Extreme Value Theory versus traditional GARCH approaches applied to financial data: a comparative evaluation[1]

Dolores Furió
Department of Financial Economics
University of Valencia (Spain)
Avenida de los Naranjos, s/n, 46022 Valencia (Spain)
Tel.: +34-963-828-369; Fax: +34-963-828-370
m.dolores.furio@uv.es
Francisco J. Climent
Department of Financial Economics
University of Valencia (Spain)
Avenida de los Naranjos, s/n, 46022 Valencia (Spain)
f.jose.climent@uv.es

Abstract

Although stock prices fluctuate, the variations are relatively small and are frequently assumed to be normal distributed on a large time scale. But sometimes these fluctuations can become determinant, especially when unforeseen large drops in asset prices are observed that could result in huge losses or even in market crashes. The evidence shows that these events happen far more often than would be expected under the generalized assumption of normal distributed financial returns. Thus it is crucial to properly model the distribution tails so as to be able to predict the frequency and magnitude of extreme stock price returns. In this paper we follow the approach suggested by McNeil and Frey (2000) and combine the GARCH-type models with the Extreme Value Theory (EVT) to estimate the tails of three financial index returns DJI, FTSE 100 and NIKKEI 225 representing three important financial areas in the world. Our results indicate that EVT-based conditional quantile estimates are much more accurate than those from conventional AR-GARCH models assuming normal or Student's t-distribution innovations when doing out-of-sample estimation (within the in-sample estimation, this is so for the right tail of the distribution of returns).

Key words: conditional extreme value theory; tails estimation; backtesting

[1] Financial support from CICYT project ECO2009-14457-C04-04 and the Cátedra Finanzas Internacionales-Banco Santander is gratefully acknowledged.

Extreme Value Theory and conventional methods applied to financial data: a comparative evaluation

1. Introduction

Although stock prices fluctuate, the variations are relatively small and are frequently assumed to be normal distributed on a large time scale. But sometimes these fluctuations can become determinant, especially when unforeseen large drops in asset prices are observed that could result in huge losses or even in market crashes. Besides, based on the quite generalized assumption of the normal distribution for financial returns, these "extreme" variations are expected to occur with an almost negligible probability. The reason is that the normal density function has exponentially decaying tails which assign very small probability to values far from the mean of the distribution. Thus, for instance, with independent realizations that are observed once a day, we should not expect a "4-sigma event" occurring with a frequency lower than 86 years, nor a "7-sigma event" with a frequency lower than 56 times the age of the universe i.e. 13.7 millions of years (Dowd et al. 2008). Of course, the evidence shows that these events happen far more often than would be expected under this assumption.

Therefore, the key is how to distinguish between extreme and non-extreme events. With the aim of answering this question, it is crucial to properly model the distribution tails so as to be able to predict the frequency and magnitude of extreme stock price returns. Moreover, as the extreme (price fluctuations) events will be defined as those exceeding a predetermined threshold, determining such a threshold becomes an essential step in embracing the analysis.

In this paper we use the Extreme Value Theory (EVT) to estimate the tails of three financial index returns. The modeling of extreme events is the central issue in EVT and the main purpose of the theory is to provide asymptotic models for the tails of a distribution. This theory has been increasingly playing a role in many research areas such as hydrology and climatology where extreme events are not infrequent and can involve important negative (or positive) consequences and, more recently, there has been a number of extreme value studies in the finance literature. Some examples include Embrechts et al. (1999), who present a broad basis for understanding the extreme value theory with applications to finance and insurance; Liow (2008), who compares the extreme behaviour of securitized real state and equity market indices representing Asian, European and North American markets; Danielsson and de Vries (1997), who test the predictive performance of various VaR[2] methods for simulated portfolios of seven US stocks concluding that EVT is particularly accurate as tails become more extreme whereas the conventional variance-covariance and the historical simulation methods under- and over-predict losses, respectively; similar results are found in Longin (2000)[3], Assaf (2009)[4] and Bekiros and Georgoutsos (2005)[5]; Danielsson and Morimoto (2000) apply EVT to Japanese financial data to confirm the accuracy and stability of this methodology over the GARCH-type techniques; Byström (2004) focuses on the negative distribution tails of the Swedish AFF and the U.S. DOW indices to compare EVT with generalized ARCH approaches and finds EVT to be a generally superior approach both for standard and more extreme VaR quantiles. Nevertheless, Fernández (2005) uses a sample comprised of several financial indices from the United States, Europe, Asia and Latin America and finds that conditional EVT gives the most accurate estimates when compared with traditional methods. Finally, Lee and Saltoglu (2001) concentrate on five Asian stock market indices and come to somewhat inconclusive results in the sense that conventional methods turn out to have more consistent performance but none of the methods used in that paper is shown to produce a superior VaR forecast.

In some papers, the focus is on the marginal or unconditional distribution of the process, without accounting for the conditional heteroscedasticity of most financial data (Christoffersen and Diebold, 2000; Danielsson and de Vries, 2000**;** Longin, 2000**;** Bekiros and Georgoutsos, 2005**;** Gilli and Këllezi, 2006**;** Assaf, 2009**)**. In this paper, however, we follow McNeil and Frey (2000) to overcome this shortcoming and proceed

---

[2] Value at Risk (VaR) is a generalized measure of market risk which tells you the maximum loss, with a given probability, over a certain time horizon. More formally, given some confidence level $\alpha \in (0,1)$, VaR at the confidence level α is given by the smallest number l such that the probability of the loss L exceeds l is no longer than (1-α). Thereby, in probabilistic terms, VaR is a quantile of the loss distribution.

[3] Longin (2000) compute the VaR of single and bivariate portfolio positions by applying the EVT methodology to S&P 500 index and the SBF 240 index.

[4] Assaf (2009) focuses on four emerging financial markets (Egypt, Jordan, Morocco and Turkey) to provide estimates of their tail index behaviour.

[5] In Bekiros and Georgoutsos (2008) the focus is on returns of the Dow Jones Industrial Average and the Cyprus Stock Exchange indices finding that at confidence levels higher (lower) than 99% the EVT-based methodology (conventional methods) produces the most accurate forecasts for extreme losses.

in two steps. First, we fit a GARCH-model to the return series with the aim of obtaining estimates of the conditional volatility. Second, we use the extreme value theory, in particular, the Peak Over Threshold (POT) approach, to estimate the distribution of the standardized normal residuals.

In contrast to Normal and Student's t distributions which are symmetric and therefore not able to capture differences between the upper and lower tails, the unconditional EVT estimator has the advantage of treating the tails separately. By applying the unconditional POT method to the residuals from the normal AR-GARCH model what we get are time-varying tail quantiles according to periods of high (low) volatility. Estimates of the tails of the residuals from models with a normal and a t-Student distributed conditional return distribution are additionally presented for comparative purposes.

This paper contributes to the literature by applying the methods proposed by McNeil and Frey (2000) to three financial indices representing the three main financial areas in the world, i.e., USA, UK and Japan, covering a sample period from 1964 (variable depending on the stock index) to 2009. Our sample extends those from previous studies focusing on (at least) one of these three financial indices and following the approach used in this paper[6]. To do so, we are concerned not only with in-sample estimation but also and most relevant to portfolio management, out-of-sample one-day prediction. Moreover, apart from considering the lower tail of the distributions, which is the most frequent choice, we additionally analyze the upper tail of the distribution. The reason is that the former represents losses for an investor with a long position in the financial index, whereas the latter represents losses for an investor being short on the index. Therefore, although throughout the paper we talk about tail quantile estimates, we distinguish between the lower and the upper tail, the lower tail quantiles estimates being direct VaR estimations, as usually defined in literature. Finally, to deal with the controversial issue of the threshold choice (necessary to define an observation as extreme), we use the standard method based on the mean residual life plot. This graphical tool is frequently employed to determine the threshold directly from visual inspection. In this paper we use the likelihood test ratio as a robustness check.

---

[6] For instance, Byström (2004) uses the time period January 2, 1980 to September 8, 1999 for DJI while Fernández (2005) uses the time period 1990 to 2002 for DJI and 1980 to 2002 for NIKKEI 225.

The remainder of this paper is organized as follows. Section 2 describes and carries out a preliminary analysis of the data set. In Section 3 the theoretical framework of the extreme value theory as well as the methods proposed by McNeil and Frey (2000) called conditional EVT are presented. Section 4 is concerned with the estimation of the GARCH-type models and the fitting of the GPD model to standardized normal returns for each of the financial indices involved in this study. In section 5, tail quantile estimates are obtained by applying the different methodologies considered in this study with comparative purposes. The empirical exercise is divided into an in-sample and an out-of-sample estimation. Finally, section 6 summarises the results and concludes.

2. Data

The data used are the historical daily log return series on three financial indices referring to three relevant financial areas such as USA, London and Japan. The selected financial indices are DJI, FTSE 100 and NIKKEI 225. Our sample respectively covers the following periods: January 2, 1964 to October 30, 2009, November 2, 1987 to October 30, 2009 and January 4, 1984 to October 30, 2009. The data has been taken from the Reuters database.

Table 1 reports some statistics on the log return series and the Ljung-Box test statistic for autocorrelation in returns and squared returns. As can be observed, all three series are stationary according to the Augmented Dickey Fuller statistics. Note the very high kurtosis and the negative value of skewness denoting wider lower tails.

According to the Ljung-Box test, the log return series display strong autocorrelation with the only exceptions being the FTSE 100 and NIKKEI 225 log returns which are not autocorrelated of order one. Though not shown, they present autocorrelation up until any other lag exceeding one. From a visual inspection of Figures 1-3 a noticeable degree of volatility clustering can be detected. To confirm such an intuition, the Ljung-Box test has been additionally applied to squared log returns. As can be observed in Table 1, the p-values for the Ljung-Box tests are below 0.05, indicating there is heteroscedasticity in the series.

Thus, two stylized facts for return series are detected: (i) the nonnormality of the unconditional distribution of returns suggested by the commented values of kurtosis and skewness and evidenced by highly significant Jarque-Bera statistics[7] and (ii) the time-varying volatility of returns indicated by the significant Ljung-Box test statistics showing strong autocorrelation in squared returns.

Table 1. Descriptive Statistics

Log returns on DJI, FTSE 100 and NIKKEI 225. ADF is the Augmented Dickey Fuller test statistic (without trend) and the 99% critical value is -3.43. Q(1) [$Q^2(1)$] and q(5) [$Q^2(5)$] are the Ljung-Box tests for autocorrelation at lags 1 and 5 in the log return series [in the squared log return series], their p-values are shown. A p-value less than or equal to 0.05 is interpreted as evidence against the null hypothesis that there is no autocorrelation up to lag shown in parenthesis. ** (*) denotes statistical significance at 1% (5%) level.

	DJI	**FTSE 100**	**NIKKEI 225**
Mean (%)	0.02	0.02	0.00
Median (%)	0.03	0.00	0.03
Standard Deviation (%)	1.03	1.11	1.46
Minimum	-0.25	-0.09	-0.16
Maximum	0.10	0.09	0.13
Skewness	-1.32	-0.139	-0.25
Kurtosis	42.62	9.82	11.40
Jarque-Bera (p-value)	759 878 (0.0000)	11 162 (0.0000)	18 770 (0.0000)
		t-Statistic	

[7] The Jarque-Bera statistic is χ_2^2 distributed under the null of normality.

ADF	-78.68**	-33.88**	-76.81**
		p-value	
Q(1)	0.00*	0.41	0.17
Q(5)	0.00*	0.00*	0.00*
$Q^2(1)$	0.00*	0.00*	0.00*
$Q^2(5)$	0.00*	0.00*	0.00*

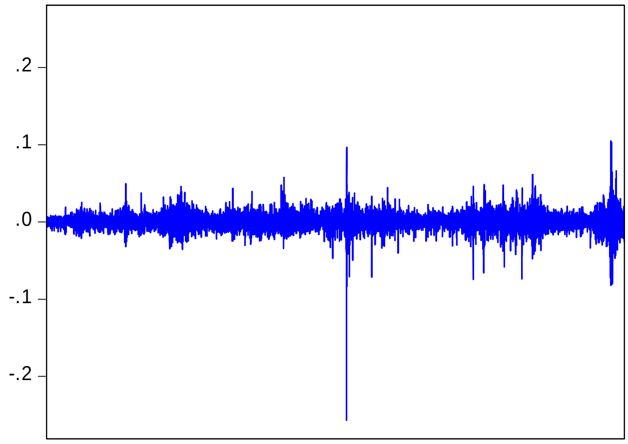


Figure 1. DJI index log returns (01/02/1964 – 30/10/2009)

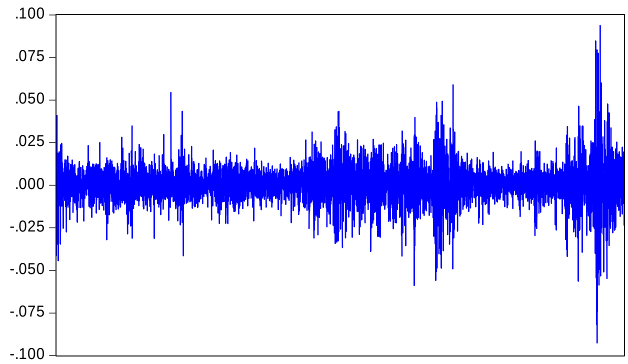


Figure 2. FTSE 100 index log returns (11/02/1987– 30/10/2009)

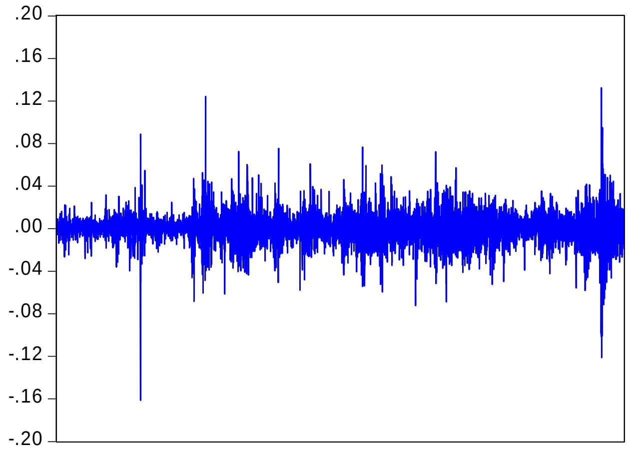


Figure 3. NIKKEI 225 index log returns (01/04/1984– 30/10/2009)

3. Methodology

The extreme value theory[8] relies on two main general definitions of extreme events. Following the so-called *Block Maxima* (BM) approach, data are taken to be the maxima (or minima) over certain blocks of time. In this context, it is appropriate to use the Generalized Extreme Value distribution. Instead, the *Peak Over Threshold* (POT) methodology considers as extreme those observations (X_i) that exceed a properly chosen high threshold *u*. These excesses, when independent, follow a Generalized Pareto Distribution. The BM approach compared to the POT approach presents a shortcoming: as just one extreme per block is chosen, completeness of the statistical population is not guaranteed. In fact, the former implies a loss of information that may be important, since the latter allows for more data to inform the analysis. Therefore, the threshold method uses data more efficiently and, for that reason, it is the method of choice in this paper.

Let X_1, X_2, … be a sequence of independent and identically distributed random variables, having marginal distribution function F. Under the POT approach, extremes are regarded as those of the X_i that exceed some high threshold u. If F were known, the distribution of threshold excesses would also be known. Since in practice this is not the

[8] See Leadbetter et al. (1983), Embrechts et al. (1999) and Coles (2003) for more details of extreme value theory.

case, approximations applicable for high values of the threshold are needed. According to Pickands (1975), for large enough *u,* the distribution function of y=X- *u*, conditional on X > *u*, belongs to the family of distributions called the generalized Pareto family and is approximately

$$H(y) = 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-\frac{1}{\xi}} \tag{1}$$

defined as $\left\{ y : y > 0 \ \ and \ \ (1 + \frac{\xi y}{\tilde{\sigma}}) > 0 \right\}$, where $\tilde{\sigma} = \sigma + \xi(u - \mu)$. "y" are the excesses of a threshold, σ is a scale parameter and ξ a shape parameter.

H(y) gives the probability of a random variable exceeding a high value given that it already exceeds a high threshold, say u. Thus, y = X-*u*, may be regarded as independent realizations of a random variable whose distribution can be approximated by a member of the Generalized Pareto family. Inference consists of fitting the generalized Pareto family to the observed threshold excesses. The result, which is stated for maxima, can be applied to minima by taking the sequence $–X_n$ instead of the sequence X_n (Coles, 2003).

The threshold choice is controversial and, according to McNeil and Frey (2000), the most important implementation issue in EVT. So far, no automatic algorithm with satisfactory performance for the selection of the threshold *u* is available. If we choose too low a threshold we might get biased estimates because the limit theorems do not apply any more, while high thresholds generate estimates with high variance due to the limited number of observations. Thus, the issue of threshold choice implies a balance between bias and variance.

In this paper, the issue of threshold choice has been handled through the standard method based on the mean residual life plot (Davison and Smith, 1990). The mean residual life plot is made up of the locus of points

$$(u, \frac{1}{n_u} \sum_{i=1}^{n_u} (x_{(i)} - u)) : u < x_{\max} \ , \tag{2}$$

where $x_{(1)}, \ldots, x_{(n_u)}$ consist of the n_u observations that exceed u, and x_{max} is the largest of the series to be fitted, X_i. Above a specific threshold u at which the generalized

Pareto distribution provides a valid approximation to the excess distribution, the mean residual life plot should be approximately linear in u. When applying this method, the choice of the threshold is frequently done directly from visual inspection (for instance, see Gilli and Këllezi, 2006, Coles, 2003); however, in this paper we use the likelihood test ratio as a robustness check.

The likelihood test ratio is defined as follows. Suppose that M_1 is a model with parameter vector θ, and M_0 is the subset of model M_1 obtained by constraining k of the components of θ to be zero. Let $l_0(M_0)$ and $l_1(M_1)$ be the maximized value of the log-likelihood for models M_0 and M_1 respectively. M_0 can be rejected in favour of M_1 at the α level of significance if $D=2[l_1(M_1) - l_0(M_0)] > c_\alpha$, where c_α is the (1-α) quantile of the x_k^2 distribution and k is the difference in the dimensionality of M_1 and M_0. D is known as the deviance statistic.

In our case, M_0 is identified with the linear model whereas M_1 corresponds to the quadratic model. Thus,

Linear model: $$mrl_t = \alpha + \beta \cdot u_t + \varepsilon_t \qquad (3)$$

Quadratic model: $$mrl_t = \varpi + \lambda \cdot u_t + \gamma u_t^2 + \phi_t \qquad (4)$$

where mrl_t denote the mean residual life computed as the sample mean of the n_u observations that exceed the corresponding threshold (u_t).

The deviance statistic is calculated at each specific threshold around the potential threshold identified by visual inspection, with the aim of determining the value of the specific u from which there is evidence of the linear model explaining better than the quadratic model the variation in the data.

As indicated in the introduction, there are previous studies in the literature that apply EVT-based methods directly to the series of returns, following the unconditional approach. However, the EVT requires the series to be identically and independently distributed (i.i.d.) and, given the conditional heteroscedasticity of most financial data, this approach is hardly appropriate. In fact, the presence of stochastic volatility implies that returns are not necessarily independent over time. Besides, financial time series generally show clusters of volatility. Therefore, we must look more carefully into the issue of de-clustering the extreme values so that they appear as approximately independent (McNeil, 1998).

Thereby, following the methods proposed by McNeil and Frey (2000), we use historical simulation for estimating the conditional mean and volatility of the log return series and threshold methods from EVT to estimate the distribution of the residuals (which are approximately independent).

Let $(X_t, t \in Z)$ be a strictly stationary time series representating daily observations of the log return on a financial index. Assuming that the dynamics of X are given by

$$X_t = \mu_t + \sigma_t Z_t \tag{5}$$

where the innovations Z_t are a strict white noise process with zero mean, unit variance and marginal distribution function $F_Z(z)$, our aim is to estimate the conditional quantiles in the tails of the predictive distributions. As well, we do obtain in-sample extreme quantiles estimates but also, and more relevantly, out-of-sample extreme quantiles estimates.

For 0 < q < 1, a conditional quantile is a quantile of the predictive distribution for the return over the next h days denoted by

$$x_q^t(h) = \inf(x \in \Re : F_{X+1+\ldots+X+h/\wp_T}(x) \geq q) \tag{6}$$

where $F_{X+1+\ldots+X+h/\wp_T}(x)$ denote the predictive distribution of the return over the next h days, given knowledge of returns up to and including day t. In particular, we are interested in quantiles for the 1-step predictive distribution which we denote by x_q^t. Being $F_{X+1/\wp_T}(x) = P(\sigma_{t+1}Z_{t+1} + \mu_{t+1} \leq x / \wp_t) = F_Z((X - \mu_{t+1})/\sigma_{t+1})$, the calculus of the conditional quantile simplifies to

$$x_q^t = \mu_{t+1} + \sigma_{t+1} z_q \tag{7}$$

where z_q is the upper qth quantile of the marginal distribution of Z_t which by assumption does not depend on t.

Firstly, we need a particular model for the dynamics of the conditional mean and volatility in order to obtain iid residual series which EVT will be applied to. In this paper, differentiated model specifications for the three studied index series are chosen so as to pre-whiten the returns**.**

We use maximum likelihood to estimate both the conditional mean and volatility from the corresponding GARCH-type model by assuming that the innovation distribution is standard normal. For comparative purposes we repeat the estimation procedure although this time considering that the distribution of the innovations is more heavier-tailed than is the normal, i.e. t-Student's.

To obtain conditional POT estimates, we follow McNeil and Frey (2000) and, firstly, fit a GARCH-type model to the return data by quasi-maximum likelihood, that is, maximize the log-likelihood function of the sample assuming normal innovations[9]. Secondly, we consider the resulting standardized residuals to be white noise process and estimate the tails of innovations using POT in order to finally compute the corresponding quantiles.

Then, conditional 95%, 97.5%, 99% and 99.5% tail quantiles (z_q^t) of the financial index log return series are estimated by multiplying the corresponding GARCH volatilities with quantiles from the standard normal, t-distribution and GPD (in this latter case by means of the application of the POT approach to standardized normal residuals) and adding the conditional mean return.

4. Estimation Results

In this section we present the models selected to capture the dependencies shown in the log return series as well as the corresponding estimation results. Then we apply the POT approach to the residuals from the AR-GARCH model that assumes normal innovations by fitting the GPD to the excesses over a predetermined threshold, which can be different according to the series.

4.1. AR-GARCH models

When looking for the best fitted AR-GARCH model to data, differences in the dynamics of the considered index log return series need individual analysis. Therefore,

[9] Even if innovations are not truly normally distributed, this way of proceeding still provides consistent and asymptotically normal estimates (see for instance Engle and González-Rivera, 1991).

according to the volatility clustering as well as the different seasonality pattern observed in the series, three differentiated model specifications have finally been chosen.

Thus, the DJI index log return series seems to be a realisation from an AR(1)-GARCH(1,1) process, while for the FTSE 100 index we are able to remove autocorrelation both in the returns themselves as well as in squared returns by simply fitting a GARCH(1,1) model. Lastly, the NIKKEI 225 index log return series requires an AR(1), AR(10)-GARCH(1,1) model[10]. Maximum likelihood estimates for each of the involved index series are reported in Table 2.

Table 2. AR-GARCH Model

Panel A, B and C respectively display AR(1)-GARCH(1,1) parameters estimates for the DJI, GARCH(1,1,) parameters estimates for the FTSE 100 and AR(1) AR(10)-GARCH(1,1) parameter estimates for the NIKKEI 225 indices. d.f. is degrees of freedom. * (**) denotes statistical significance at a 1% (10%) level.

Panel A: DJI index		
AR(1)-GARCH(1,1) $X_t = \phi_0 + \phi_1 X_{t-1} + \sigma_t \varepsilon_t$ $\sigma^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$	Normal innovations	Student´s t innovations
ϕ_0	0.4E04*	0.4E04*
ϕ_1	0.081*	0.074*
α_0	6.74E-07*	5.02E-07*
α_1	0.067*	0.054*
β_1	0.928*	0.941*

[10] In McNeil and Frey (2000) an AR(1) model for the mean and a GARCH(1,1) process for the volatility are used. AR(1), AR(24) and AR(168) terms combined with a GARCH(1,1) model are included in Byström(2005).

d.f.		8.04*
Panel B: FTSE 100 index		
GARCH(1,1) $X_t = \phi_0 + \sigma_t \varepsilon_t$ $\sigma^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$	Normal innovations	Student´s t innovations
ϕ_0	0.4E04*	0.5E04*
α_0	1.18E-06*	1.12E-06*
α_1	0.079*	0.076*
β_1	0.909*	0.913*
d.f.		10.77*
Panel C: NIKKEI 225 index		
AR(1),AR(10)-GARCH(1,1) $X_t = \phi_0 + \phi_1 X_{t-1} + \phi_2 X_{t-10} + \sigma_t \varepsilon_t$ $\sigma^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$	Normal innovations	Student´s t innovations
ϕ_0	0.7E04*	0.7E04*
ϕ_1	0.027**	0.004
ϕ_2	0.036*	0.034*
α_0	2.36E-06*	1.26E-06*
α_1	0.132*	0.095*
β_1	0.867*	0.904*
d.f.		7.07*

As we have said, we look at both the standardized normal and Student's t residuals distribution of each of the involved series. Table 3 displays the descriptive statistics of these series distinguishing between in-sample and out-of-sample estimation. Note that, in contrast to the log return series, the standardized residuals are approximately independent according to the Ljung-Box tests on the residuals and the squared residuals (in particular, Ljung-Box tests for one and five lags are presented).

Table 3. Descriptive Statistics of standardized normal residuals

IS and OS are in-sample and out-of-sample estimation. Q(1) [$Q^2(1)$] and q(5) [$Q^2(5)$] are the Ljung-Box tests for autocorrelation at lags 1 and 5 in the log return series [in the squared log return series], their p-values are shown. A p-value less than or equal to 0.05 is interpreted as evidence against the null hypothesis that there is no autocorrelation up to lag shown in parenthesis. * denotes statistical significance at a 5% level, indicating significant serial correlation in the residuals.

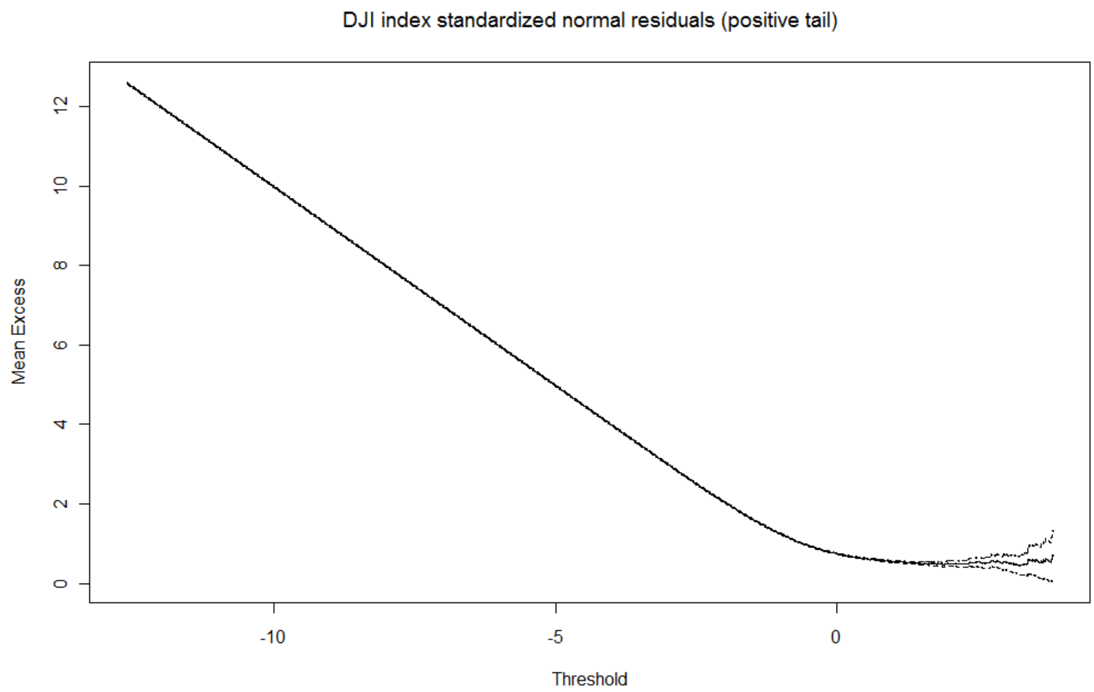
	DJI		**FTSE 100**		**NIKKEI 225**	
	IS	OS	IS	OS	IS	OS
Mean (%)	-0.02	-0.02	-0.02	-0.02	-0.05	-0.05
Median (%)	-0.01	-0.01	-0.03	-0.02	-0.03	-0.04
Standard Deviation (%)	1.00	1.01	1.00	1.00	1.00	1.00
Minimum	-12.61	-9.54	-5.57	-4.70	-13.32	-6.22
Maximum	5.45	5.47	6.09	5.94	10.92	6.39
Skewness	-0.43	-0.23	-0.22	-0.21	-0.55	-0.10
Kurtosis	7.86	5.27	4.00	3.90	11.02	4.52
Q(1)	0.85	0.26	0.20	0.36	0.55	0.78
Q(5)	0.44	0.84	0.06	0.06	0.55	0.98
$Q^2(1)$	0.92	0.26	0.38	0.96	0.16	0.03*

$Q^2(5)$	0.47	0.64	0.50	0.14	0.69	0.25

4.2. POT methodology applied to the upper tail (maxima)

Maximized value of the log likelihood for the quadratic and linear models together with deviance statistics calculated at different potential thresholds for each of the standardized residuals series involved in the study are shown in Figures 4-6 and in Figures 7-9 respectively for in-sample estimation and out-of-sample estimation.

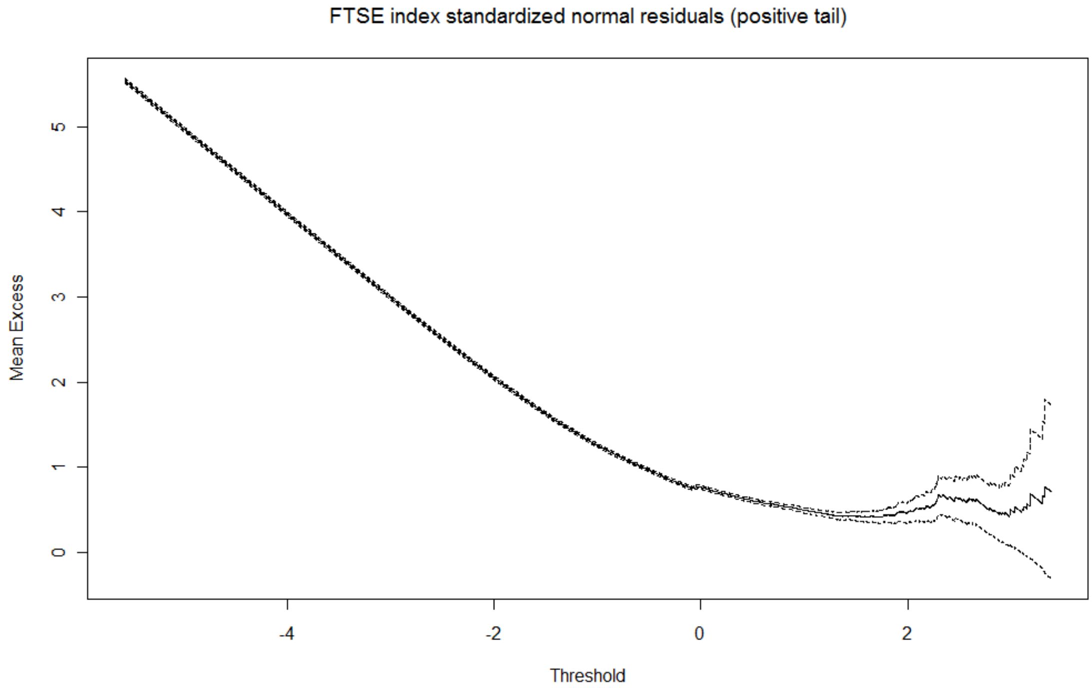
The particular thresholds from which the linear model fits better than the quadratic one are highlighted in bold in all the cases. Thereby, according to the test likelihood ratio, the thresholds for the in-sample estimation should be the following: $u^*_{DJI} = 1.65$, $u^*_{FTSE} = 1.77$ and $u^*_{NIKKEI} = 2.18$, considering as extreme values 4.4%, 2.9% and 1.2% of data, respectively (Figures 4-6).



Log likelihood	u=1.62	u=1.63	u=1.64	**u=1.65**	u=1.66	u=1.67
Linear relationship	332.64	330.22	327.75	325.35	322.43	319.57

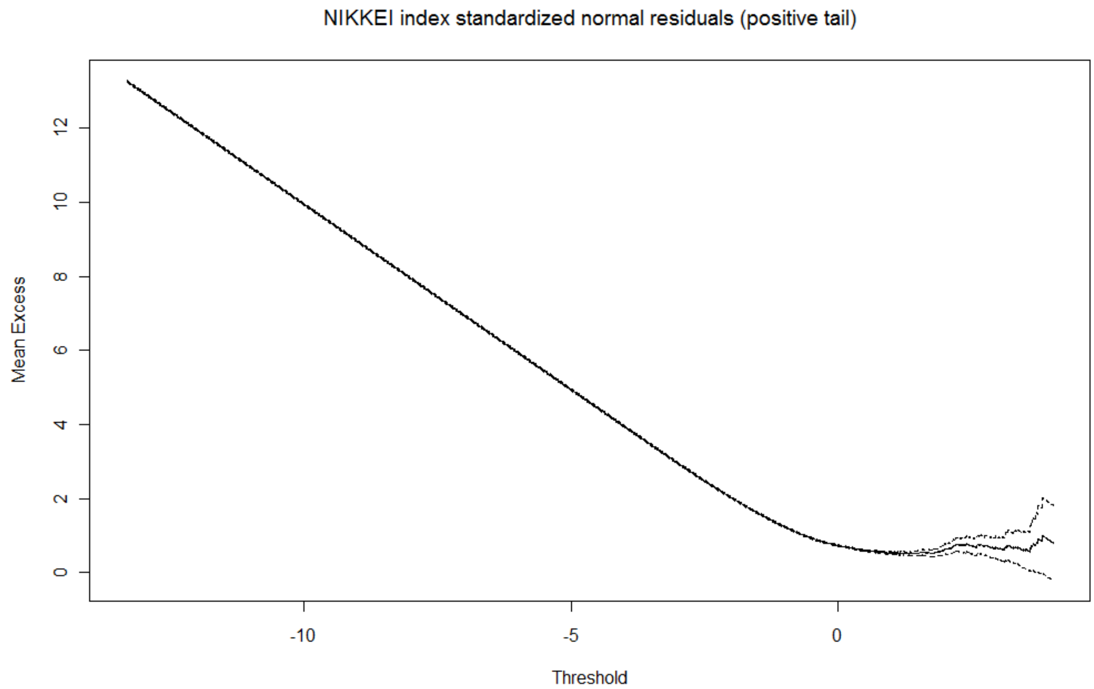
Quadratic relationship	335.91	332.88	329.86	326.95	323.71	320.54
Deviance Statistic	6.55	5.33	4.23	3.20*	2.55*	1.93*

Figure 4. In-sample mean residual life plot and likelihood ratio tests. Calculated at different potential thresholds for the DJI index standardized normal residuals (upper tail). * denotes statistical significance at 5% level.



Log likelihood	u=1.74	u=1.75	u=1.76	**u=1.77**	u=1.78	u=1.79
Linear relationship	169.81	167.26	164.66	162.01	159.83	157.58
Quadratic relationship	172.15	169.39	166.64	163.90	161.32	158.74
Deviance Statistic	4.67	4.26	3.96	3.77*	2.97*	2.31*

Figure 5. In-sample mean residual life plot and likelihood ratio tests. Calculated at different potential thresholds for the FTSE 100 index standardized normal residuals (upper tail). * denotes statistical significance at 5% level.



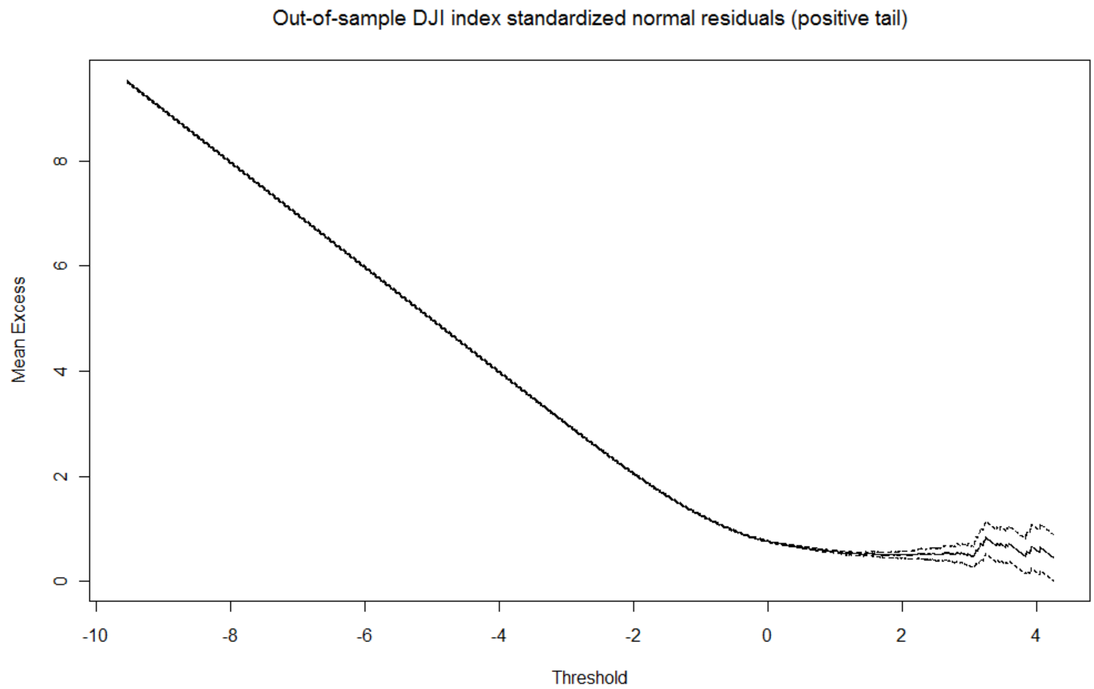
Log likelihood	u=2.15	u=2.16	u=2.17	**u=2.18**	u=2.19	u=2.20
Linear relationship	189.94	192.00	190.47	188.03	186.21	183.56
Quadratic relationship	194.43	194.86	192.56	189.72	187.34	184.45
Deviance Statistic	8.98	5.73	4.18	3.37*	2.26*	1.78*

Figure 6. In-sample mean residual life plot and likelihood ratio tests. Calculated at different potential thresholds for the NIKKEI 225 index standardized normal residuals (upper tail). * denotes statistical significance at 5% level.

Much more interesting than in-sample estimation is out-of-sample estimation, as the latter allows us to forecast tail estimates. Thus, we fix a constant memory n (n=1001 in our case) so that at the end of day t our data consist of the last 1001 log returns. On each day we fit a new AR-GARCH-type model, i.e. the AR-GARCH models selected in Section 4.1, to capture the dynamics of the three studied indices. The next step is to obtain the quantile estimates from the GPD by fitting this distribution to the excesses of the new standardized normal residuals over the corresponding thresholds.

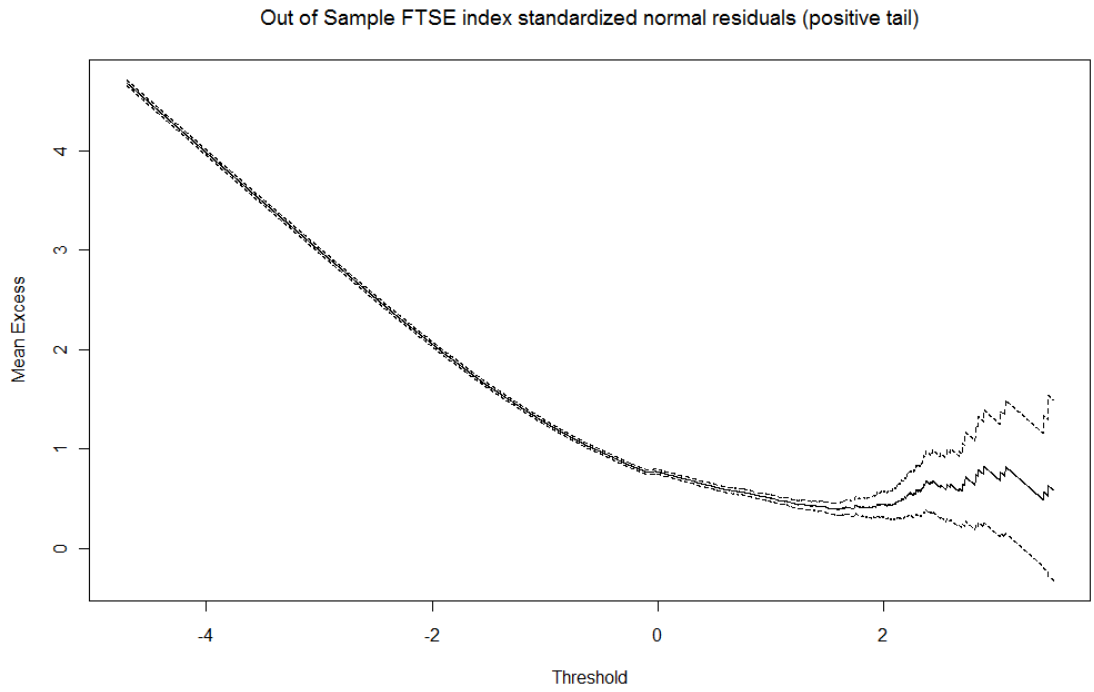
These thresholds are fixed by applying the standard method based on the mean residual life plot. Finally, we calculate the EVT conditional quantile estimates by multiplying the new estimated GARCH volatilities with quantiles from the standard normal, t-distribution and GPD (in this latter case by means of the application of the POT approach) and adding the new estimated conditional mean returns.

The thresholds suggested by the test likelihood ratio for the out-of-sample estimation are $u^*_{DJI} = 1.66$, $u^*_{FTSE} = 2.06$ and $u^*_{NIKKEI} = 2.03$, considering as extreme values 4.4%, 1.4% and 1.7% of data, respectively (Figures 7-9).



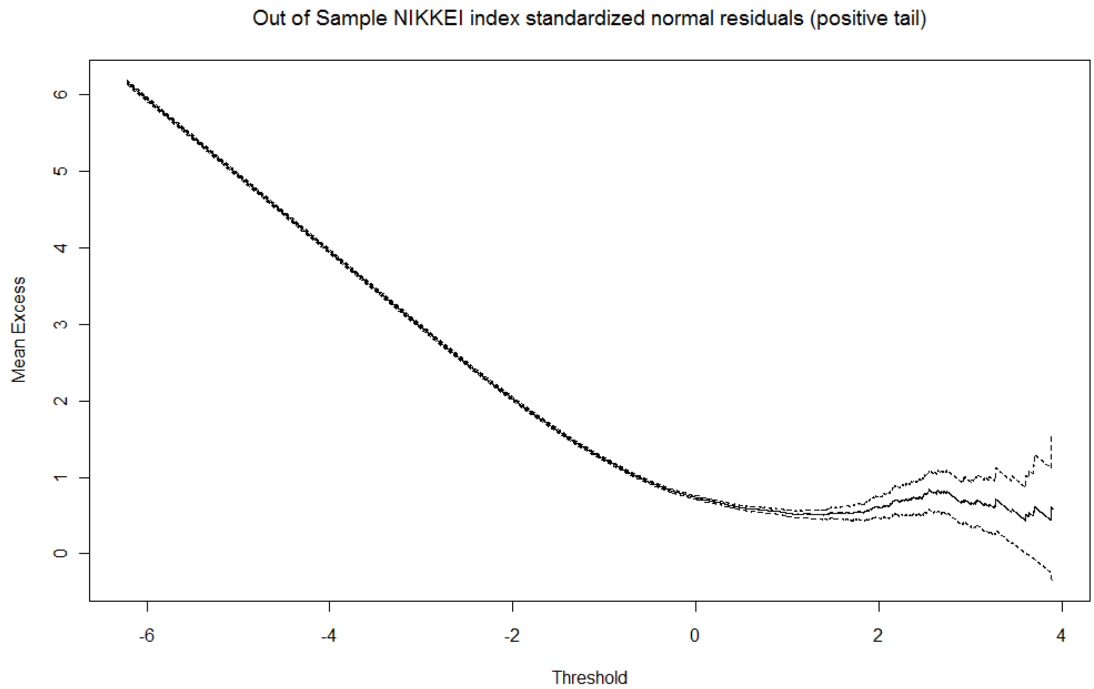
Log likelihood	u=1.63	u=1.64	u=1.65	**u=1.66**	u=1.67	u=1.68
Linear relationship	376.00	374.77	373.32	373.59	371.13	367.75
Quadratic relationship	380.72	378.40	376.01	373.59	371.13	367.75
Deviance Statistic	9.45	7.26	5.38	3.79*	2.49*	1.80*

Figure 7. Out-of-sample mean residual life plot and likelihood ratio tests. Calculated at different potential thresholds for the DJI index standardized normal residuals (upper tail). * denotes statistical significance at 5% level.



Log likelihood	u=2.03	u=2.04	u=2.05	**u=2.06**	u=2.07	u=2.08
Linear relationship	106.01	103.41	101.68	99.20	96.17	93.04
Quadratic relationship	110.34	107.03	104.16	100.99	97.69	94.43
Deviance Statistic	8.66	7.24	4.97	3.58*	3.05*	2.78*

Figure 8. Out-of-sample mean residual life plot and likelihood ratio tests. Calculated at different potential thresholds for the FTSE 100 index standardized normal residuals (upper tail). * denotes statistical significance at 5% level.



Log likelihood	u=2.00	u=2.01	u=2.02	**u=2.03**	u=2.04	u=2.05
Linear relationship	181.95	178.93	176.10	173.30	170.84	168.06
Quadratic relationship	184.31	181.26	178.18	175.12	172.24	169.22
Deviance Statistic	4.72	4.65	4.16	3.64*	2.80*	2.31*

Figure 9. Out-of-sample mean residual life plot and likelihood ratio tests. Calculated at different potential thresholds for the NIKKEI 225 index standardized normal residuals (upper tail). * denotes statistical significance at 5% level.

From a visual inspection of mean life residual plots, the selected thresholds are around the lowest values of u for which the mean residual life plots seem to be linearly related to the corresponding potential thresholds, so that we conclude that the selected thresholds do not seem unreasonable.

The excesses over the selected thresholds are fitted to the GPD in each case. Parameters under the in-sample and out-of-sample estimation have been estimated by maximum likelihood and are shown respectively in Tables 4 and 5. On the one hand, we obtain the estimates $\hat{\xi} = 0.016$ (0.171, 0.029) and $\hat{\sigma} = 0.480$ (0.360, 0.720) for the DJI (FTSE 100, NIKKEI 225) index within the in-sample estimation. On the other hand, under the out-of-sample estimation, we obtain the estimates $\hat{\xi} = 0.001$ (0.337, 0.217) and $\hat{\sigma} = 0.511$ (0.295, 0.474) for the DJI (FTSE 100, NIKKEI 225) index.

Table 4. Threshold In-sample

Panel A, B, C respectively shows in-sample maximum likelihood GPD parameter estimates (with standard errors in parenthesis) and threshold values for both tails of the standardized normal residuals distribution of the DJI, FTSE 100 and NIKKEI 225 indices.

	GPD parameters estimates	
	Upper tail	Lower tail
Panel A: DJI		
σ	0.480 (0.02)	0.451 (0.04)
ξ	0.016 (0.04)	0.340 (0.08)
u	1.65	2.18
Panel B: FTSE 100		
σ	0.360 (0.04)	0.646 (0.02)
ξ	0.171 (0.09)	-0.04 (0.02)
u	1.77	0.93

Panel C: NIKKEI 225		
σ	0.720 (0.11)	0.421 (0.12)
ξ	0.029 (0.11)	0.633 (0.29)
u	2.18	2.79

Table 5. Threshold Out-of-sample

Out-of-sample maximum likelihood GPD parameter estimates (with standard errors in parenthesis) for both tails of the standardized normal residuals distribution. Panel A, B, C respectively shows the estimates for the DJI, FTSE 100 and NIKKEI 225 indices.

	GPD parameters estimates	
	Upper tail	Lower tail
Panel A: DJI		
σ	0.511 (0.03)	0.514 (0.04)
ξ	0.001 (0.04)	0.180 (0.06)
u	1.66	1.95
Panel B: FTSE 100		
σ	0.295 (0.06)	0.711 (0.12)
ξ	0.337 (0.17)	-0.18 (0.12)
u	2.06	2.52
Panel C: NIKKEI 225		
σ	0.474 (0.08)	0.573 (0.15)
ξ	0.217 (0.15)	0.158 (0.21)

u	2.03	2.82

4.3. POT methodology applied to the lower tail (minima)

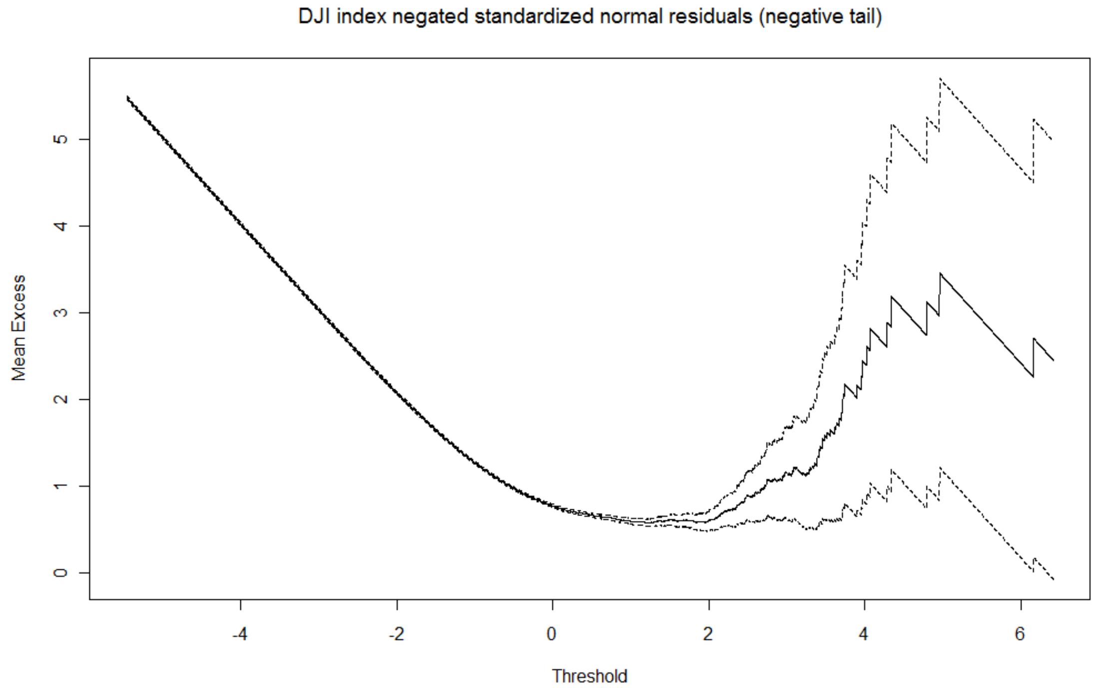
One of the advantages of the GPD approach to tail estimation is the fact that it allows for the handling of upper and lower tails separately. In contrast, normal and Student's t symmetric distributions are unable to capture any difference between them since both tails are assumed to present identical characteristics.

Of note is also the fact that the threshold level finally chosen depends on the particular series. Thus, the number of data exceeding the corresponding threshold is different according to this threshold level, which is another sign of this methodology's flexibility.

Similarly to the upper tail, the mean residual life plots together with some deviance statistics calculated at several thresholds are shown in Figures 10-12 and in Figures 13-15, respectively, under the in-sample and out-of-sample estimation.

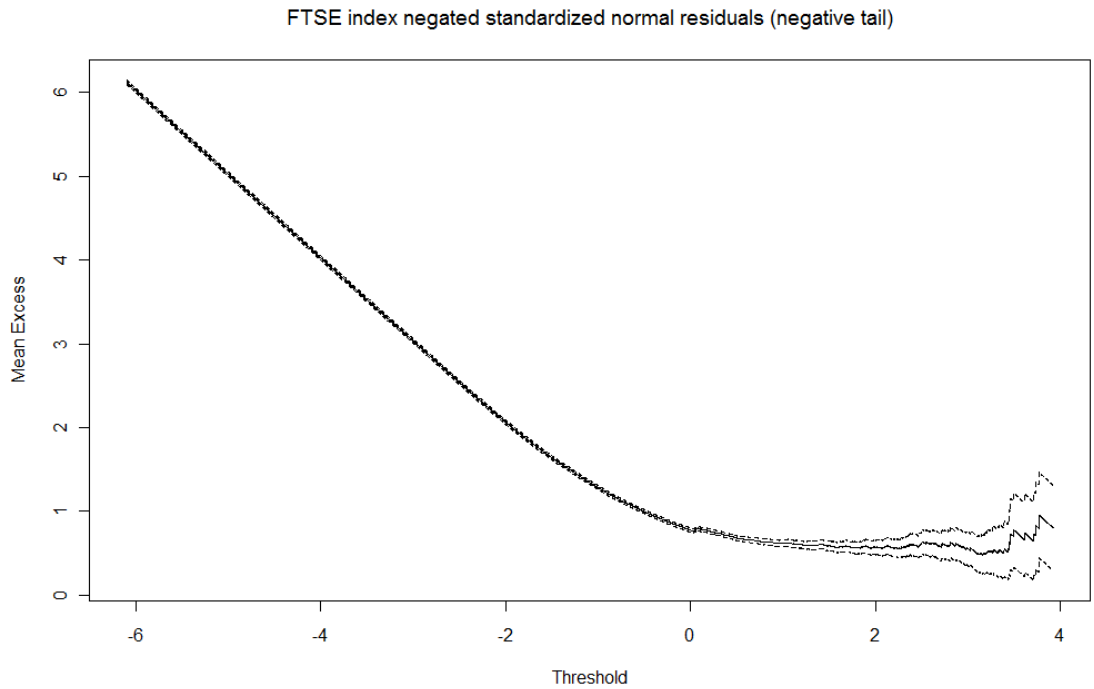
Thus, within the in-sample estimation, the thresholds suggested by the ratio likelihood test are u^*_{DJI} = -2.18, u^*_{FTSE} = -0.93 and u^*_{NIKKEI} = -2.79, leaving 1.8%, 16.8% and 0.6% of data below each of them, respectively. As can be observed, the selected thresholds for the lower tail of the DJI and the NIKKEI 225 (FTSE 100) indices are greater (lower), in absolute value, than the ones for the upper tail, which is a sign of asymmetry in the series. The estimates $\hat{\xi} = 0.340$ (-0.040, 0.633) and $\hat{\sigma} = 0.451$ (0.646, 0.421) for the DJI (FTSE 100, NIKKEI 225) index are displayed in Table 4. Since the shape parameter gives an indication of the heaviness of the tail (the larger ξ, the heavier the tail), results lead us to conclude that the upper tail of the standardized normal residuals distribution is heavier than the lower tail for the FTSE 100 index, whereas the reverse holds for the DJI and the NIKKEI 225 indices. This result is only partly consistent with that from Gilli and Këllezi (2006) who state that the left tail is heavier than the right one for the NIKKEI 225 and the FTSE 100 indices, though it is true that in that paper the unconditional approach is used and the studied samples also differ[11].

[11] Sample periods in Gilli and Këllezi (2006) cover from 01/07/1970 through 08/17/2004 and from 01/05/1984 through 08/17/2004 respectively for DJI and FTSE 100.



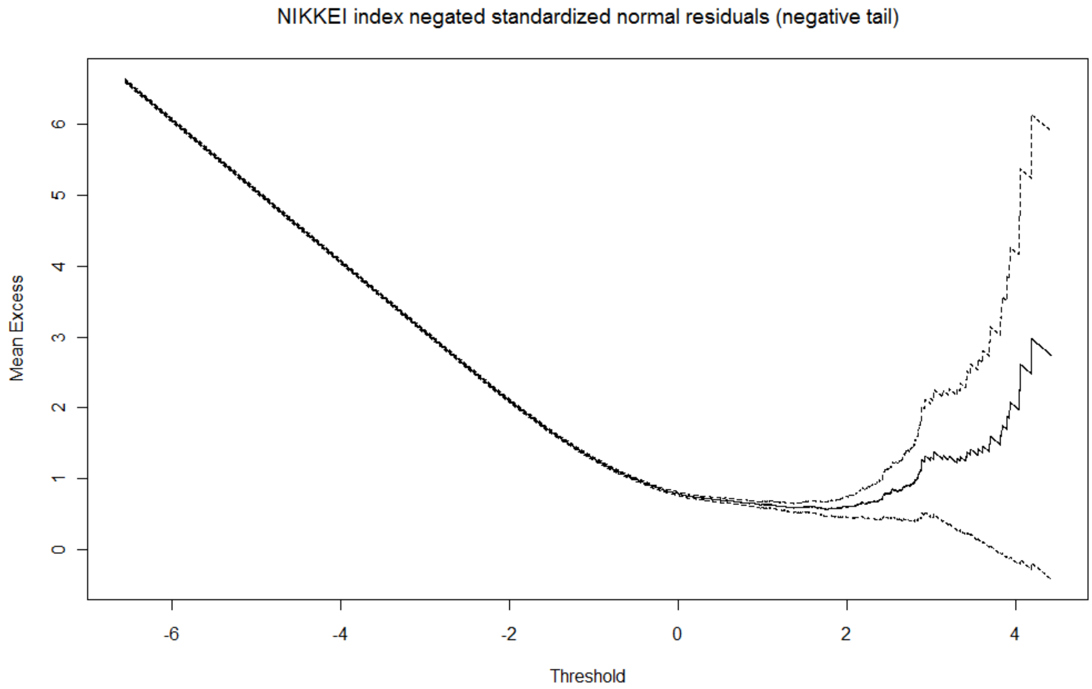
Log likelihood	u=2.15	u=2.16	u=2.17	**u=2.18**	u=2.19	u=2.20
Linear relationship	118.44	115.88	113.09	110.12	107.59	105.10
Quadratic relationship	121.51	118.36	115.19	112.02	108.98	106.02
Deviance Statistic	6.14	4.97	4.20	3.81*	2.78*	1.84*

Figure 10. In-sample mean residual life plot and likelihood ratio tests. Calculated at different potential thresholds for the DJI index negated standardized normal residuals (lower tail). * denotes statistical significance at 5% level.



Log likelihood	u=0.90	u=0.91	u=0.92	**u=0.93**	u=0.94	u=0.95
Linear relationship	520.52	518.12	515.70	513.25	510.74	508.17
Quadratic relationship	522.99	520.34	517.68	515.01	512.31	509.58
Deviance Statistic	4.94	4.42	3.95	3.51*	3.14*	2.82*

Figure 11. In-sample mean residual life plot and likelihood ratio tests. Calculated at different potential thresholds for the FTSE 100 index negated standardized normal residuals (lower tail). * denotes statistical significance at 5% level.

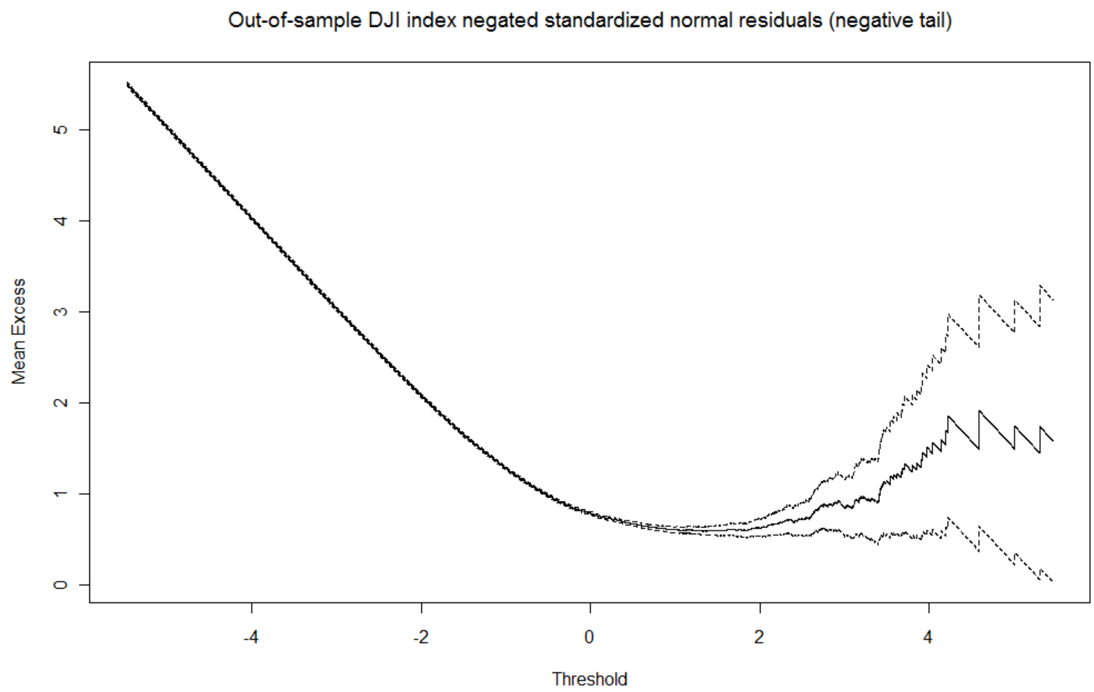


Log likelihood	u=2.76	u=2.77	u=2.78	**u=2.79**	u=2.80	u=2.81
Linear relationship	29.63	29.38	27.60	27.49	24.38	21.07
Quadratic relationship	36.38	33.35	29.87	27.84	24.41	21.12
Deviance Statistic	13.50	7.93	4.55	0.70*	0.07*	0.10*

Figure 12. In-sample mean residual life plot and likelihood ratio tests. Calculated at different potential thresholds for the NIKKEI 225 index negated standardized normal residuals (lower tail). * denotes statistical significance at 5% level.

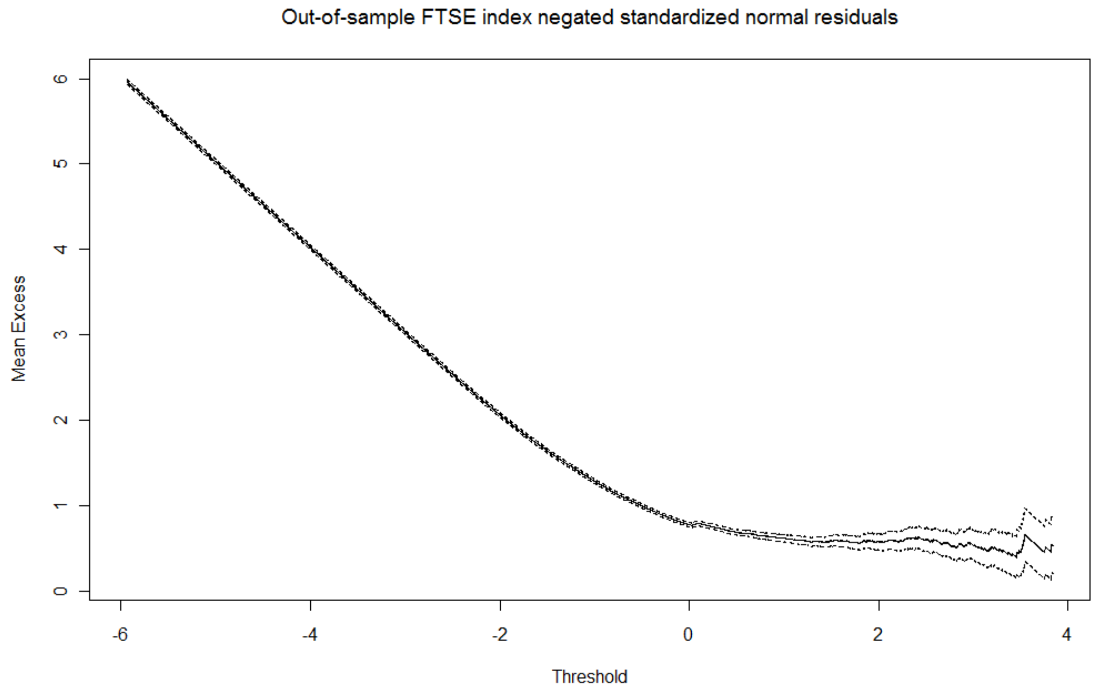
Under the out-of-sample estimation, however, the thresholds should be u^*_{DJI} = -1.95, u^*_{FTSE} = -2.52 and u^*_{NIKKEI} = -2.82, leaving 3.0%, 1.2% and 0.6% of data above each of them, respectively. In this case, the asymmetry of the distribution is also evidenced by comparing the upper and lower tails in terms of the estimated shape parameters and thresholds. The estimates $\hat{\xi} = 0.180$ (-0.180, 0.158) and $\hat{\sigma} = 0.514$ (0.711, 0.573) for the DJI (FTSE 100, NIKKEI 225) index are shown in Table 5.

On the one hand, the estimated thresholds for the lower tail within the out-of-sample estimation are always higher than the ones for the upper tail. On the other hand, regarding the estimated shape parameters, the results obtained from the in-sample estimation remain constant under the out-of-sample estimation except for the NIKKEI 225 index that exhibits a shape parameter for the upper tail (0.217) higher than the one for the lower tail (0.158), meaning that, in this case, the upper tail is heavier than the lower tail (Table 5).



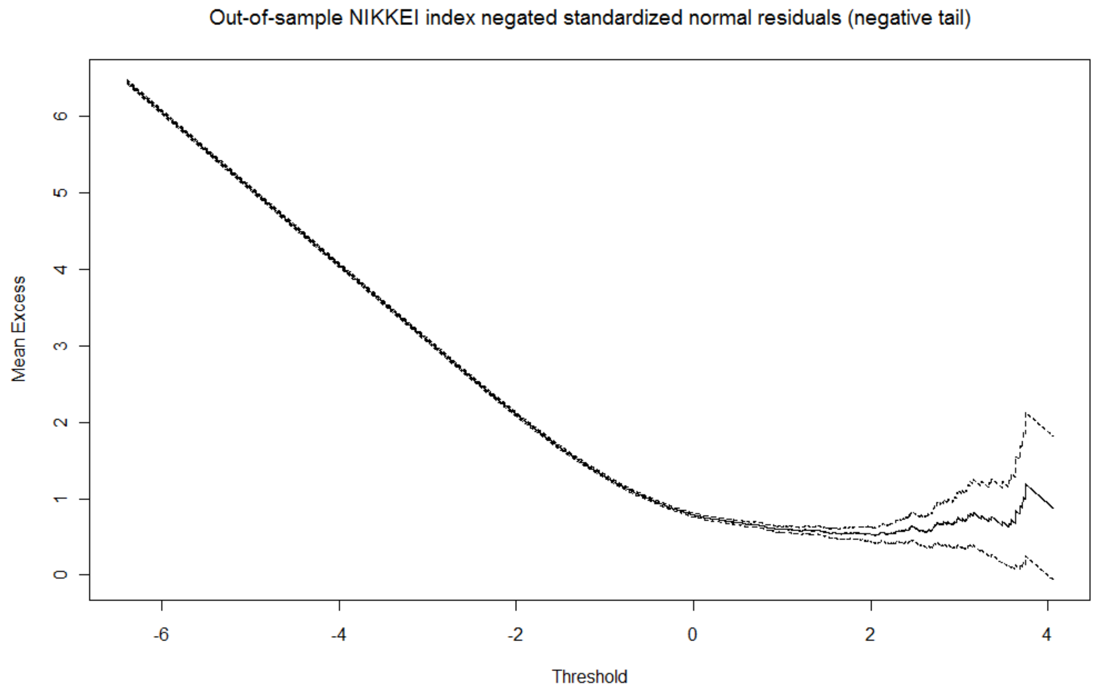
Log likelihood	u=1.92	u=1.93	u=1.94	**u=1.95**	u=1.96	u=1.97
Linear relationship	248.74	246.16	243.55	240.97	238.48	235.90
Quadratic relationship	248.74	246.16	243.55	240.97	238.48	235.90
Deviance Statistic	5.09	4.46	3.94	3.40*	2.80*	2.39*

Figure 13. Out-of-sample mean residual life plot and likelihood ratio tests. Calculated at different potential thresholds for the DJI index negated standardized normal residuals (lower tail). * denotes statistical significance at 5% level.



Log likelihood	u=2.49	u=2.50	u=2.51	**u=2.52**	u=2.53	u=2.54
Linear relationship	166.26	163.18	160.15	157.75	154.91	151.84
Quadratic relationship	168.76	165.63	162.48	159.53	156.40	153.28
Deviance Statistic	5.00	4.89	4.64	3.56*	2.98*	2.87*

Figure 14. Out-of-sample mean residual life plot and likelihood ratio tests. Calculated at different potential thresholds for the FTSE 100 index negated standardized normal residuals (lower tail). * denotes statistical significance at 5% level.



Log likelihood	u=2.79	u=2.80	u=2.81	**u=2.82**	u=2.83	u=2.84
Linear relationship	48.54	45.18	43.04	40.04	39.74	36.29
Quadratic relationship	52.05	48.79	45.17	41.46	39.82	36.93
Deviance Statistic	7.03	7.23	4.27	2.85*	0.15*	1.28*

Figure 15. Out-of-sample mean residual life plot and likelihood ratio tests. Calculated at different potential thresholds for the NIKKEI 225 index negated standardized normal residuals (lower tail). * denotes statistical significance at 5% level.

Similarly to the in-sample estimation, from a visual inspection of mean life residual plots, the selected thresholds are around the lowest values of u for which the mean residual life plots seem to be linearly related to the corresponding potential thresholds, so that we consider the selected thresholds acceptable. The excesses over the selected thresholds are fitted to the GPD in each case and maximum likelihood parameters are shown in Tables 4 and 5.

5. Tail quantile calculations and backtesting

The estimates from the previous section allow us to compute the series of conditional tail quantiles by multiplying the estimated conditional volatility with the quantiles of the normal distribution, the t-distribution or the generalized Pareto distribution and finally adding the estimated conditional mean.

The accuracy of the estimates under the distributions considered in the present study can be assessed by counting the number of actual returns that are larger than the estimated tail quantile and comparing this figure with the theoretically expected number of excesses[12] for a determined probability. Of course, the closer the empirically observed number of excesses is to the theoretically expected amount, the more preferable the method is for estimating the tail quantiles.

As a first step, we carry out an in-sample evaluation mainly to investigate the fit of the models to extreme data, followed by an out-of-sample evaluation to test how well future extreme movements can be predicted, the latter being of greater concern to risk managers.

5.1. In Sample Evaluation

Table 6 presents the number of excesses for both tails at different quantiles associated with each of the involved distributions, together with the theoretically expected number of excesses for the DJI (Panel A), FTSE 100 (Panel B) and NIKKEI 225 (Panel C) indices. To help the reader with the comparison, closer numbers of estimated excesses to theoretically expected ones are highlighted in bold. Also reported (in parenthesis) is the difference between the theoretically expected and the estimated excesses.

As can be observed in Table 6, the results do indicate that the EVT-based approach is the most successful for capturing the behaviour of the upper tail of the DJI and FTSE

[12] For example, the expected number of excesses of a 95% tail quantile over a sample of 11564 observations is 578 (0.05*11564).

100 indices at all the considered levels of probability, as well as at the most extreme levels (99% and 99.5%) in the NIKKEI 225 index.

Nevertheless, according to our results, the normal AR-GARCH and the AR-GARCH-t models generally provide more accurate estimations for the lower tails. In fact, the former performs better at the 95% and 97.5% levels whereas the latter do so at the 99% and 99.5% of the DJI and the NIKKEI 225 indices. The reason is that the Student´s t distribution is a fat-tailed distribution compared to the normal. The AR-GARCH-t model is also the one that produces better tail estimates for the FTSE 100 index, with the only exception being the most extreme tail quantile, i.e. at the 99.5% level, in which it is overcome by the EVT-based method. These results are consistent with previous studies in the sense that EVT is particularly accurate as tails become more extreme (Danielsson and de Vries, 1997; Longin (2000); Assaf (2009); Bekiros and Georgoutsos, 2005).

Table 6. In sample evaluation

In sample evaluation of estimated (positive and negative) tail quantiles at different probabilities for the DJI, FTSE 100 and NIKKEI 225 indices. Closer numbers of estimated excesses to theoretically expected ones are highlighted in bold. Also reported are the differences between the theoretically expected and the estimated excesses (in parenthesis).

Panel A: DJI							
		AR-GARCH		**AR-GARCH-t**		**Conditional GPD**	
Probability	Expected	Upper tail	Lower tail	Upper tail	Lower tail	Upper tail	Lower tail
0.95	578	506	**565**	547	613	**594**	416
		(72)	(13)	(31)	(-35)	(-16)	(162)
0.975	289	242	**312**	253	320	**300**	253

		(47)	(-23)	(36)	(-31)	(-11)	(36)
0.99	116	133	163	84	**115**	**115**	104
		(-17)	(-47)	(32)	(1)	(1)	(12)
0.995	58	75	101	41	**55**	**63**	51
		(-17)	(-43)	(17)	(3)	(-5)	(7)
Panel B: FTSE 100							
		AR-GARCH		**AR-GARCH-t**		**Conditional GPD**	
Probability	Expected	Upper tail	Lower tail	Upper tail	Lower tail	Upper tail	Lower tail
0.95	287	222	313	240	**278**	**273**	255
		(65)	(26)	(47)	(9)	(14)	(32)
0.975	144	91	172	92	**147**	**152**	126
		(53)	(-28)	(52)	(-3)	(-8)	(18)
0.99	57	36	101	29	**59**	**57**	48
		(21)	(-44)	(28)	(-2)	(0)	(9)
0.995	29	27	60	21	34	**29**	**28**
		(2)	(-31)	(8)	(-5)	(0)	(1)
Panel C: NIKKEI 225							
		AR-GARCH		**AR-GARCH-t**		**Conditional GPD**	
Probability	Expected	Upper tail	Lower tail	Upper tail	Lower tail	Upper tail	Lower tail
0.95	318	237	**360**	**256**	389	613	104
		(81)	(-42)	(62)	(-71)	(-259)	(214)
0.975	159	**108**	**195**	107	201	239	85
		(51)	(-36)	(52)	(-42)	(-80)	(74)

0.99	64	61	108	41	**71**	**62**	53
		(3)	(-44)	(23)	(-7)	(2)	(11)
0.995	32	45	64	24	**33**	**33**	29
		(-13)	(-32)	(8)	(-1)	(1)	(3)

5.2. Out-of-sample evaluation

In this case (Table 7), results indicate that EVT-based conditional quantile estimates are much more accurate than the conventional AR-GARCH models assuming normal or t-Student's innovations[13]. In fact, it occurs in 20 to 24 cases. Furthermore, at most of the considered confidence levels, the AR-GARCH model combined with normal or t-Student's innovations underestimate the upper tail and overestimate the lower tail. This is the consequence of using a symmetric distribution with data which are asymmetric in the tails. The almost perfect correspondence between the theoretically expected number of violations and the estimated number of violations provided by GPD evidences the suitability of this methodology to estimate the tails of the DJI, FTSE 100 and NIKKEI 225 indices returns distributions.

Table 7. Out-of-sample evaluation

Out-of-sample evaluation of estimated (positive and negative) tail quantiles at different confidence levels for the DJI, FTSE 100 and NIKKEI 225 indices. Closer numbers of estimated excesses to theoretically expected ones are highlighted in bold. Also reported are the differences between the theoretically expected and the estimated excesses (in parenthesis).

Panel A: DJI

[13] This result is consistent with that of Byström (2004), in which it is stated that for the lower tail of the DJI index (December 14, 1983 to September 8, 1999) and within the out-of-sample estimation the EVT-based models do a better job for confidence levels equal or higher than 99%.

		AR-GARCH		AR-GARCH-t		Conditional GPD	
Probability	Expected	Upper tail	Lower tail	Upper tail	Lower tail	Upper tail	Lower tail
0.95	528	477 (51)	**534** (-6)	521 (7)	544 (-16)	**523** (5)	496 (32)
0.975	264	241 (23)	294 (-30)	243 (21)	292 (-28)	**266** (-2)	**267** (-3)
0.99	106	129 (-23)	166 (-60)	88 (18)	115 (-9)	**103** (3)	**110** (-4)
0.995	53	82 (-29)	112 (-59)	36 (17)	64 (-11)	**52** (1)	**53** (0)
Panel B: FTSE 100							
		AR-GARCH		**AR-GARCH-t**		**Conditional GPD**	
Probability	Expected	Upper tail	Lower tail	Upper tail	Lower tail	Upper tail	Lower tail
0.95	237	197 (40)	**255** (-18)	**207** (30)	272 (-35)	142 (95)	421 (-184)
0.975	119	78 (41)	**141** (-22)	82 (37)	146 (-27)	**100** (19)	151 (-32)
0.99	48	32 (16)	80 (-32)	23 (25)	60 (-12)	**48** (0)	**49** (-1)
0.995	24	21 (3)	55 (-31)	11 (13)	38 (-14)	**22** (2)	**24** (0)
Panel C: NIKKEI 225							
		AR-GARCH		**AR-GARCH-t**		**Conditional GPD**	

Probability	Expected	Upper tail	Lower tail	Upper tail	Lower tail	Upper tail	Lower tail
0.95	268	199	302	220	320	**234**	**237**
		(69)	(-34)	(48)	(-52)	(34)	(31)
0.975	134	94	162	98	164	**132**	**143**
		(40)	(-28)	(36)	(-30)	(2)	(-9)
0.99	54	**51**	87	33	61	**51**	**57**
		(3)	(33)	(21)	(-7)	(3)	(-3)
0.995	27	31	57	21	23	**27**	**28**
		(-4)	(-30)	(6)	(4)	(0)	(1)

6. Concluding remarks

In this paper we follow McNeil and Frey´s (2000) two-step estimation procedure (conditional EVT) with the aim of comparing this methodology with other conventional methods such as those that combine GARCH models with Student's t or normal distributions for tail estimation of financial data. In step one, we fit a GARCH-type model to the return data by maximizing the log-likelihood function of the sample assuming normal innovations. In step two, the standardized residuals computed in step one are considered as realizations of a white noise process and EVT (in particular, the POT approach) is used to estimate the tails of innovations.

Specifically, we test the conditional EVT approach and the above mentioned traditional methods by applying them to the log return series of DJI, FTSE 100 and NIKKEI 225 stock indices. To do so, both in-sample and out-of-sample estimations are conducted.

According to our results, within the in-sample estimation, the EVT methodology produces the most accurate estimates for the upper tail of the three considered financial indices. In particular, this is true at all the contemplated confidence levels for the DJI and the FTSE 100 indices and at the 99% and 99.5% levels for the NIKKEI 225 index. However, the normal AR-GARCH and the AR-GARCH-t models generally provide more accurate estimations for the lower tails. Specifically, the former performs better at the 95% and 97.5% levels whereas the latter do so at the 99% and 99.5% levels of the DJI and the NIKKEI 225 indices.

More interestingly for a risk manager, whose aim is to know how well he is able to predict future extreme events rather than to model the past, the superiority of the conditional EVT methodology over the other two conventional methods is clearly evidenced under the out-of-sample estimation.

Thus, the accuracy of the conditional EVT tail estimates is confirmed for both the upper and the lower tails, since it provides the most accurate estimates in 11 (9) out of 12 cases for the right (left) tail. Such a result has been obtained when applied to the extreme returns for the three financial indices involved in this study, and by extension, it should be applied to other financial assets. In fact, these financial indices were chosen because they can be considered as representative of three important financial areas and no remarkable differences in terms of the accuracy of the estimates have arisen between them.

To conclude, on the one hand the results found in this paper should be useful to investors in general, since their goal is to be able to forecast unforeseen price movements and take advantage of them by positioning themselves in the market according to these predictions. On the other hand, precise (out-of-sample) predictions of the probability of extreme returns are of great importance for risk traders who implement dynamic portfolio hedging and need to design active strategies on a daily basis.

References

Assaf, A., 2009. Extreme observations and risk assessment in the equity markets of MENA region: Tail measures and Value-at-Risk. *International Review of Financial Analysis* 18, 109-116

Bekiros, S.D. and Georgoutsos, D. A. 2005. Estimation of Value-at-Risk by extreme value and conventional methods: a comparative evaluation of their predictive performance. *International Financial Markets, Institutions and Money*, 15, 209-228.

Byström, H.N.E., 2004. Managing extreme risks in tranquil and volatile markets using conditional extreme value theory. *International Review of Financial Analysis*, 13, 133-152.

Byström, H.N.E., 2005. Extreme value theory and extremely large electricity price changes. *International Review of Economics and Finance* 14, 41-55.

Christoffersen, P. and Diebold, F. 2000. How relevant is volatility forecasting for financial risk management. *Review of Economics and Statistics*, 82, 1-11.

Coles, S., 2003. An Introduction to Statistical Modeling of Extreme Values. *Springer Series in Statistics*. Second Edition.

Danielsson, J. and de Vries, C., 1997. Beyond the Sample: Extreme Quantile and Probability Estimation. *Mimeo*.

Danielsson, J. and de Vries, C. 2000. Value-at-Risk and Extreme Returns. *Annales d´Economie et de Statistique*, 60, 239-270.

Danielsson, J. and Morimoto, Y., 2000. Forecasting Extreme Financial Risk: a Critical Analysis of Practical Methods for the Japanese Market, Disc. Paper No. 2000-E-8, *Institute for Monetary and Economic Studies, Bank of Japan*.

Davison, A.C. and Smith, R.L., 1990. Models for exceedances over high thresholds (with discussion). *J.R. Statist. Soc.* B 52, 393-442.

Dowd, K.., J. Cotter, C. Humphrey and Woods M., 2008: How unlucky is 25-sigma. *Journal of Portfolio Management*, 34 (4), 76-80.

Embrechts, P., Klüppelberg, C. and Mikosch, T., 1999. Modelling Extremal Events for Insurance and Finance. *Springer, Berlin*.

Engle, R., González-Rivera, G., 1991. Semi-parametric ARCH models. *Journal of Business and Economic Statistics*, 9 (4), 345-359.

Fernández, V., 2005. Risk management under extreme events. *International Review of Financial Analysis*, 14, 113-148.

Gilli, M. and Këllezi, E., 2006. An Application of Extreme Value Theory for Measuring Financial Risk. *Computational Economics*, 27, 207-228.

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

Lee, T.-H., and Saltoglu, B. 2001. Evaluating Predictive Performance of Value-at-Risk Models in Emerging Markets: A Reality Check, *Mimeo, Marmara University*.

Liow, K.H., 2008. Extreme returns and value at risk in international securitized real estat markets. *Journal of Property Investment & Finance*, 26 (5), 418-446.

Longin, F., 2000. From Value at Risk to stress testing: the Extreme Value approach. *Journal of Banking and Finance* 24, 1097-1130.

McNeil, A. J., 1998. Calculating Quantile Risk Measures for Financial Return Series using Extreme Value Theory. *ETH Zürich*.

McNeil, A.J., and 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.

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