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

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

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 S&P 500, FTSE 100 and NIKKEI 225 representing three important financial areas in the world. Our results indicate that EVT-based conditional quantile estimates are more accurate than those from conventional GARCH models assuming normal or Student's t distribution innovations when doing not only in-sample but also out-of-sample estimation. Moreover, these results are robust to alternative GARCH model specifications. The findings of 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.

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 spikes 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.

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 the financial literature, the study of the tails of the distributions has mostly focused on the lower tail, with the estimation of the so-called Value at Risk (VaR) being the main application. VaR is a generalized measure of market risk which indicates 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 $(1-\alpha)$. Thereby, in probabilistic terms, VaR is a quantile of

the loss distribution. However, positive extreme observations are also of relevance beyond the estimation of huge benefits for long investors, since they mean losses for short investors.

In fact, the main goal of rational investors is to make money during both bearish and bullish trends in the market. Long position traders invest as buying a stock, holding it while it appreciates in price, and eventually sell it for profit. They encounter risk when the price of the stock decreases. On the other hand, the short trading position traders reacts exact opposite where they firstly sell the stock with the intention to later buy it back at a lower price. Therefore, the risk comes from a rise in the price of the stock. Thus, both the upper and lower tails are relevant to trading positions.

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; Danielsson and de Vries (1997), who test the predictive performance of various VaR methods for simulated portfolios of seven US stocks and conclude 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)[2], Assaf (2009)[3] and Bekiros and Georgoutsos (2005)[4]. Neftci (2000) uses EVT to forecasting the tails of the distributions of interest rate and exchange rate changes and also obtains that extreme value theory provides much more accurate results than the standard value at risk calculated from the

---

[2] 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.

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

[4] 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.

normal distribution. Danielsson and Morimoto (2000) apply EVT to Japanese financial data to confirm the accuracy and stability of this methodology over the GARCH-type techniques. McNeil and Frey (2000) combine fitting of GARCH models to estimate the current volatility and EVT for estimating the lower tail of the innovation distribution of the GARCH model. To backtest the method, they focus on the series of negative returns of the Standard and Poors index, the DAX index, the BMW share, the US dollar British pound exchange rate and the gold. Their approach outperforms both the unconditional EVT models and the GARCH models with normally distributed innovations and Student's t-innovations. Following McNeil and Frey (2000), Byström (2004) and Fernández (2005) lead to similar results. The former 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 above all for more extreme VaR quantiles. The latter, also focusing on the left tail of the distribution of returns, 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. 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.

Bali and Neftci (2003) analyze the fluctuations of the maximal and minimal changes in short-term interest rates and test the significance of time-varying paths followed by the mean and volatility of extremes. A conditional extreme value approach to calculating value at risk by specifying the location and scale parameters of the generalized Pareto distribution as a function of past information is proposed. Based on the estimated VaR thresholds, the statistical theory of extremes is found to provide more accurate estimates of the rate of occurrence and the size of extreme observations. Chavez-Demoulin et al. (2005) propose an approach that models within cluster behaviour, involving a self-exciting process for the exceedance times. More recently, Bali and Weinbaum (2007) introduce a conditional extreme value volatility estimator based on

high-frequency intraday returns and conclude that EVT provides more accurate forecasts than the implied volatility index and GARCH volatility models, and that implied volatility index generally yields a less accurate characterization of realized volatility than EVT and GARCH models. Finally, Bellini and Figà-Talamanca (2007) study the tail behaviour of eight major market indexes stratifying data according to the violation of a high threshold on the previous day. However, their results suggest that EVT-based methodologies are not sufficient conservative and produce too many violations[5] which may be due, according to the authors, to the choice of the threshold after which the generalized Pareto distribution is fitted.

Thus, while 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 (e.g. Danielsson and de Vries, 2000; Longin, 2000; Bekiros and Georgoutsos, 2005; Gilli and Këllezi, 2006; Assaf, 2009) others employ a conditional EVT approach (e.g. McNeil and Frey (2000), Bali and Neftci (2003); Bali and Weinbaum (2007) and Bellini and Figà-talamanca (2007)). In this paper, we follow McNeil and Frey (2000) and proceed 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 here on forth) 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 EVT estimator has the advantage of treating the tails separately. By applying the POT method to the standardized residuals from the normal 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 that assume normal- and Student's t-distributed innovations are additionally presented for comparative purposes. It should be noted that Chavez-Demoulin et al (2005) claimed that one drawback of this methodology is the fact that, as being a two-stage procedure, the results of the EVT analysis would be sensitive to the fitting of the GARCH model to the dataset in the first stage.

---

[5] A violation is defined to be an extreme observation being not identified by a particular method.

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 1987 (variable depending on the stock index) to 2011. Our sample extends that from McNeil and Frey (2000) focusing on two other relevant financial index such as FTSE 100 and NIKKEI 225. 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. Furthermore, to address the previously mentioned supposed drawback of this methodology, we fit alternative GARCH model specifications in the first stage of the estimation and show that the EVT based method is able to identify extreme observations much better than the rest of the models involved in the present study, regardless the particular GARCH model selected to filter data.

Also, apart from considering the lower tail of the distribution, which is the most frequent choice, we additionally analyze the upper tail of the distribution[6]. 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. Thereby, 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. Our empirical results evidence that the EVT provides a convenient framework for asymmetric properties in the distributions. This finding is important because asymmetry is often present in financial time series data.

Finally, in order 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.

---

[6] The upper tail of several securitized real estated and stock market returns is also analysed in Gilli and Këllezi (2006) and in Liow (2008). Gilli and Këllezi (op.cit.)) apply both the EVT *Block Maxima* and the *Peak Over Threshold* approaches to estimate both the left- and right-tail distribution of a set of financial series of returns. In contrast to our study, their aim is simply to illustrate such an estimation and not to make a comparative evaluation between different methodologies. Liow (op.cit.) models the maxima and minima of ten returns financial indices using the EVT *Block Maxima* approach, although the accuracy of the EVT estimates for the upper tail is not tested. The *Block Maxima* approach is different from the one chosen in this study, namely the *Peak over Threshold*. A brief explanation of the *Block Maxima* methodology and the reason of our choice is included in Section 3.

This graphical tool is frequently employed to determine the threshold directly from visual inspection. In this paper we additionally use the likelihood test ratio as a robustness check.

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, UK and Japan. The selected financial indices are S&P 500, FTSE 100 and NIKKEI 225. Our sample respectively covers the following periods: December 4, 1987 to October 17, 2011; December 11, 1987 to October 17, 2011 and October 7, 1987 to October 17, 2011. 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. Of note is 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.

[Insert Table 1 about here]

[Insert Figure 1 about here]

[Insert Figure 2 about here]

[Insert Figure 3 about here]

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

---

[7] The Jarque-Bera statistic is χ^2_2 distributed under the null of normality.

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

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 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

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

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). 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.

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. 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, three alternative GARCH 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. the Student's t.

To obtain conditional POT estimates, the standardized residuals resulting from fitting the 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]) are used to estimate the tails of innovations applying POT.

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 GARCH-type models selected to capture the dependencies shown in the log return series as well as the corresponding estimation results. We fit a GARCH(1,1), EGARCH(1,1) and TGARCH(1,1) as three alternative GARCH specifications in order to check whether the EVT based estimation results appear to be robust, or whether contrarily, they are sensitive to the choice of the GARCH model. Then we apply the POT approach to the residuals from the 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[10].

4.1. GARCH models

When looking for the best fitted AR-GARCH model to data, differences in the dynamics of the considered index log return series such as volatility clustering or seasonality patterns need individual analysis. Our results indicate that autocorrelation in both the returns themselves and in the squared returns can be mostly removed by simply fitting a GARCH (1,1) model for all the financial indexes involved in the study, i.e. S&P 500, FTSE 100 and NIKKEI 225 indexes[11]. As previously mentioned, an EGARCH (1,1) and a TGARCH(1,1) models have also been fitted to

---

[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).
[10] The computations presented in this study are conducted by means of Eviews (first stage estimations) and R software (EVT estimations).
[11] 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).

data. Maximum likelihood estimates for each of the involved index series are reported in Table 2.

[Insert Table 2 about here]

Table 3 displays the descriptive statistics of the standardized normal residuals 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 ten lags are presented), since they indicate some remaining autocorrelation both for the returns in levels and for the squared returns at lag 10 just within the out-of-sample estimation for the S&P 500 index and also for the squared returns at lag 10 within the out-of-sample estimation for the NIKKEI 225 index.

[Insert Table 3 about here]

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

Firstly, it is crucial the choice of a proper threshold so as to consider as extreme any observation exceeding it. According to the mean residual life plot, in which the mean excess over a threshold is potted as a function of the threshold itself, and to the test likelihood ratio, the threshold should be the one from which the linear model fits better than the quadratic one. 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 from the GARCH (1,1) specifications[12] are shown in Figures 4-6 and in Figures 7-9 respectively for in-sample estimation and out-of-sample estimation. Thus, the thresholds for the in-sample estimation under the GARCH(1,1) specification, should be the following: $u^{*}_{S\&P}=$ 1.85, $u^{*}_{FTSE}=1.42$ and $u^{*}_{NIKKEI}=2.26$, considering as extreme values 3%, 7% and 1% of data, respectively.

[Insert Figure 4 about here]

---

[12] Similar figures for the E-GARCH and T-GARCH specifications are available from the authors upon request.

13

[Insert Figure 5 about here]

[Insert Figure 6 about here]

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 GARCH model 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 which are fixed by applying the 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 mean-residual-life-plot method and reinforced by the test likelihood ratio, under the GARCH(1,1) specification[13], for the out-of-sample estimation are displayed in Figures 7-9. Thus, these thresholds are $u^*_{S\&P}$ = 0.55, u^*_{FTSE} = 2.10 and u^*_{NIKKEI} = 2.26, considering as extreme values 26%, 1% and 1% of data, respectively.

[Insert Figure 7 about here]

[Insert Figure 8 about here]

[Insert Figure 9 about here]

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

---

[13] Similar figures for the E-GARCH and T-GARCH specifications are available from the authors upon request.

are shown respectively in Tables 4 and 5. Also, the corresponding thresholds are shown. For instance, under the EGARCH specification, within the in-sample estimation, we obtain the estimates $\hat{\xi} = -0.170(-0.047, 0.024)$ and $\hat{\sigma} = 0.638$ (0,679, 0,628) for the S&P 500 (FTSE 100, NIKKEI 225) index. On the other hand, under the TGARCH specification, within the out-of-sample estimation, we obtain the estimates $\hat{\xi} = -0.277$ (0.102, 0,161) and $\hat{\sigma} = 0.593$ (0.364, 0.541) for the S&P 500 (FTSE 100, NIKKEI 225) index.

[Insert Table 4 about here]

[Insert Table 5 about here]

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. Thus, the threshold level finally chosen depends on the particular series and the number of data exceeding the corresponding threshold is logically different according to this threshold level, 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.

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, within the in-sample and out-of-sample estimation under the GARCH(1,1) specification[14]. Thus, within the in-sample estimation, the thresholds suggested by the mean residual life plot method together with the ratio likelihood test are $u^*_{S\&P} = -2.01$, $u^*_{FTSE} = -2.45$ and $u^*_{NIKKEI} = -2.76$, leaving 2%, 1% and 4% of data below each of them, respectively. As can be observed, most of the selected thresholds for the lower tail are greater, in absolute value, than the ones for the upper tail, which is a sign of asymmetry in the series. The corresponding GPD parameter estimates are displayed in Table 4. Since the shape parameter, ξ, gives an indication of the heaviness of the tail (the larger ξ, the heavier the tail), results lead us to conclude that the lower

---

[14] Similar figures for the E-GARCH and T-GARCH specifications are available from the authors upon request.

15

tail of the standardized normal residuals distribution is heavier than the upper tail for the three financial indices considered in the study, with the exceptions of the S&P 500 and FTSE 100 when the TGARCH specification is used.

[Insert Figure 10 about here]

[Insert Figure 11 about here]

[Insert Figure 12 about here]

Under the out-of-sample estimation, however, the thresholds for the GARCH(1,1) specification should be $u^*_{S\&P}$ = -2.46, u^*_{FTSE} = -2.74 and u^*_{NIKKEI} = -1.85, leaving 3.0%, 1.2% and 0.6% of data above each of them, respectively. These thresholds together with GPD parameter estimates are shown in Table 5. The asymmetry of the distribution is again evidenced by comparing the upper and lower tails in terms of the estimated shape parameters and thresholds. On the one hand, the estimated thresholds for the lower tail within the out-of-sample estimation are always higher, in absolute value, than the ones for the upper tail with the exceptions being the NIKKEI 225 under both the GARCH and EGARCH specification. On the other hand, regarding the estimated shape parameters, the results obtained from the in-sample estimation generally remain constant under the out-of-sample estimation except for the FTSE 100 index that exhibits shape parameters for the upper tail higher than the ones for the lower tail, meaning that, in this case, the upper tail is heavier than the lower tail.

[Insert Figure 13 about here]

[Insert Figure 14 about here]

[Insert Figure 15 about here]

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.

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 for a determined probability[15]. 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 S&P 500 (Panel A), FTSE 100 (Panel B) and NIKKEI 225 (Panel C) indices. Also reported (in brackets) is the p-value for the binomial test developed by McNeil and Frey (2000) of the success of these quantile estimation methods. This test is based on the number of violations defined as the difference between the theoretically expected and the estimated excesses. It is a two-sided binomial test of the null hypothesis that a method correctly estimates the conditional quantiles against the alternative that a method has a systematic estimation error and gives too few or too many violations. A p-value less than or equal to 0.05 will be interpreted as evidence

---

[15] For example, the expected number of excesses of a 95% tail quantile over a sample of 6020 observations is 301 ($0.05 \cdot 6020$)·

against the null. To help the reader with the comparison, closer numbers of estimated excesses to theoretically expected ones are highlighted in bold.

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 S&P 500 at all the considered levels of probabilitiy since in all the cases (including the GARCH(1,1), EGARCH(1,1) and TGARCH(1,1) specifications) the conditional GPD approach correctly estimates the conditional quantiles according to the binomial test and in 11 out of 12 cases it is the closest to the benchmark. On no occasion does this approach fail. In contrast, the conditional normal (t) approach fails in 5 (9) out of 12 cases. Regarding the FTSE 100, the conditional GPD approach correctly estimates the conditional quantiles in 7 out of 12 cases whereas the conditional normal (t) does in 5 (4) out of 12 cases. Finally, for the NIKKEI 225, the EVT-based estimated conditional quantiles would be correct in 8 out of 12 cases, being so in 6 (3) out of 12 cases for the normal (t) approach.

Moving to the lower tail, the conditional GPD approach correctly estimates the conditional quantiles for the S&P 500 in 10 out of 12 cases, being closest to the mark in 8 out of these 10 cases (in one case the EVT and the conditional t approaches are joint best). The corresponding figures for the normal and the t are, respectively, 7 and just 4 out of 12 cases. Similar results are obtained for the FTSE 100. Finally, the conditional GPD and the t approach under any of the three GARCH specifications provide similar results for the lower tail of the NIKKEI 225. They would be correctly estimating in 9 out of 12 cases each, both being closest to the mark in 5 out of the 12 cases, whereas the normal approach would only provide correct estimates in 3 out of 12 cases.

[Insert Table 6 about here]

5.2.   Out-of-sample evaluation

Within the out-of-sample estimation (Table 7), results indicate that the GPD conditional quantile estimates are even better than within the in-sample analysis. In fact, for both the upper and lower tails, they would be correctly estimated according to the binomial test in 20 (22, 16) out of 24 cases for the S&P 500 (FTSE 100, NIKKEI 225), being more accurate than the conventional GARCH models assuming normal- or Student's t-innovations in 18 (21, 12) out of 24 cases.

Of note are the results for the lower tail of the S&P 500. Whereas the GPD conditional quantiles would be correctly estimated in 9 out of 12 cases, these figures contrast with the ones for the normal approach which fails in all the cases and with the t approach which correctly estimates the conditional quantiles in just 1 out of 12 cases. Our results for the lower tail of the FTSE 100 give even stronger support to the conditional GPD approach since it provides correctly estimated conditional quantiles in 11 out of 12 cases and they are always closest to the theoretical mark. On the other hand, just in 4 (2) out of 12 cases the conditional normal (t) returns correctly estimates according to the binomial test. With regards to the lower tail of NIKKEI 225, in spite of the fact that GPD conditional quantile estimates are, according to the binomial test, correctly estimated in 10 out of 12 cases, however, in this case, the conditional t approach does perform quite well too, providing correct estimations in 8 out of 12 cases. The conditional normal approach provides correct quantile estimates in just 3 out of 12 cases.

[Insert Table 7 about here]

On the other hand, at most of the considered confidence levels, both for the in-sample and the out-of-sample estimation, the GARCH models combined with normal or Student's t-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. In this sense, it should be pointed out that the alternative GARCH-type specifications considered in this study, i.e. the EGARCH(1,1) and the TGARCH (1,1), differ from the GARCH(1,1) in that they are able to capture an asymmetric behaviour of the conditional volatility meaning that a negative shock

leads to a relatively higher conditional variance than a positive shock. However, there is no evidence for any substantial difference between the EVT empirical results resulting from fitting alternative GARCH-type specifications to data in the first stage of the empirical exercise. This finding is important because one drawback that has been argued against this methodology was the fact that the empirical results were sensitive to the particular GARCH specification used to pre-whiten the residuals. We have employed three alternative GARCH specifications in the first stage and obtain a better performance of the EVT-based approach in most of the cases, regardless the GARCH specification finally chosen, what may lead us to conclude that the second stage EVT results in detecting extreme observations are robust.

As an example, backtesting from January 4, 2010 to October 17, 2011 for the FTSE 100 index within the out-of-sample estimation is graphically illustrated in Figure 16. It shows the negative log returns, the GPD, Normal and t conditional quantile estimates at the 99 per cent of probability.

[Insert Figure 16 about here]

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 EVT (in particular, the POT approach) is used to estimate the tails of innovations.

One criticism of this methodology that was pointed out by Chavez-Demoulin et al. (2005) is that, as being a two-stage procedure, the results of the EVT analysis are sensitive to the fitting of a GARCH model in the first stage. In order to address this issue, we fit alternative GARCH specifications so as to compare the estimation results. Specifically, we test the conditional EVT

approach and the above mentioned traditional methods under a GARCH(1,1), EGARCH(1,1) and TGARCH(1,1) specifications by applying them to the log return series of S&P 500, FTSE 100 and NIKKEI 225 stock indices. Both in-sample and out-of-sample estimations are conducted.

According to our results, within the in-sample estimation, the EVT methodology generally produces the most accurate estimates of the three considered financial indices. But more interestingly for a risk manager, whose aim is to know how well she 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 better performance of the conditional EVT tail estimates is confirmed for both the upper and the lower tails, within in- and out-of-sample estimation, since, according to the binomial test proposed by McNeil and Frey (2000), it provides correct estimations in 56 (58) out of 72 cases for the upper (lower) tail, contrasting with the 39 (22) of the conditional normal or the 23 (28) of the conditional t. No evidence of any difference in the conditional EVT quantile estimates due to the particular GARCH specification is found. Then, despite the claimed drawback of the methodology used in this paper, the empirical results are shown to be robust. Such results have been achieved when applied to the extreme returns for the three financial indices involved in this study, and by extension, it may 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* *n*18, 109-116

Bali, T.G. and Neftci, S.C., 2003. Disturbing extremal behavior of spot rate dynamics. Journal of Empirical Finance, 10 (4), 455-477.

Bali, T.G. and Weinbaum, D., 2007. A conditional extreme value volatility estimator based on high-frequency returns. Journal of Economic Dynamics & Control, 31, 361–397.

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.

Bellini, F. and Figà-Talamanca, G., 2007. Conditional tail behaviour and Value at Risk. Quantitative Finance, 7 (6), 599–607.

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.

Chavez-Demoulin, V., Davison, A.C. and McNeil, A. J. (2005). Estimating value-at-risk: a point process approach. Quantitative Finance, 5 (2), 227–234.

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

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

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.

Neftci, S.N., 2000. Value at Risk Calculations, Extreme Events, and Tail Estimation. Journal of Derivatives, 7 (3), 23-37.

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

Table 1. Descriptive Statistics

Log returns on S&P 500, 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(10) [Q^2(10)] are the Ljung-Box tests for autocorrelation at lags 1 and 10 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.

	S&P 500	FTSE 100	NIKKEI 225
Range Dates	12/4/1987 10/17/2011	12/11/1987 10/17/2011	10/7/1987 10/17/2011
N. Obs.	6020	6015	5913
Mean (%)	0.0002	0.0001	0.0001
Median (%)	0.0005	0.0004	0.00004
Standard Deviation (%)	0.0116	0.0113	0.0153
Minimum	-0.0946	-0.0926	-0.1613
Maximum	0.1095	0.0938	0.1323
Skewness	-0.2889	-0.1321	-0.2642
Kurtosis	11.63298	9.1194	11.71
Jarque-Bera	1877.8	9401.3	14723
(p-value)	(0.0000)	(0.0000)	(0.0000)
	t-statistic		
ADF	-59.19**	-34.82**	-58.29**
	p-value		
Q(1)	0.00*	0.402	0.087
Q(10)	0.00*	0.00*	0.00*
Q^2(1)	0.00*	0.00*	0.00*
Q^2(10)	0.00*	0.00*	0.00*

Table 2. Alternative GARCH Models

Panel A, B, C and D display GARCH(1,1), EGARCH(1,1) and TGARCH (1,1) parameters estimates for the S&P 500, FTSE 100 and NIKKEI 225 indices. d.f. is degrees of freedom. * (**) denotes statistical significance at a 1% (5%) level.

GARCH(1,1)

$$X_t = \phi_0 + \sigma_t \varepsilon_t$$

$$\sigma^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

EGARCH(1,1)

$$X_t = \phi_0 + \sigma_t \varepsilon_t$$

$$log(\sigma_t^2) = \alpha_0 + \alpha_1 \cdot \left| \frac{\varepsilon_{t-1}}{\sigma_{t-1}} \right| + \alpha_2 \cdot \frac{\varepsilon_{t-1}}{\sigma_{t-1}} \cdot$$

TGARCH(1,1)

$$X_t = \phi_0 + \sigma_t \varepsilon_t$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \cdot \varepsilon_{t-1}^2 + \alpha_2 \cdot \varepsilon_{t-1}^2 \cdot d_{t-1} + \beta$$

	Normal	Student´s t		Normal	Student´s t		Normal	Student´s t
Panel A: S&P 500								
ϕ_0	4.89E04*	6.31E04*	ϕ_0	2.22E04**	4.19E04*	ϕ_0	2.62E04**	4.56E04*
α_0	1.00E-06*	5.57E-07*	α_0	-0.2266*	-0.1864	α_0	1.30E-06*	8.68E-07*
α_1	0.0635*	0.0597*	α_1	0.1072*	0.1075*	α_1	0.0212**	0.0146**
			α_2	-0.0868	-0.0905*	α_2	0.1022*	0.1048*
β_1	0.9286*	0.9376*	β_1	0.9843*	0.9889	β_1	0.9334*	0.9369*
d.f.		6.2119*	d.f.		6.7592*	d.f.		6.7971*
Panel B: FTSE 100								
ϕ_0	4.11E04*	4.77E04*	ϕ_0	2.70E04**	2.57E04**	ϕ_0	1.99E04**	2.81E04*
α_0	1.27E-06*	1.15E-06*	α_0	-0.2237*	-0.2126*	α_0	1.36E-06*	1.26E-06*
α_1	0.0848*	0.0787*	α_1	0.1306*	0.1248*	α_1	0.0217*	0.0162*
			α_2	-0.0700*	-0.0737	α_2	0.0882	0.0923
β_1	0.9054*	0.9120*	β_1	0.9869*	0.9877*	β_1	0.9207*	0.9244*
d.f.		12.7905*	d.f.		14.1443*	d.f.		13.4661*
Panel C: NIKKEI 225								
ϕ_0	4.91E04*	4.06E04*	ϕ_0	3.42E04**	3.17E04**	ϕ_0	2.07E04**	2.76E04**
α_0	2.63E-06*	1.87E-06*	α_0	-0.3530*	-0.3115*	α_0	3.09E-06*	2.44E-06*
α_1	0.1102*	0.0964*	α_1	0.1741*	0.1605*	α_1	0.0314*	0.0275*
			α_2	-0.1055	-0.0980	α_2	0.1427*	0.1272
β_1	0.8846*	0.9003*	β_1	0.9746*	0.9785*	β_1	0.8868*	0.8996*
d.f.		7.6088*	d.f.		8.8466*	d.f.		8.4533*

Table 3. Descriptive Statistics of standardized normal residuals

IS and OS are in-sample and out-of-sample estimation. Q(1) [$Q^2(10)$] and Q(10) [$Q^2(10)$] are the Ljung-Box tests for autocorrelation at lags 1 and 10 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.

	S&P500						FTSE 100						NIKKEI 225					
	GARCH		EGARCH		TGARCH		GARCH		EGARCH		TGARCH		GARCH		EGARCH		TGARCH	
	IS	OS	IS	OS	IS	OS	IS	OS	IS	OS	IS	OS	IS	OS	IS	OS	IS	OS
Mean (%)	-0.02	-0.02	0.002	0.008	-0.002	0.006	-0.02	-0.02	0.002	0.006	-0.001	0.004	-0.05	-0.05	-0.01	-0.009	-0.02	-0.01
Median (%)	0.01	0.01	0.04	0.04	0.03	0.04	0.005	0.007	0.03	0.03	0.03	0.03	-0.03	-0.03	0.001	0.006	-0.003	0.004
St.Dev. (%)	0.99	1.01	1.00	1.04	1.00	1.03	1.00	1.00	1.00	1.03	1.00	1.01	1.00	1.01	1.00	1.00	1.003	1.01
Maximum	4.11	3.41	3.53	4.41	3.78	3.46	6.01	6.21	5.86	6.23	6.18	6.28	6.26	7.44	5.74	5.86	6.04	6.42
Minimum	-9.85	-7.27	-9.65	-7.99	-9.97	-7.44	-4.91	-4.72	-5.33	-5.47	-5.50	-5.99	-11.41	-7.13	-10.69	-6.75	-10.70	-6.95
Skewness	-0.55	-0.45	-0.55	-0.47	-0.57	-0.468	-0.21	-0.21	-0.18	-0.21	-0.19	-0.24	-0.41	-0.15	-0.29	-0.13	-0.33	-0.14
Kurtosis	6.27	4.95	6.05	5.00	6.28	4.95	3.79	3.89	3.76	3.95	3.84	4.01	7.30	5.01	6.25	4.59	6.44	4.66
Q(1)	0.39	0.28	0.39	0.29	0.51	0.48	0.24	0.35	0.21	0.35	0.24	0.48	0.40	0.56	0.47	0.28	0.38	0.49
Q(10)	0.07	0.03*	0.11	0.03*	0.15	0.55	0.38	0.15	0.40	0.14	0.43	0.21	0.36	0.27	0.28	0.40	0.30	0.46
$Q^2(1)$	0.52	0.48	0.17	0.09	0.08	0.05	0.38	0.99	0.76	0.60	0.39	0.49	0.22	0.36	0.57	0.58	0.96	0.55
$Q^2(10)$	0.64	0.03*	0.66	0.00*	0.84	0.04*	0.51	0.07	0.34	0.07	0.45	0.51	0.96	0.97	0.99	0.01*	0.99	0.41

Table 4. Threshold and In-sample estimation

Panel A, B, C respectively show as 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 S&P 500, FTSE 100 and NIKKEI 225 indices.

| | GPD parameters estimates | | | | | |
| | Upper tail | | | Lower tail | | |
	GARCH	EGARCH	TGARCH	GARCH	EGARCH	TGARCH
	Panel A: S&P 500					
σ	0.467	0.638	0.579	0.541	0.516	0.577
	(0.04)	(0.02)	(0.02)	(0.05)	(0.06)	(0.02)
ξ	-0.100	-0.170	-0.144	0.211	0.272	-0.146
	(0.07)	(0.02)	(0.02)	(0.08)	(0.09)	(0.02)
u	1.85	0.87	1.06	2.01	2.07	1.15
	Panel B: FTSE 100					
σ	0.380	0.679	0.557	0.602	0.617	0.528
	(0.02)	(0.13)	(0.11)	(0.09)	(0.12)	(0.03)
ξ	0.082	-0.047	0.045	0.083	-0.037	0.024
	(0.05)	(0.11)	(0.13)	(0.11)	(0.14)	(0.05)
u	1.42	2.41	2.47	2.45	2.54	1.56
	Panel A: NIKKEI 225					
σ	0.718	0.628	0.464	0.490	0.451	0.419
	(0.14)	(0.13)	(0.03)	(0.14)	(0.05)	(0.05)
ξ	0.006	0.024	0.007	0.536	0.213	0.294
	(0.13)	(0.15)	(0.04)	(0.27)	(0.08)	(0.10)
u	2.26	2.41	1.36	2.76	2.02	2.20

Table 5. Threshold and Out-of-sample estimation

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 S&P 500, FTSE 100 and NIKKEI 225 indices.

| | GPD parameters estimates | | | | | |
| | Upper tail | | | Lower tail | | |
	GARCH	EGARCH	TGARCH	GARCH	EGARCH	TGARCH
	Panel A: S&P 500					
σ	0.749	0.601	0.593	0.580	0.755	0.545
	(0.02)	(0.02)	(0.05)	(0.09)	(0.10)	(0.16)
ξ	-0.202	-0.132	-0.277	0.170	0.054	0.363
	(0.02)	(0.02)	(0.06)	(0.11)	(0.09)	(0.25)
u	0.55	1.08	1.75	2.46	2.33	3.00
	Panel B: FTSE 100					
σ	0.3558	0.309	0.364	0.575	0.521	0.543
	(0.07)	(0.03)	(0.02)	(0.12)	(0.07)	(0.06)
ξ	0.273	0.230	0.102	-0,091	0.092	0.074
	(0.17)	(0.08)	(0.05)	(0.15)	(0.10)	(0.09)
u	2.10	1.71	1.49	2.74	2.08	1.98
	Panel A: NIKKEI 225					
σ	0.591	0.856	0.541	0.518	0.383	0.466
	(0.13)	(0.198)	(0.09)	(0.05)	(0.062)	(0.08)
ξ	0.176	-0.090	0.161	0.123	0.258	0.205
	(0.17)	(0.16)	(0.14)	(0.07)	(0.12)	(0.13)
u	2.26	2.43	2.11	1.85	2.27	2.39

Table 6. In sample evaluation

In sample evaluation of estimated (positive and negative) tail quantiles at different probabilities for the S&P 500, FTSE 100 and NIKKEI 225 indices (Panel A, B and C). Closer numbers of estimated excesses to theoretically expected ones are highlighted in bold. The binomial probability of statistical difference between violations are in parenthesis.

		Panel A: S&P 500					
		Upper tail			**Lower tail**		
Prob.	Expected	**GARCH**	**GARCH-t**	**Cond. GPD**	**GARCH**	**GARCH-t**	**Cond. GPD**
0.95	301	252 (0.00)	275 (0.13)	**301 (1.00)**	**311 (0.55)**	359 (0.00)	257 (0.00)
0.975	151	120 (0.01)	119 (0.00)	**153 (0.84)**	192 (0.00)	**159 (0.48)**	160 (0.43)
0.99	60	56 (0.65)	25 (0.00)	**59 (0.95)**	121 (0.00)	42 (0.02)	**54 (0.48)**
0.995	30	29 (0.93)	7 (0.00)	**27 (0.65)**	70 (0.00)	13 (0.00)	**27 (0.65)**
		EGARCH	**EGARCH-t**	**Cond. GPD**	**EGARCH**	**EGARCH-t**	**Cond. GPD**
0.95	301	269 (0.06)	291 (0.57)	**301 (1.00)**	**307 (0.70)**	352 (0.00)	259 (0.01)
0.975	151	118 (0.00)	121 (0.01)	**138 (0.32)**	171 (0.09)	182 (0.01)	**148 (0.9)**
0.99	60	**59 (0.95)**	23 (0.00)	**59 (0.95)**	105 (0.00)	69 (0.24)	**60 (1.00)**
0.995	30	**23 (0.23)**	9 (0.00)	**23 (0.23)**	70 (0.00)	35 (0.36)	**27 (0.65)**
		TGARCH	**TGARCH-t**	**Cond. GPD**	**TGARCH**	**TGARCH-t**	**Cond. GPD**
0.95	301	255 (0.00)	277 (0.16)	**299 (0.93)**	286 (0.39)	346 (0.00)	**296 (0.76)**
0.975	151	120 (0.01)	117 (0.00)	**151 (0.97)**	139 (0.36)	**146 (0.74)**	146 (0.74)
0.99	60	**60 (1.00)**	25 (0.00)	60 (0.79)	67 (0.36)	31 (0.00)	**59 (0.95)**
0.995	30	25 (0.41)	7 (0.00)	**24 (0.52)**	**31 (0.85)**	9 (0.00)	23 (0.23)

Table 6. In sample evaluation (continued)

		Panel B: FTSE 100					
		Upper tail			Lower tail		
Prob.	Expected	**GARCH**	**GARCH-t**	**Cond. GPD**	**GARCH**	**GARCH-t**	**Cond. GPD**
0.95	301	223 (0.00)	233 (0.00)	**314 (0.42)**	**331 (0.08)**	352 (0.00)	368 (0.00)
0.975	150	87 (0.00)	89 (0.00)	**160 (0.41)**	180 (0.02)	187 (0.00)	**164 (0.26)**
0.99	60	41 (0.01)	32 (0.00)	**56 (0.65)**	101 (0.00)	84 (0.00)	**59 (0.95)**
0.995	30	29 (0.93)	22 (0.17)	**30 (1.00)**	62 (0.00)	48 (0.00)	**33 (0.58)**
		EGARCH	**EGARCH-t**	**Cond. GPD**	**EGARCH**	**EGARCH-t**	**Cond. GPD**
0.95	301	244 (0.00)	253 (0.00)	1117 (0.00)	**319 (0.27)**	335 (0.04)	437 (0.00)
0.975	150	96 (0.00)	98 (0.00)	464 (0.00)	**169 (0.13)**	176 (0.04)	210 (0.00)
0.99	60	**49 (0.17)**	34 (0.00)	85 (0.00)	89 (0.00)	71 (0.15)	**63 (0.70)**
0.995	30	**32 (0.71)**	25 (0.45)	33 (0.58)	54 (0.00)	41 (0.05)	**30 (1.00)**
		TGARCH	**TGARCH-t**	**Cond. GPD**	**TGARCH**	**TGARCH-t**	**Cond. GPD**
0.95	301	236 (0.00)	250 (0.00)	508 (0.00)	314 (0.42)	332 (0.07)	**304 (0.84)**
0.975	150	94 (0.00)	94 (0.00)	211 (0.00)	162 (0.34)	170 (0.11)	**151 (0.93)**
0.99	60	47 (0.09)	37 (0.09)	**59 (0.95)**	94 (0.00)	73 (0.10)	**61 (0.9)**
0.995	30	**33 (0.85)**	24 (0.31)	**33 (0.85)**	61 (0.00)	41 (0.05)	**32 (0.71)**

Table 6. In sample evaluation (continued)

		Upper tail			Lower tail		
		Panel C: NIKKEI 225					
Prob.	Expected	**GARCH**	**GARCH-t**	**Cond. GPD**	**GARCH**	**GARCH-t**	**Cond. GPD**
0.95	301	215 (0.00)	245 (0.00)	722 (0.00)	340 (0.00)	361 (0.00)	**129 (0.06)**
0.975	150	94 (0.00)	102 (0.00)	279 (0.00)	184 (0.00)	182 (0.00)	96 (0.00)
0.99	60	48 (0.17)	35 (0.00)	**62 (0.69)**	95 (0.00)	**57 (0.84)**	52 (0.39)
0.995	30	34 (0.41)	21 (0.14)	**30 (0.92)**	55 (0.00)	26 (0.58)	**28 (0.85)**
		EGARCH	**EGARCH-t**	**Cond. GPD**	**EGARCH**	**EGARCH-t**	**Cond. GPD**
0.95	301	243 (0.00)	252 (0.00)	431 (0.00)	**298 (0.88)**	323 (0.11)	252 (0.00)
0.975	150	111 (0.00)	111 (0.00)	172 (0.04)	154 (0.59)	155 (0.53)	**149 (0.9)**
0.99	60	**58 (0.95)**	43 (0.04)	56 (0.74)	91 (0.00)	**53 (0.47)**	**53 (0.47)**
0.995	30	40 (0.06)	27 (0.71)	**28 (0.85)**	48 (0.00)	**27 (0.71)**	26 (0.58)
		TGARCH	**TGARCH-t**	**Cond. GPD**	**TGARCH**	**TGARCH-t**	**Cond. GPD**
0.95	296	236 (0.00)	256 (0.02)	**308 (0.46)**	319 (0.00)	352 (0.00)	212 (0.00)
0.975	148	103 (0.00)	103 (0.00)	**143 (0.74)**	163 (0.21)	161 (0.28)	**136 (0.36)**
0.99	59	**57 (0.84)**	39 (0.00)	52 (0.39)	93 (0.00)	**57 (0.84)**	56 (0.74)
0.995	30	37 (0.16)	26 (0.58)	**31 (0.78)**	53 (0.00)	**29 (1.00)**	28 (0.85)

Table 7. Out-of-sample evaluation

Out-of-sample evaluation of estimated (positive and negative) tail quantiles at different confidence levels for the S&P 500, FTSE 100 and NIKKEI 225 indices ((Panel A, B and C). Closer numbers of estimated excesses to theoretically expected ones are highlighted in bold. The binomial probability of statistical difference between violations are in parenthesis.

		Panel A: S&P 500					
		Upper tail			Lower tail		
Prob.	Expected	**GARCH**	**GARCH-t**	**Cond. GPD**	**GARCH**	**GARCH-t**	**Cond. GPD**
0.95	251	215 (0.02)	**243 (0.63)**	236 (0.35)	273 (0.00)	310 (0.00)	193 (0.00)
0.975	125	109 (0.02)	110 (0.17)	**120 (0.65)**	174 (0.00)	172 (0.00)	**124 (0.93)**
0.99	50	54 (0.57)	31 (0.00)	**49 (0.94)**	113 (0.00)	76 (0.00)	**49 (0.94)**
0.995	25	33 (0.13)	7 (0.00)	**29 (0.42)**	77 (0.00)	33 (0.13)	**23 (0.76)**
		EGARCH	**EGARCH-t**	**Cond. GPD**	**EGARCH**	**EGARCH-t**	**Cond. GPD**
0.95	251	259 (0.60)	290 (0.01)	**258 (0.65)**	281 (0.00)	320 (0.00)	**276 (0.10)**
0.975	125	**120 (0.65)**	119 (0.59)	**120 (0.65)**	166 (0.00)	167 (0.00)	**131 (0.62)**
0.99	50	65 (0.04)	34 (0.02)	**47 (0.72)**	112 (0.00)	81 (0.00)	**49 (0.94)**
0.995	25	33 (0.13)	7 (0.13)	**23 (0.76)**	81 (0.00)	43 (0.00)	**24 (0.92)**
		TGARCH	**TGARCH-t**	**Cond. GPD**	**TGARCH**	**TGARCH-t**	**Cond. GPD**
0.95	251	**244 (0.67)**	280 (0.06)	291 (0.01)	281 (0.00)	317 (0.00)	122 (0.00)
0.975	125	**113 (0.67)**	116 (0.42)	119 (0.59)	172 (0.00)	181 (0.00)	87 (0.00)
0.99	50	69 (0.01)	32 (0.00)	**53 (0.67)**	113 (0.00)	74 (0.00)	**47 (0.72)**
0.995	25	32 (0.16)	5 (0.00)	**28 (0.55)**	73 (0.00)	37 (0.02)	**24 (0.92)**

Table 7. Out-of-sample evaluation (continued)

		Panel B: FTSE 100					
		Upper tail			Lower tail		
Prob.	Expected	**GARCH**	**GARCH-t**	**Cond. GPD**	**GARCH**	**GARCH-t**	**Cond. GPD**
0.95	251	195 (0.00)	202 (0.00)	181 (0.00)	275 (0.12)	288 (0.02)	**256 (0.72)**
0.975	125	80 (0.00)	82 (0.00)	**114 (0.32)**	147 (0.05)	154 (0.01)	**112 (0.24)**
0.99	50	35 (0.03)	23 (0.00)	**51 (0.89)**	86 (0.00)	67 (0.02)	**49 (0.94)**
0.995	25	21 (0.48)	14 (0.03)	**23 (0.76)**	57 (0.00)	45 (0.00)	**23 (0.76)**
		EGARCH	**EGARCH-t**	**Cond. GPD**	**EGARCH**	**EGARCH-t**	**Cond. GPD**
0.95	251	**239 (0.48)**	244 (0.7)	238 (0.44)	540 (0.07)	544 (0.05)	413 (0.00)
0.975	125	89 (0.48)	92 (0.00)	**127 (0.86)**	288 (0.02)	302 (0.00)	**254 (0.82)**
0.99	50	42 (0.29)	30 (0.00)	**51 (0.89)**	150 (0.03)	162 (0.00)	**122 (0.20)**
0.995	25	25 (1.00)	15 (0.04)	**25 (1.00)**	82 (0.00)	69 (0.01)	**50 (1.00)**
		TGARCH	**TGARCH-t**	**Cond. GPD**	**TGARCH**	**TGARCH-t**	**Cond. GPD**
0.95	251	209 (0.00)	220 (0.05)	**265 (0.35)**	**269 (0.23)**	236 (0.36)	253 (0.87)
0.975	125	81 (0.00)	87 (0.00)	**128 (0.86)**	**141 (0.16)**	98 (0.01)	**117 (0.5)**
0.99	50	37 (0.06)	30 (0.00)	**47 (0.72)**	80 (0.00)	37 (0.06)	**53 (0.67)**
0.995	25	19 (0.27)	13 (0.01)	**20 (0.37)**	61 (0.00)	15 (0.04)	**26 (0.84)**

Table 7. Out-of-sample evaluation (continued)

|		Panel C: NIKKEI 225					
		Upper tail			Lower tail		
Prob.	Expected	**GARCH**	**GARCH-t**	**Cond. GPD**	**GARCH**	**GARCH-t**	**Cond. GPD**
0.95	246	210 (0.02)	208 (0.01)	317 (0.00)	295 (0.00)	316 (0.00)	**272 (0.09)**
0.975	123	91 (0.02)	95 (0.01)	158 (0.00)	152 (0.00)	157 (0.00)	**142 (0.08)**
0.99	49	47 (0.83)	27 (0.00)	**49 (1.00)**	89 (0.00)	57 (0.25)	**53 (0.57)**
0.995	25	27 (0.61)	16 (0.08)	**24 (1.00)**	51 (0.00)	30 (0.26)	**24 (1.00)**
		EGARCH	**EGARCH-t**	**Cond. GPD**	**EGARCH**	**EGARCH-t**	**Cond. GPD**
0.95	246	211 (0.02)	**225 (0.19)**	1102 (0.00)	**251 (0.72)**	286 (0.01)	152 (0.00)
0.975	123	90 (0.00)	94 (0.00)	364 (0.00)	**135 (0.27)**	137 (0.2)	108 (0.18)
0.99	49	**54 (0.57)**	33 (0.02)	61 (0.09)	82 (0.00)	**49 (1.00)**	46 (0.72)
0.995	25	31 (0.19)	**23 (0.84)**	26 (0.77)	42 (0.00)	23 (0.84)	**24 (1.00)**
		TGARCH	**TGARCH-t**	**Cond. GPD**	**TGARCH**	**TGARCH-t**	**Cond. GPD**
0.95	246	219 (0.08)	**229 (0.29)**	276 (0.05)	**267 (0.16)**	294 (0.00)	175 (0.00)
0.975	123	94 (0.00)	96 (0.013)	**129 (0.55)**	144 (0.05)	143 (0.07)	**109 (0.22)**
0.99	49	60 (0.13)	31 (0.00)	**52 (0.67)**	88 (0.00)	57 (0.25)	**48 (0.94)**
0.995	25	32 (0.13)	20 (0.42)	**25 (0.92)**	50 (0.00)	**24 (1.00)**	23 (0.84)

Figures
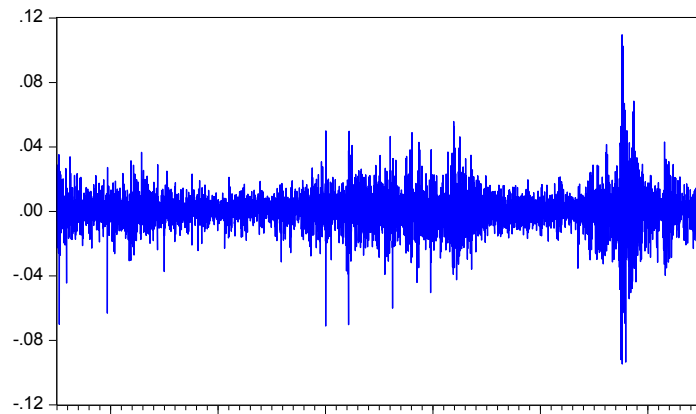
Figure 1. S&P 500 index log returns (12/4/1987 - 10/17/2011**)**
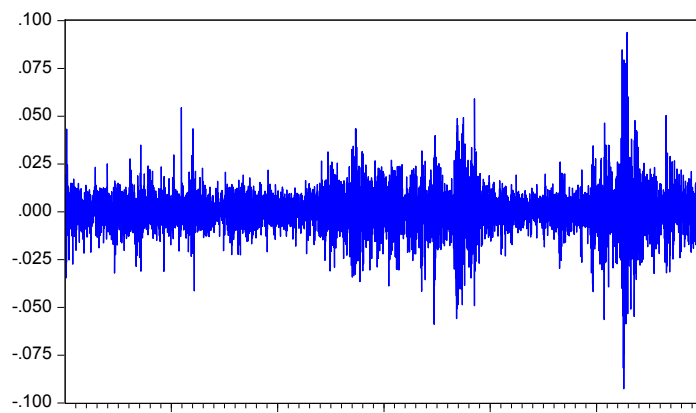
**Figure 2. FTSE 100 index log returns (**12/11/1987 - 10/17/2011**)**
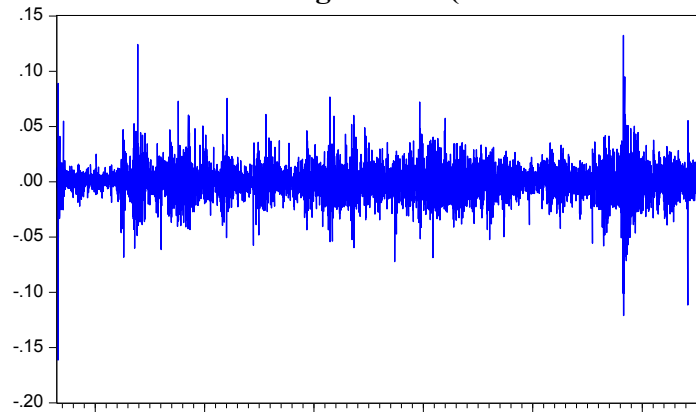
Figure 3. NIKKEI 225 index log returns (10/7/1987- 10/17/2011**)**

S&P 500 In-sample standardized normal residuals
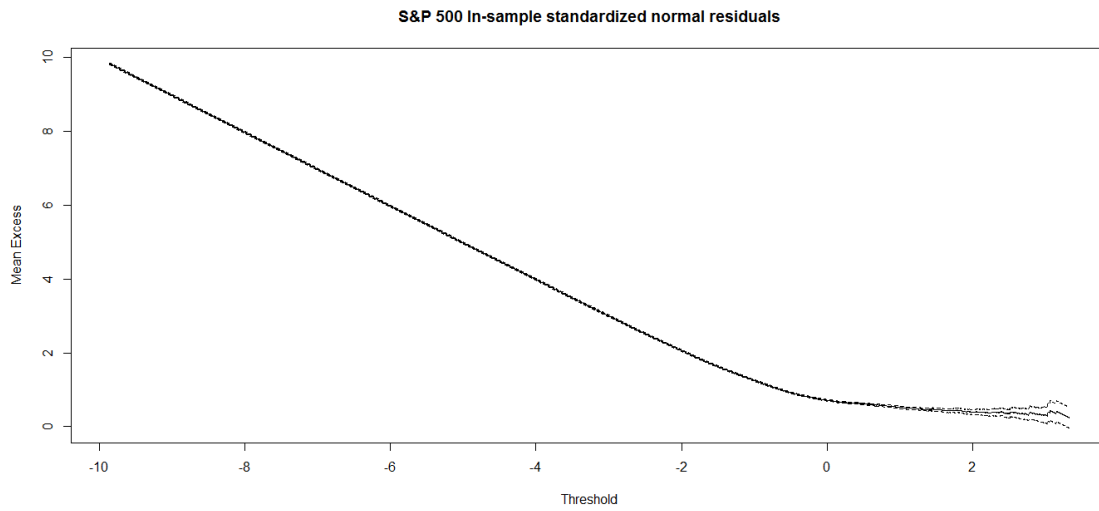
Log likelihood	u=1.83	u=1.84	**u=1.85**	u=1.86
Linear relationship	217.69	327.75	215.33	213.02
Quadratic relationship	221.30	216.21	216.85	213.95
Deviance Statistic	7.23	5.09	3.03*	1.87*

Figure 4. In-sample mean residual life plot and likelihood ratio tests.
Calculated at different potential thresholds for the S&P 500 standardized normal residuals (upper tail). * denotes statistical significance at 5% level.

FTSE 100 In-sample standardized normal residuals
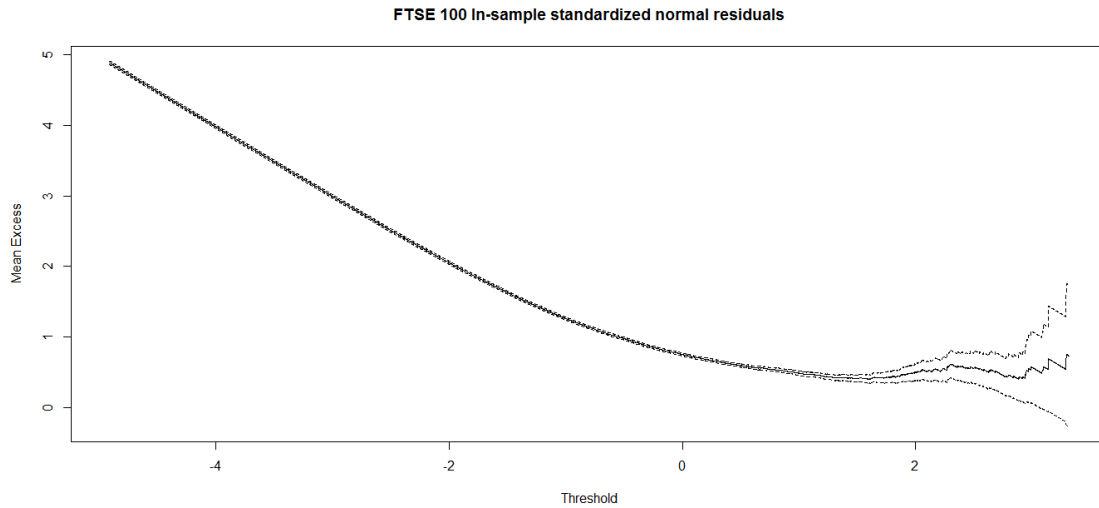
Log likelihood	u=1.40	u=1.41	**u=1.42**	u=1.43
Linear relationship	279.19	278.36	277.23	275.71
Quadratic relationship	282.69	280.97	279.10	277.01
Deviance Statistic	7.01	5.21	3.74*	2.60*

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.

**NIKKEI 225 In-sample standardized normal residuals**
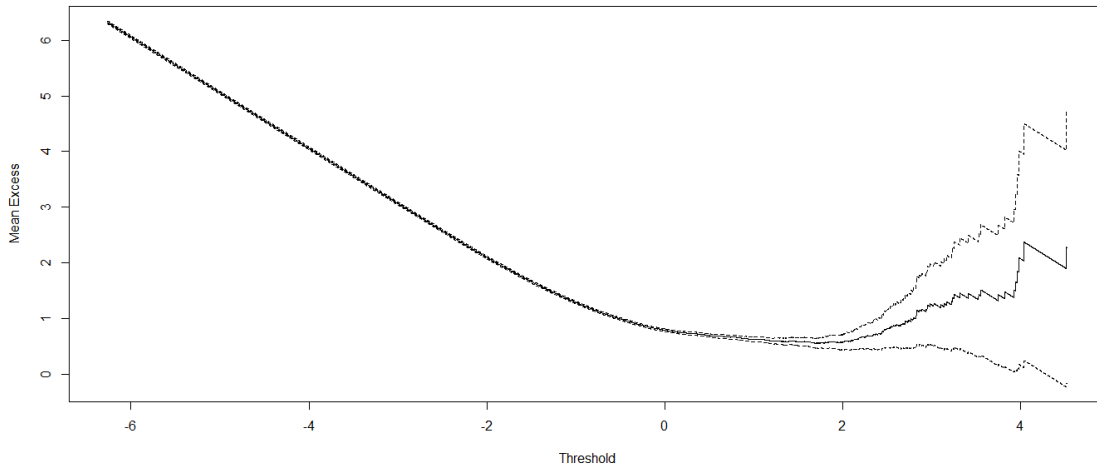
Log likelihood	u=2.24	u=2.25	**u=2.26**	u=2.27
Linear relationship	124.25	123.19	120.62	119.22
Quadratic relationship	127.37	125.13	122.02	119.85
Deviance Statistic	6.24	3.87	2.79*	1.26*

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.

S&P 500 Out-of-sample standardized normal residuals
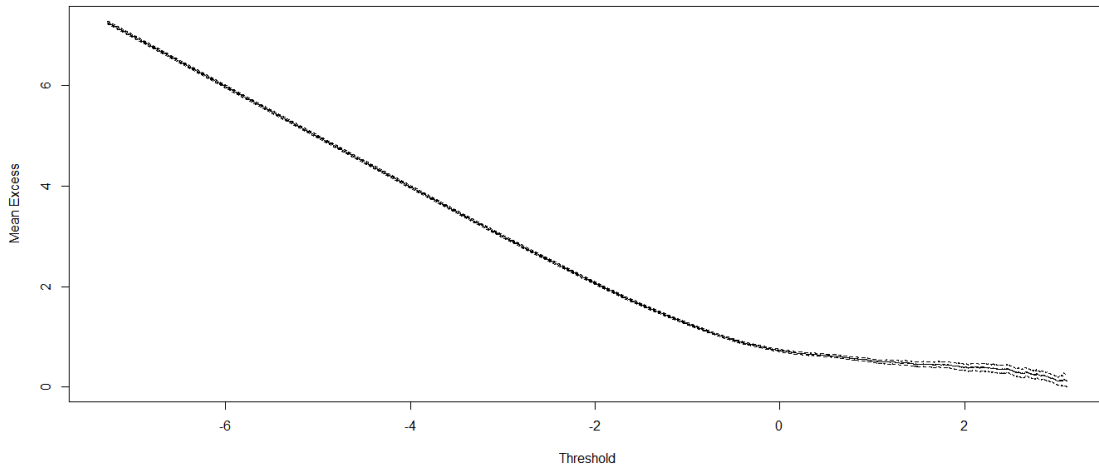
Log likelihood	u=0.53	u=0.54	**u=0.55**	u=0.56
Linear relationship	588.71	586.69	584.71	582.63
Quadratic relationship	591.53	589.03	586.60	584.13
Deviance Statistic	5.65	4.69	3.79*	3.01*

Figure 7. Out-of-sample mean residual life plot and likelihood ratio tests
Calculated at different potential thresholds for the S&P 500 index standardized normal residuals (upper tail). * denotes statistical significance at 5% level.

FTSE 100 Out-of-sample standardized normal residuals
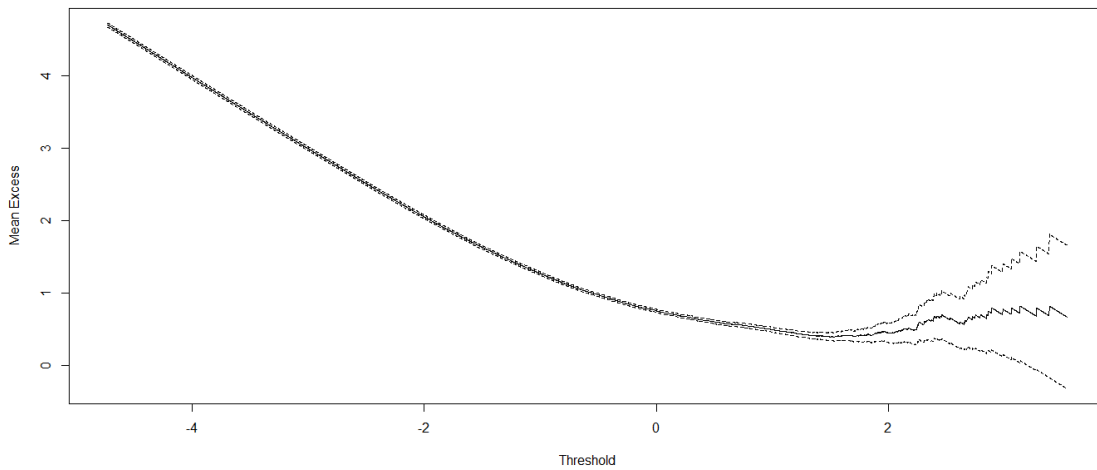
Log likelihood	u=2.08	u=2.09	**u=2.10**	u=2.11
Linear relationship	74.93	72.31	69.76	67.04
Quadratic relationship	77.37	74.39	71.42	68.48
Deviance Statistic	4.88	4.17	3.32*	2.88*

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.

NIKKEI 225 Out-of-sample standardized normal residuals
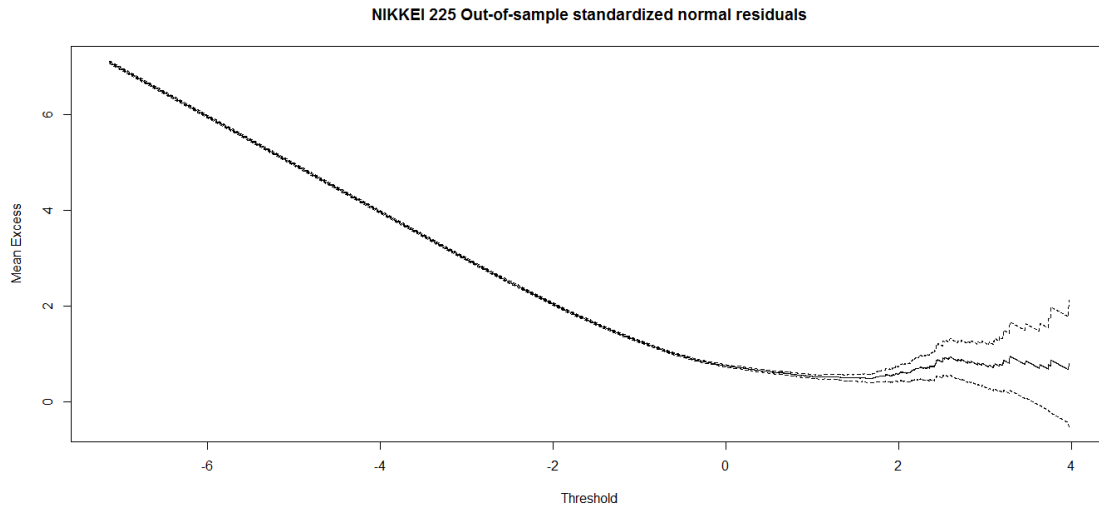
Log likelihood	u=2.24	u=2.25	**u=2.26**	u=2.27
Linear relationship	73.52	71.32	70.32	69.63
Quadratic relationship	76.52	73.91	71.87	70.29
Deviance Statistic	6.00	5.16	3.10*	1.32*

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.

S&P 500 In-sample negated standardized normal residuals
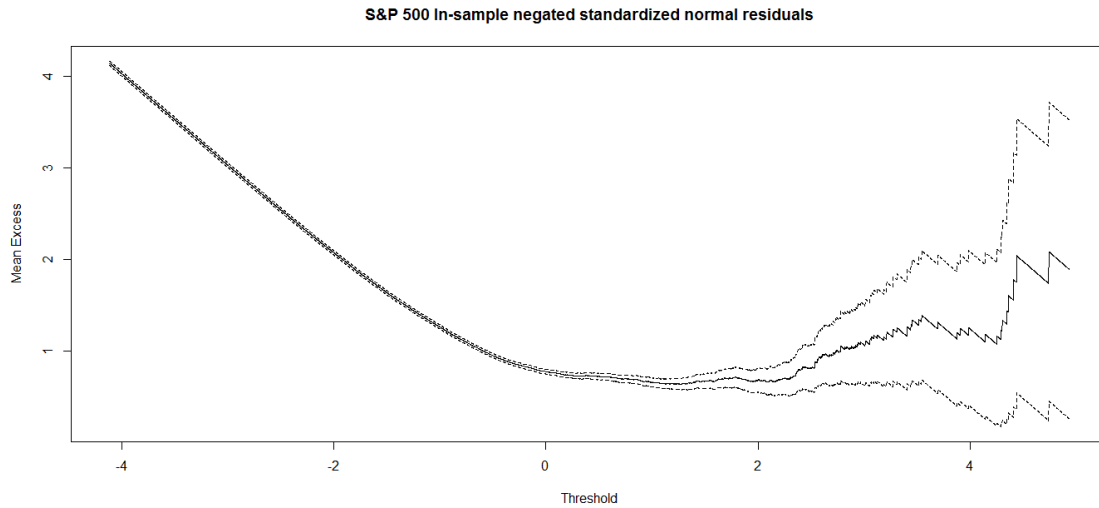
| Log likelihood	u=1.99	u=2.00	**u=2.01**	u=2.02
Linear relationship	232.89	233.20	233.75	234.99
Quadratic relationship	236.80	236.04	235.64	236.02
Deviance Statistic	7.81	5.69	3.77*	2.05*

Figure 10. In-sample mean residual life plot and likelihood ratio tests
Calculated at different potential thresholds for the S&P 500 index negated standardized normal residuals (lower tail). * denotes statistical significance at 5% level.

FTSE 100 In-sample negated standardized normal residuals
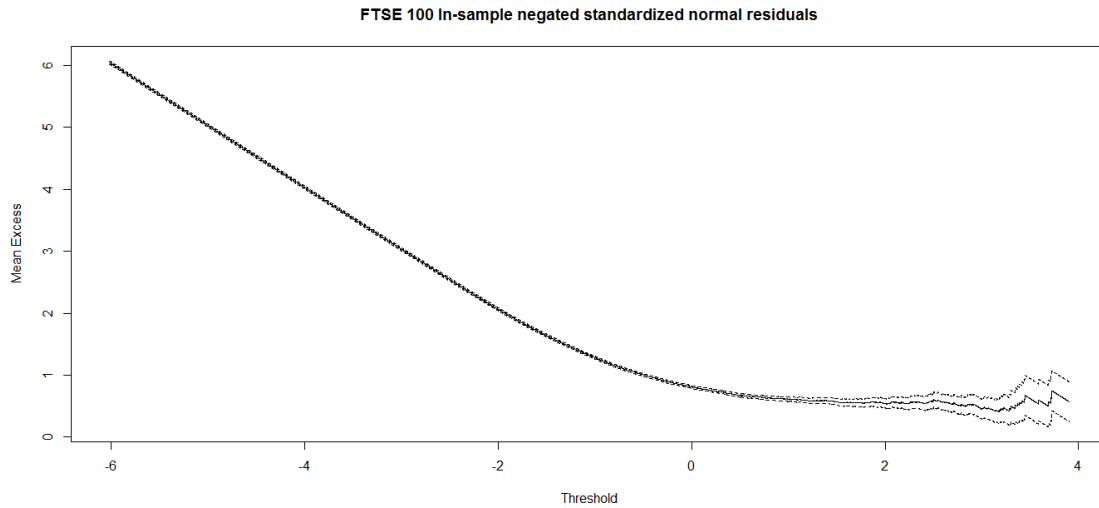
Log likelihood	u=2.43	u=2.44	**u=2.45**	u=2.46
Linear relationship	170.60	170.75	172.01	171.67
Quadratic relationship	174.44	173.18	173.16	172.08
Deviance Statistic	7.69	4.87	2.29*	0.80*

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.

NIKKEI 225 In-sample negated standardized normal residuals
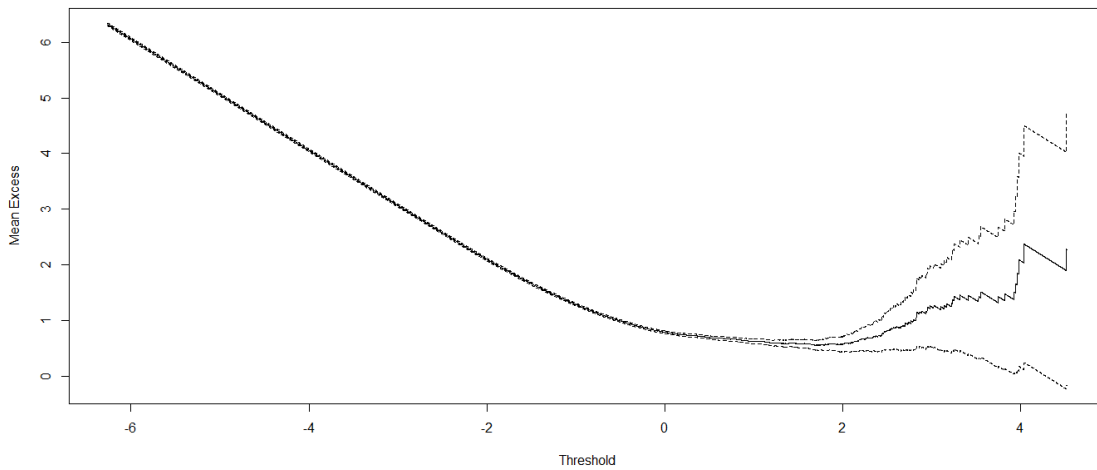
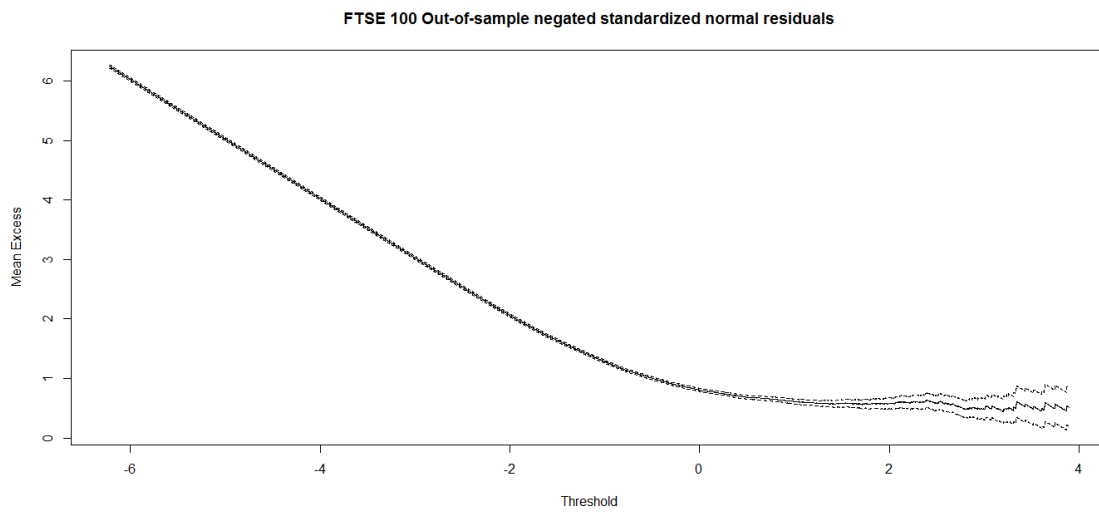| Log likelihood	u=2.74	u=2.75	**u=2.76**	u=2.77
Linear relationship	27.74	25.52	23.55	21.02
Quadratic relationship	30.20	27.46	24.70	22.17
Deviance Statistic	4.93	3.88	2.30*	2.32*

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.

S&P 500 Out-of-sample negated standardized normal residuals
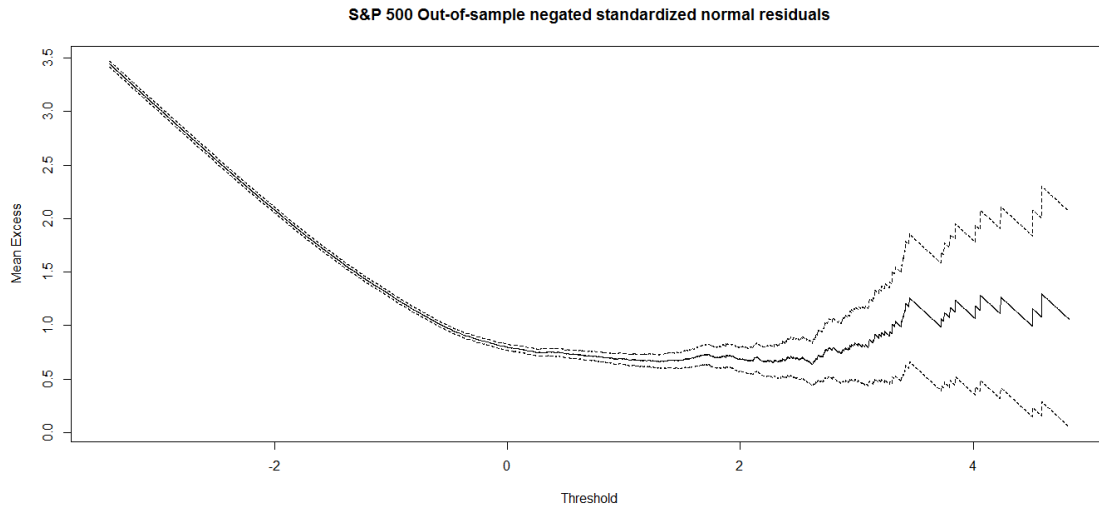
Log likelihood	u=2.44	u=2.45	**u=2.46**	u=2.47
Linear relationship	168.52	165.21	166.39	167.44
Quadratic relationship	171.91	170.03	168.16	166.30
Deviance Statistic	7.12	5.17	3.52*	2.18*

Figure 13. Out-of-sample mean residual life plot and likelihood ratio tests
Calculated at different potential thresholds for the S&P 500 index negated standardized
normal residuals (lower tail). * denotes statistical significance at 5% level.

FTSE 100 Out-of-sample negated standardized normal residuals

| Log likelihood	u=2.72	u=2.73	**u=2.74**	u=2.75
Linear relationship	96.14	95.30	93.69	96.14
Quadratic relationship	99.47	97.21	94.66	99.47
Deviance Statistic	6.66	3.82	1.93*	6.66*

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.

NIKKEI 225 Out-of-sample negated standardized normal residuals
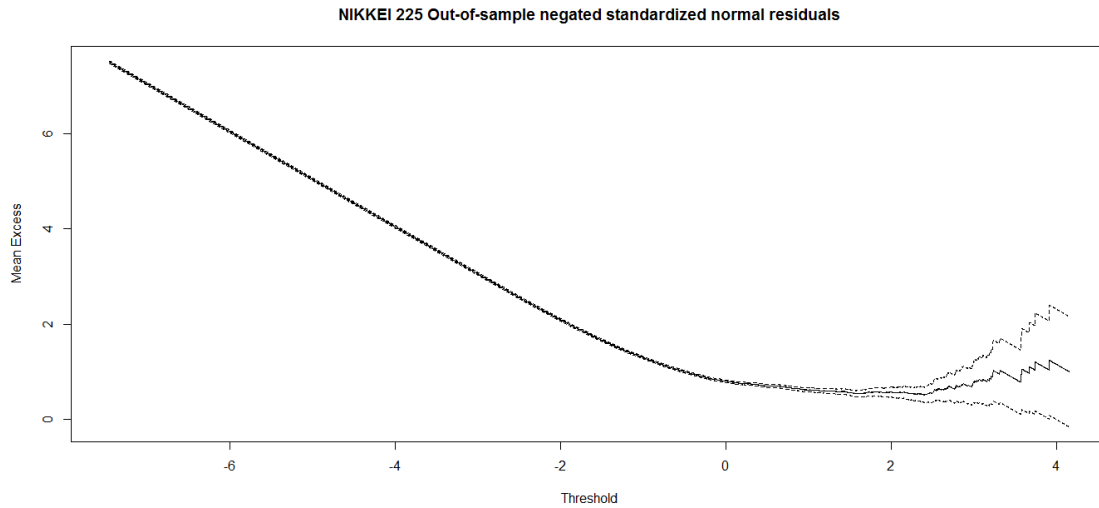
Log likelihood	u=1.83	u=1.84	**u=1.85**	u=1.86
Linear relationship	301.49	296.98	298.59	296.98
Quadratic relationship	304.12	297.88	299.96	297.88
Deviance Statistic	5.26	1.80	2.75*	1.80*

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.
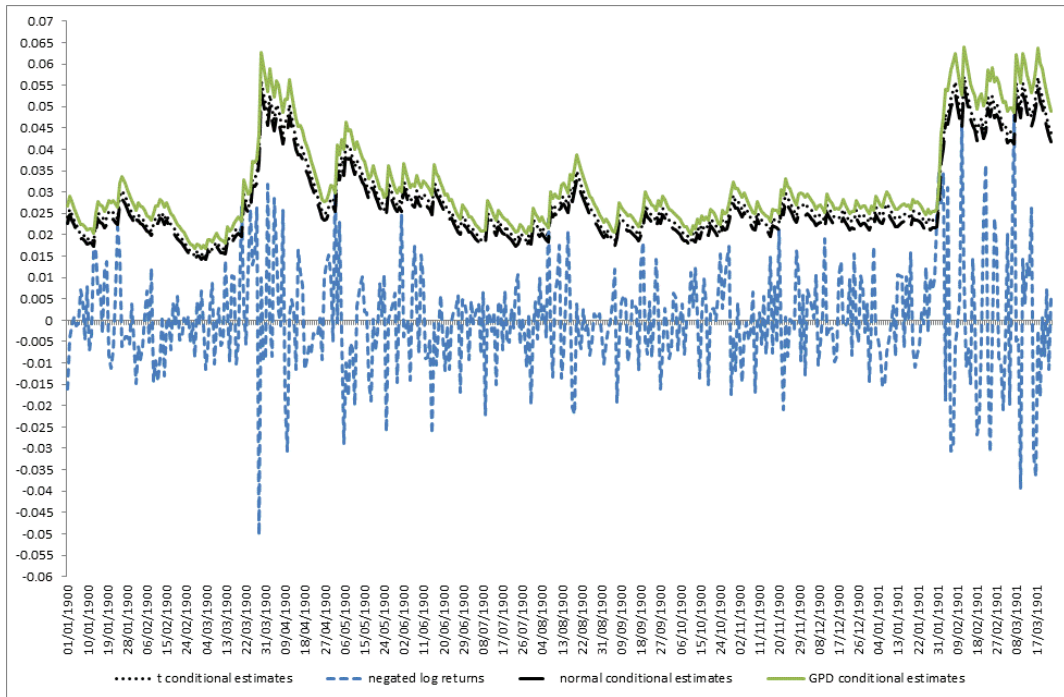
Figure 16. Backtesting from January 4, 2010 to October 17, 2011 for the FTSE 100 index. Conditional quantiles estimates of the GPD, Normal and t approaches at the 99 per cent of probability are superimposed to the negative log returns.