into the tail. Apart from comparing bias and standard deviation across different values of α it is also worth noticing what happens when the 2nd order parameter β changes for given fixed values of α. Only in the Burr distribution case, we can let β evolve independently from α. The results clearly reveal that the bias and standard error decrease for higher values of β,i.e., the quicker the 2nd order term in the expansion (4.1) converges to zero the smaller will be the estimation risk. Finally, notice that the bias and variance properties for dependent data (AR(1), stochastic volatility model and GARCH(1,1)) are not dramatically different from the i.i.d. results. Summarizing, the quantile estimator

[8]Bivariate estimation risk for tail-βs has already been assessed previously by Hartmann *et al.* (2005). They applied the same EVT framework as ours to daily bank stock returns in order to calculate tail-βs w.r.t. macro factors and interbank co-exceedance probabilities. Monte Carlo simulations for a sample of size n$\approx 3,000$ render reasonable results for the bias and variance of the tail dependence parameter and tail-β and for a number of bivariate parametric models (bivariate Normal, bivariate Pareto and bivariate Gumbel-Pareto).

performs relatively well when judged by the mean and standard deviation. Moreover, the performance of the quantile estimator is fairly consistent across distributions and stochastic processes.

Table 5 reports simulation results for the critical values of the equality tests in the paper's empirical section. The tests are used for detecting structural change/asymmetry in tail indices (tail dependence parameters) and tail quantiles (tail-βs).

[Insert table 5]

The table reports the benchmark values of the normal quantile (top row) as well as the critical values for 5,000 Monte Carlo replications and for samples of 16,000 draws (equal subsamples of 8,000 draws). The same data generating processes are employed as for assessing estimation risk. The table shows that size distortions are small for all unconditional models characterized by i.i.d. draws. Critical values for the AR(1) process and the persistent GARCH(1,1) process suggests that our tests may be size distorted, especially the quantile equality test. However, a large part of testing outcomes in the paper's empirical section would remain significant upon application of size-corrected testing procedures because a lot of t-statistics in the empirical section are bigger than 3 or 4 in absolute value! An avenue of future research could be to fit an intraday version of a GARCH or other Stochastic Volatility model and to use this as a vehicle to simulate or bootstrap small sample critical values.

5. Robustness to GARCH

One of the authors did filter for GARCH in a previous paper on systemic risk in the U.S. vs. the European banking sector, see Hartmann *et al.* (2005). In that paper, basically the same type of co-exceedance probabilities were estimated; the co-exceedance probabilities between banks or w.r.t aggregate shocks were tested for endogenous structural change (no exogenously fixed breakpoint as in the 9/11 paper) using the Quintos *et al.* (2001) technique. The structural change tests were complemented by cross sectional tests comparing the magnitude of the co-exceedance probabilities across different pairs of banks. Hartmann *et al.* (2005) both presented raw and GARCH filtered results. Filtering for GARCH, they observe that the co-exceedance probabilities nearly all decrease; however, they are still way above the marginal exceedance probability p, the value the co-exceedance probability would take under complete independence. Thus, GARCH seems to induce part - but not all - of the extreme dependence that determines the level of the co-exceedance probabilities. Second, the break results and cross sectional tests do change upon filtering but not in a dramatic way. Only a minor part of the testing results change and breaks are still abundantly present for the filtered data pairs. Part of the breaks disappear. In a version of the 9/11 paper that even preceded the 1st submission we did filter for GARCH and the effects were comparable to the described Hartmann *et al*. observations. However, the problem of applying the Poon *et al.* (2004) and Hartmann *et al.* (2005) GARCH filters lie in the intraday nature of our data (Poon *et al.* and Hartmann *et al.* worked with daily data). It is by now generally accepted that modelling of intraday volatility is much more complex than daily volatility modelling. More specifically, it would also require the modelling of intraday seasonal effects on the volatility which is a nontrivial exercise, see e.g. Andersen and Bollerslev (1997, 1998) or Bollerslev, Cai and Song (2000). We think that modelling the intraday volatility seasonalities - although very interesting - lies outside the scope of this paper.

Also, notice that our tests should be able to disentangle changes in the unconditional tail features from clusters of high and low conditional volatility (given a stationary stochastic volatility model). We therefore estimate the asymptotic variances of the tail index and quantile estimates by means of a block bootstrap that captures the nonlinear dependence in the data. Indeed, the parameters governing the unconditional distribution (scale, tail index) should necessarily be related to the parameters governing the conditional distribution (e.g. GARCH and seasonal volatility parameters). However, there is no closed form solution known of this relation except in some very simple cases.[9] Consequently, changes in the tail index/quantile necessarily reflect changes in the GARCH or seasonal volatility parameters. How exactly the conditional parameters are changing is unknown and would require a full fledged intraday GARCH model together with structural breaks tests and asymmetry tests.

6. Details on the block bootstrap

The test statistics in eqs. (2.11) and (2.12) the paper were calculated using (block) bootstrapped standard deviations. The blocks are meant to take account of temporal dependence in the data. If financial returns would be completely independent over time we could have exploited the fact that $m^{1/2}(\widehat{\alpha}-\alpha) \to N(0,\alpha^2)$ and $\frac{\sqrt{m}}{\ln\left(\frac{m}{pn}\right)} \ln \frac{\widehat{q}(p)}{q(p)} \to N(0,\gamma^2)$

[9]Closed forms of the relation between conditional and unconditional parameters are only known for simple cases like the ARCH(1) or GARCH(1,1) models with conditionally normal innovations.

and estimate the asymptotic variances of the left-hand side expressions with $\widehat{\alpha}^2$ and $\widehat{\gamma}^2 = 1/\widehat{\alpha}^2$, respectively. However, financial returns, and especially high frequency data like ours, may exhibit strong nonlinear dependencies which may change the asymptotic variance (it is probably upward biased by the temporal dependence although this should not necessarily be the case). It can be shown that consistency and asymptotic normality still holds if data are temporally dependent, albeit with a different asymptotic variance, see e.g. Drees (2002). The latter author suggests estimators for this asymptotic variance under general (nonspecified) dependence but this asymptotic variance can only be used if one wants to test the estimated tail index or quantile values against a specific value, i.e., $H_0 : \alpha = \alpha_0$ or $H_0 : q = q_0$. For the structural change and asymmetry tests at hand ($H_0 : \alpha_1 = \alpha_2$ and $H_0 : q_1 = q_2$), however, estimates of the asymptotic variances alone are insufficient because one also needs the covariance terms for (α_1, α_2) and (q_1, q_2) in order to calculate standard deviations for $(\widehat{\alpha}_1 - \widehat{\alpha}_2)$ and $(\widehat{q}_1 - \widehat{q}_2)$. This covariance may both be present in the structural change test (due to temporal dependence) as well as the asymmetry test (due to cross sectional dependence). The block bootstrap seems a straightforward way to estimating the standard deviation of the difference. However, the application of nonparametric bootstraps (let alone block bootstraps) in EVT analysis is still in its infancy and clear-cut techniques for the determination of the optimal block length are unknown as to date.[10] In order to get an idea of the optimal block length's magnitude in our case, we simulated from a GARCH(1,1) with $\beta_0 + \beta_1 \approx 1$ (persistent; close to nonstationarity). Making abstraction of intraday seasonalities - that also play a role in driving the volatility at the intraday frequency - we might consider this model as a crude approximation of our data. The GARCH model has the advantage that it is known how to estimate the scaling factor of the Hill estimator's and quantile estimator's variance, i.e., *scale* such that $\sigma^2(\widehat{\alpha}) = scale \times \widehat{\alpha}^2$, see e.g. Hsing (1991). In the simulations, we always found estimates of this scaling factor close to 2 for $\beta_0 + \beta_1 \approx 1$. Moreover, the scaling factor seemed invariant upon varying the value of the sum $\beta_0 + \beta_1$ from 0.9 to 0.95 till 0.99 and upon changing the trade-off between (β_0, β_1) for $\beta_0 + \beta_1 = c$ when c is fixed. By trial and error, we next searched

[10]Hall *et al.* (1995) determine the optimal block length as a function of the sample size for some specific parametric examples with temporal dependence. They advice to set block lengths equal to either $aT^{1/3}$ or $aT^{1/4}$ where T is the sample size and a some scaling factor that depends on the model at hand. Neither of their considered parametric examples, however, is relevant for finance and they also do not give estimators for the exponent and the scaling factor a.

for the optimal block length for this GARCH process, i.e., the block length that can reproduce variances for the Hill statistic that are approximately twice as high as their theoretical values under absence of GARCH. We were able to get quite good results for block lengths around 50-75 (the former value is used).

Table 1. Unconditional and conditional (truncated) CAPM beta's

Indices	full sample		mean				tail (p=5%)			
	< 9/11	> 9/11	<9/11		>9/11		<9/11		>9/11	
			l	*u*	*l*	*u*	*l*	*u*	*l*	*u*
INDU	1.160	1.159	1.1	1.1	1.110	1.112	1.108	1.018	1.040	1.020
TRANS	0.680	1.019	0.598	0.566	0.991	0.900	0.321	0.378	0.882	0.733
UTIL	0.573	0.818	0.542	0.572	0.908	0.900	0.362	0.534	0.811	1.149
PC	2.181	1.567	1.673	2.039	1.304	1.339	0.955	1.928	0.698	0.838
BIO	1.859	1.562	1.571	1.700	1.247	1.281	0.732	1.780	0.285	0.502
INSUR	0.440	0.612	0.385	0.342	0.527	0.566	0.220	0.043	0.246	0.525
TEL	1.942	1.532	1.549	1.825	1.275	1.299	0.973	1.726	0.647	0.788
BANK	0.544	0.657	0.477	0.507	0.589	0.639	0.393	0.508	0.343	0.594
FIN	0.806	0.822	0.728	0.745	0.738	0.814	0.634	0.715	0.482	0.776
OFIN	1.215	0.878	1.172	1.075	0.747	0.763	1.009	0.754	0.380	0.501
INTER	2.568	1.673	1.963	2.510	1.380	1.403	1.251	2.583	0.527	0.741
PHARMA	0.864	0.981	0.805	0.727	0.910	0.932	0.630	0.494	0.782	0.770
AIR	0.654	1.248	0.581	0.536	1.207	0.950	0.300	0.411	1.002	0.409
OIL	0.577	0.782	0.510	0.502	0.763	0.674	0.322	0.236	0.684	0.480
SCAP	0.694	0.839	0.672	0.627	0.763	0.725	0.571	0.529	0.322	0.568
MCAP	0.977	0.953	0.866	0.873	0.899	0.858	0.651	0.771	0.598	0.709
GROWTH	1.481	1.169	1.331	1.397	1.098	1.096	1.133	1.267	0.970	0.924
VALUE	1.026	1.174	0.969	0.994	1.142	1.140	0.865	0.988	1.051	1.086

Note: The table reports estimates of nontruncated and truncated CAPM beta's. We distinguish between pre-9/11 and post-9/11 subsamples. The nontruncated estimates are based on all observations in these sample periods. "Mean-truncated" results are based on return data below (above) their sample averages (denoted by l and u, respectively). "Tail-truncated" results are based on the 0.05% smallest (largest) return observations (denoted by l and u, respectively).

Table 2. 9/11 structural change/asymmetry tests for CAPM beta's

indices	structural change					asymmetry			
		mean		tail(p=5%)		mean		tail (p=5%)	
	$\beta_1 = \beta_2$	$l_1 = l_2$	$u_1 = u_2$	$l_1 = l_2$	$u_1 = u_2$	$l_1 = u_1$	$l_2 = u_2$	$l_1 = u_1$	$l_2 = u_2$
IND	0.09	-0.46	-0.49	0.94	-0.03	0	-0.09	1.20	0.2
TRAN	***-13.23	***-5.69	***-8.62	***-2.61	***-3.78	0.83	1.32	-0.71	0.6
UTIL	***-5.77	***-4.94	***-2.95	***-2.01	***-2.55	-0.55	0.07	-1.37	-1.1
PC	***11.77	***4.82	***5.93	0.87	***2.91	***-2.77	-0.73	**-2.18	-0.8
BIO	***4.78	***3.22	***2.89	*1.79	***3.18	-0.83	-0.40	***-2.35	-1.3
INSUR	***-8.13	***-3.68	***-4.74	-0.20	***-4.38	0.91	-1.00	1.41	***-2.4
TEL	***8.34	***3.42	***4.68	1.15	***3.64	***-2.34	-0.34	**-2.22	-0.7
BANK	***-7.17	***-3.42	***-3.31	0.65	-0.82	-0.67	-1.96	-0.99	***-4.1
FIN	-0.92	-0.31	***-1.79	1.39	-0.71	-0.41	***-2.66	-0.66	***-4.5
OFIN	***7.84	***5.45	***3.96	***2.43	1.47	0.92	-0.49	0.86	-1.3
INTER	***13.79	***6.04	***6.69	**2.25	***4.47	**-3.28	-0.24	***-2.74	-1.1
PHARMA	***-4.77	***-2.27	***-3.63	-1.02	**-1.97	1.29	-0.54	0.83	0.0
AIR	***-14.10	***-4.64	***-5.25	-1.45	0.01	0.88	*1.74	-0.79	1.2
OIL	***-8.59	***-5.97	***-4.53	***-2.91	***-2.84	0.20	**2.24	1.14	1.5
SCAP	***-6.79	***-2.56	***-2.54	***2.52	-0.38	0.96	1.59	0.35	***-3.1
MCAP	1.35	-0.20	0.43	0.43	0.71	-0.17	0.25	-0.85	*-1.9
GROWTH	***14.57	***9.08	***5.90	**2.04	***2.64	-1.32	0.07	-1.01	0.6
VALUE	***-15.43	***-11.30	***-8.32	***-5.06	***-2.49	-1.33	0.14	***-3.12	-0.9

Note: Structural change tests for the full (non-truncated) pre-9/11 and post-9/11 subsamples are contained in column 1. The rest of the table reports truncated (subsample) structural change and asymmetry tests. "Mean-truncated" testing results are based on return data below or above sample averages (denoted by l and u, respectively). "Tail truncated" testing results are based on the 0.05% smallest (largest) return observations (denoted by l and u, respectively). The tests are calculated using Newey-West standard errors robust to heteoskedasticity and autocorrelation. Moreover, the tests are normally distributed and two-sided rejections at the 10, 5 and 2 percent significance level are denoted by *, ** and ***, respectively.

Table 3. Dekkers-Einmahl-de Haan estimates of the tail index: estimation and testing results

Indices	DEDH $\widehat{\gamma}$				$H_0 : \gamma = 0$			
	<9/11		>9/11		<9/11		>9/11	
	l	u	l	u	l	u	l	u
INDU	0.279	0.266	0.155	0.183	***3.361	***2.409	**2.206	***2.817
TRANS	0.177	0.190	0.346	0.228	***3.157	***2.376	***3.298	***4.351
UTIL	0.301	0.219	0.354	0.382	***4.174	**2.250	***7.310	***5.843
PC	0.184	0.303	0.141	0.152	*1.806	***4.236	**1.977	*1.773
BIO	0.194	0.181	0.019	0.095	1.171	*1.182	0.178	1.203
INSUR	0.333	0.290	0.100	0.262	***2.660	***3.408	1.104	***3.549
TEL	0.205	0.302	0.176	0.097	***2.763	***4.458	***2.546	0.846
BANK	0.207	0.284	0.022	0.250	***3.336	***3.931	0.173	***3.898
FIN	0.187	0.205	-0.114	0.258	1.634	***2.615	-0.824	***4.301
OFIN	0.285	0.262	0.098	0.204	**2.028	***3.691	1.387	***2.464
INTER	0.259	0.317	0.122	0.170	***2.107	***5.488	1.567	***2.199
PHARMA	0.234	0.164	0.175	0.174	***3.910	**2.254	**2.147	**1.983
AIR	0.220	0.173	0.409	0.170	***3.647	*1.895	***3.109	**2.104
OIL	0.179	0.200	0.225	0.132	***2.366	***3.707	***3.410	1.558
SCAP	0.209	0.262	0.010	0.227	***3.111	***3.294	0.146	***3.615
MCAP	0.233	0.265	-0.010	0.153	**2.220	***4.069	-0.070	1.558
GROWTH	0.206	0.283	0.141	0.163	1.439	***4.912	*1.692	**2.194
VALUE	0.212	0.320	0.127	0.216	***2.375	***3.012	1.299	***3.582
NYCOMP	0.246	0.293	0.140	0.185	***2.664	**2.212	1.392	***2.245

Note: The table reports estimates of the Dekkers-Einmahl-De Haan (DEDH) estimator and corresponding co-exceedance probabilities for the pre-9/11 and post-9/11 subsample and for the lower (l) and upper (u) bivariate tail separately. The co-exceedance probabilities are calculated for a marginal significance level p=0.02%

Table 4. Tail index and quantile estimation risk for representative data generating processes

	Tail index $\widehat{\alpha}$		Quantile $\widehat{q}_p$					
			$p = 0.05\%$			$p = 0.02\%$		
	aver.	s.e.	aver.	s.e.	q_p	aver.	s.e.	q_p
Student$(\alpha = 2)$	1.90	0.12	34.01	4.75	31.60	55.40	9.44	49.98
Student$(\alpha = 4)$	3.57	0.42	8.82	0.88	8.61	11.48	1.48	10.91
Frechet$(\alpha = 2)$	1.94	0.07	47.33	4.64	44.71	76.00	8.72	70.71
Frechet$(\alpha = 4)$	3.90	0.13	6.85	0.33	6.69	8.67	0.48	8.41
Burr$((\alpha, \beta) = (2, 2))$	1.94	0.08	47.07	5.20	44.71	75.72	9.92	70.70
Burr$((\alpha, \beta) = (4, 2))$	3.67	0.31	6.80	0.57	6.61	8.75	0.92	8.35
Burr$((\alpha, \beta) = (2, 4))$	1.97	0.05	46.15	3.57	44.72	73.48	6.55	70.71
Burr$((\alpha, \beta) = (4, 4))$	3.88	0.17	6.86	0.38	6.69	8.70	0.58	8.41
Stab$(\alpha = 1.2)$	1.23	0.06	189.42	37.98	-	402.86	97.09	-
Stab$(\alpha = 1.5)$	1.60	0.15	52.54	11.26	-	94.71	25.62	-
AR$((\rho, \alpha) = (0.1, 1.2))$	1.25	0.07	188.67	41.33	-	397.02	104.07	-
AR$((\rho, \alpha) = (0.1, 1.5))$	1.66	0.16	49.30	11.04	-	86.99	24.50	-
SV$((\alpha, \beta, \gamma) = (2, 0.1, 0.9))$	1.95	0.12	32.32	4.46	31.60	52.03	8.68	49.98
SV$((\alpha, \beta, \gamma) = (4, 0.1, 0.9))$	3.76	0.45	8.43	0.83	8.61	10.83	1.37	10.91
GARCH$((\alpha, \beta_0, \beta_1) = (4, 0.3, 0.59))$	3.96	0.66	1.80E-2	3.76E-3	-	2.31E-2	6.51E-3	-

Note: The table reports average estimated values and true (analytic) values of the tail index and tail quantile. Averages are calculated for samples of n=8,000 and for 250 replications. The optimal number of highest order statistics m is determined analytically for all unconditional models by minimizing the Asymptotic Mean Squared Error (AMSE). For Garch(1,1) models we choose m on basis of Hill plots and the Beirlant et al. (1999) method. The univariate quantiles are calculated for marginal exceedance probabilities equal to 0.05% and 0.02%.

Table 5. Small sample critical values for (univariate) tail index and tail quantile equality test

	p=1%	p=2.5%	5%	p=95%	p=97.5%	p=99%
as. nor.	-2.33	-1.96	-1.64	1.64	1.96	2.33
			Panel A: $H_0 : \alpha_1 = \alpha_2$			
Student($\alpha = 2$)	-2.27	-1.90	-1.61	1.60	1.91	2.24
Student($\alpha = 4$)	-2.24	-1.89	-1.57	1.60	1.92	2.27
Frechet($\alpha = 2$)	-2.33	-1.90	-1.61	1.62	1.92	2.31
Frechet($\alpha = 4$)	-2.33	-1.93	-1.64	1.61	1.95	2.29
Burr($(\alpha, \beta) = (4, 2)$)	-2.17	-1.85	-1.55	1.60	1.85	2.17
Burr($(\alpha, \beta) = (2, 2)$)	-2.26	-1.95	-1.63	1.61	1.93	2.25
Burr($(\alpha, \beta) = (4, 4)$)	-2.25	-1.91	-1.61	1.59	1.91	2.21
Burr($(\alpha, \beta) = (2, 4)$)	-2.20	-1.93	-1.65	1.66	1.97	2.28
Stab($\alpha = 1.2$)	-2.35	-1.99	-1.66	1.66	2.00	2.30
Stab($\alpha = 1.5$)	-2.50	-1.98	-1.69	1.66	1.98	2.31
AR($(\rho, \alpha) = (0.1, 1.2)$)	-2.51	-2.10	-1.76	1.75	2.09	2.49
AR($(\rho, \alpha) = (0.1, 1.5)$)	-2.41	-2.04	-1.75	1.68	2.01	2.36
SV($\alpha, \beta, \gamma) = ((2, 0.1, 0.9)$)	-2.22	-1.89	-1.58	1.60	1.90	2.17
SV($(\alpha, \beta, \gamma) = (4, 0.1, 0.9)$)	-2.20	-1.83	-1.54	1.53	1.81	2.17
GARCH($(\alpha, \beta_0, \beta_1) = (4, 0.3, 0.59)$)	-2.33	-1.96	-1.63	1.76	2.08	2.40
			Panel B: $H_0 : q_1 = q_2$			
student ($\alpha = 2$)	-2.32	-1.95	-1.67	1.62	1.90	2.26
student ($\alpha = 4$)	-2.39	-1.91	-1.64	1.61	1.94	2.32
Frechet ($\alpha = 2$)	-2.33	-1.97	-1.64	1.63	1.92	2.34
Frechet ($\alpha = 4$)	-2.30	-1.95	-1.65	1.60	1.95	2.33
Burr($(\alpha, \beta) = (4, 2)$)	-2.25	-1.93	-1.63	1.57	1.86	2.27
Burr($(\alpha, \beta) = (2, 2)$)	-2.29	-1.94	-1.64	1.64	1.95	2.25
Burr($(\alpha, \beta) = (4, 4)$)	-2.27	-1.90	-1.61	1.63	1.94	2.26
Burr($(\alpha, \beta) = (2, 4)$)	-2.32	-1.99	-1.66	1.66	1.94	2.22
Stable($\alpha = 1.2$)	-2.39	-2.00	-1.67	1.72	1.99	2.37
Stable($\alpha = 1.5$)	-2.36	-2.03	-1.74	1.73	2.08	2.52
AR($(\rho, \alpha) = (0.1, 1.2)$)	-2.60	-2.15	-1.83	1.86	2.21	2.60
AR($(\rho, \alpha) = (0.1, 1.5)$)	-2.56	-2.13	-1.81	1.82	2.11	2.52
SV($\alpha, \beta, \gamma) = ((2, 0.1, 0.9)$)	-2.23	-1.93	-1.62	1.61	1.93	2.25
SV($(\alpha, \beta, \gamma) = (4, 0.1, 0.9)$)	-2.14	-1.85	-1.59	1.63	1.92	2.30
GARCH($(\alpha, \beta_0, \beta_1) = (4, 0.3, 0.59)$)	-2.82	-2.39	-2.02	1.93	2.27	2.68

Note: The table reports small sample critical values for the tail index equality test (Panel A) and the tail quantile equality test (panel B). Test statistics are based on samples of size n=8,000 and small sample critical values are obtained as quantiles averaged over 5,000 replications. The optimal number of highest order statistics m is determined analytically for all unconditional models by minimizing the Asymptotic Mean Squared Error (AMSE). For Garch(1,1) models we set m using Hill plots and the Beirlant et al. (1999) method. The standard deviation in the test statistics' denominators is either determined via a wild bootstrap (unconditional i.i.d. models) or a bootstrap with block size 50 (GARCH(1,1)). The quantile equality test is performed for a marginal exceedance probability of 0.02%.

References

[1] Andersen, TG, Bollerslev, T. 1997. Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance* **4**: 115-198.

[2] Andersen, TG, Bollerslev, T. 1998. Towards a unified framework for high and low frequency return volatility modeling. *Statistica Neerlandica* **52**(3): 273-302.

[3] Ang A, Chen J. 2002. Asymmetric correlations of equity portfolios. *Journal of Financial Economics* **63**(3): 443-494.

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

[5] Bollerslev, T. Cai, J. Song, F. 2000. Intraday periodicity, long memory volatility, and macroeconomic announcement effects in the US Treasury bond market. *Journal of Empirical Finance* **7**: 37-55.

[6] Danielsson J, de Vries CG. 1997. Tail index and quantile estimation with very high frequency data. *Journal of Empirical Finance* **4**: 241-257.

[7] Danielsson J, de Haan L, Peng L, de Vries CG. 2001. Using a Bootstrap Method to Choose the Sample Fraction in Tail Index Estimation. *Journal of Multivariate Analysis* **76**: 226-248.

[8] Dekkers ALM, Einmahl JHJ, de Haan L. 1989. A Moment Estimator for the Index of an Extreme-Value Distribution. *The Annals of Statistics* **17**(4): 1833-1855.

[9] Drees H. 2002. Extreme Quantile Estimation for Dependent Data with Applications to Finance. *Bernouilli* **9**: 617-657.

[10] Embrechts P, Klüppelberg C, Mikosch T. 1997. *Modelling Extremal Events.* Springer: Berlin.

[11] Goldie CM, Smith R. 1987. Slow variation with remainder: Theory and applications. *Quarterly Journal of Mathematics* **38**: 45-71.

[12] Hall P, Horowitz J, Jing B. 1995. On blocking rules for the bootstrap with dependent data. *Biometrika* **82**(3): 561-574.

[13] Hartmann P, Straetmans S, de Vries CG. 2005. Banking System Stability: A Cross-Atlantic Perspective. NBER Working Paper nr. 11698.

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

[15] Pickands III J. 1975. Statistical inference using extreme order statistics. The Annals of Statistics **3**(1): 119-131.

[16] Poon SH, Rockinger M, Tawn J. 2004. Extreme Value Dependence in Financial Markets: Diagnostics, Models, and Financial Implications. *Review of Financial Studies* **17** (2): 581-610.

[17] Quintos C, Fan Z, Phillips PCB. 2001. Structural Change Tests in Tail Behavior and the Asian Crisis. *Review of Economic Studies* **68**(3): 633-663.

[18] Resnick S, Stărică C. 1998. Tail index estimation for dependent data. *Annals of Applied Probability* **8**: 1156-1183.

[19] Samorodnitsky G, Taqqu M. 1994. *Stable Non-Gaussian Random Processes.* Chapman and Hall: New York.