LUND UNIVERSITY
School of Economics and Management

Measuring Financial Risks by Peak Over Threshold Method

An application of Value-at-Risk and Expected Shortfall

by

Bing Zhang

August 2016

Master's Programme in Finance

Supervisor: Birger Nilsson

Abstract

Assessing the probability of rare and extreme events is an important issue in the risk management of financial portfolios. Extreme value theory provides the solid fundamentals needed for the statistical modelling of such events and the computation of extreme risk measures. The focus of the paper is on the use of the peak over threshold method under extreme value theory to compute right tail risk measures using Value at Risk and Expected Shortfall and to understand how models perform in different economic situations by back-testing, applying it to the S&P 500 Index and one of the Index - Ford Motor Company. Both unconditional and conditional GARCH(1,1) models with zero or non-zero parametric ξ are tested during different financial economical periods (2007-2011). Three different confidence levels with different thresholds are applied to two distributions. All models are evaluated by back-testing procedures. Extreme high confidence level or conditional models improve the results, thus conclude that the POT method does better at a higher confidence level and market volatility capturing is important in financial risk measurement.

Keywords: Extreme Value Theory, Peak Over Threshold, Value at Risk, Expected Shortfall, Back-testing, GARCH(1,1)

Acknowledgements

I would like to appreciate my supervisor, Birger Nilsson, for his guidance and valuable advice to my research and to appreciate my family for all supports during my master study. Furthermore, I would like to appreciate Ford Motor Company where I experienced my important five years and decided to have a further education. Last but not least, I would like to thank all professors, coordinators, classmates, and friends who encouraged, coached and helped me during my stay in Lund.

Table of Contents

1 Introduction

It has been eight years since last global financial crisis and about twenty years since last Asia financial crisis. It is interesting to know: “If things go wrong again, how wrong can they go?” The problem is then how to model the rare phenomena that lie outside the range of available observations. Extreme value theory (EVT) provides a robust theoretical foundation on which we can build statistical models describing extreme events and absolutely it has been a well-founded methodology since an interesting discussion about the potential of extreme value theory in risk management in Diebold et al. (1998).

Extreme value theory is the term used to describe the science of estimating the tails of a distribution. EVT can be used to improve Value at Risk (VaR) and Expected Shortfall (ES) estimates and to help in situations where analysts want to estimate VaR with a very high confidence level. It is a way of smoothing and extrapolating the tails of an empirical distribution.

The Generalized Pareto distribution (GPD) is the key distributions of EVT. The idea underlying EVT is to model the extreme outcomes rather than all outcomes because it is exactly these large losses that are relevant for estimating VaR and ES. However, the Peak Over Threshold (POT) has become the preferred extreme value approach in finance. By using all losses in a sample larger than some pre-specified threshold value, POT solves the problem of information loss that happens in traditional EVT.

This paper focus on the use of the peak over threshold method to compute the right tail risk measures using VaR and ES at different confidence levels and attempts to understand which model perform better during different economic situations by comparing the back-testing results using Kupiec tests and the 2^{nd} method proposed by Acerbi and Szekely (2015) respectively. Both unconditional and conditional GARCH(1,1) models are constructed. Furthermore, different threshold u is tested instead of using the rule of thumb to set u approximately equal to the 95^{th} percentile of the empirical distribution, thus in order to know how the choice of u will impact the models. As a limitation, only right tail is considered in the paper.

According to my research and application results, models perform better at high confidence levels - 99% or 99.9% and the estimation error is larger for ES than for VaR when fat-tailed distributions are used. The GARCH(1,1) models improve the test results, especially during financial crisis periods when there are more fluctuation log-returns. Moreover, the threshold u does impact on the results, but not significant.

The paper continues as follows: Section 2 presents the definitions of the risk measures and empirical methods that applied in this paper. Section 3 introduces the data selection and the methodology of testing models. In Section 4, a practical application result is presented where data and models are analyzed and compared. Final, the conclusion and limitation is presented in Section 5.

2 Theoretical Review

2.1 Value at Risk

Value at Risk (VaR) measures the size of an amount at risk of underlying assets or liabilities at a specific probability α and within a certain time period T. In other words, we are interested in making a statement of the following form:

"We are α percent certain that we will not lose more than l dollars in time T."

The variable L is the VaR of the portfolio. It is a function of two parameters: the time horizon, T, and the confidence level, α percent. It is the loss level during a time period of length T that we are α% certain will not be exceeded. The following equation defines value-at-risk (VaR) mathematically:

$$\mathrm{VaR}_{\alpha}(\mathrm{L}) = \min\{\mathrm{t}: \Pr(\mathrm{L} > \mathrm{l}) \leq 1 - \alpha\} \qquad (1)$$

When estimating VaR, a time horizon must be stated, in which market-related losses of the assets or liabilities might occur. This horizon depends on different circumstances; for instance the firm's tolerance in relation to actual risks, how often the underlying assets or liabilities require risk evaluations, and how easily the firm can liquidate or hedge large losses. One day, one week or quarterly are time horizons that are commonly used when calculating VaR. Furthermore, a relevant confidence level, α must be assumed for the underlying assets or liabilities, given the probability of a loss. The confidence interval is equivalent to $(1 - \alpha)$%, where α is the probability of the left tail of a risk distribution, the normal distribution as an example. More specific, α is called the VaR level or the critical probability, meaning that VaR should not be exceeded more than α% of the time period. The two most common confidence levels are 1 % and 5 %, depending on the probability in which losses occur.

2.2 Expected Shortfall

A coherent risk measure satisfies certain requirements, which we may think as “desirable” properties. Not all risk measures are coherent. The properties are:

1. Monotonicity: If a portfolio produces a worse result than another portfolio for every state of the world, its risk measure should be greater.
2. Translation Invariance: If an amount of cash K is added to a portfolio, its risk measure should go down by K.
3. Homogeneity: Changing the size of a portfolio by a factor λ while keeping the relative amounts of different items in the portfolio the same, should result in the risk measure being multiplied by λ.
4. Subadditivity: The risk measure for two portfolios after they have been merged should be no greater than the sum of their risk measures before they were merged.

VaR satisfies the first three conditions, but does not always satisfy the fourth one. Expected Shortfall (ES) is always coherent and it can produce better incentives for traders than VaR. This is also sometimes referred to as conditional value at risk, conditional tail expectation, or expected tail loss. Whereas VaR asks the question: “How bad can things get?” ES asks: “If things do get bad, what is the expected loss?” ES, like VaR, is a function of two parameters: T (the time horizon) and α (the confidence level). It is the expected loss during time T conditional on the loss being greater than the α^{th} percentile of the loss distribution and it is mathematically defined as the average value-at-risk for confidence levels larger than or equal to α:

$$ES_\alpha(L) = \frac{1}{1-\alpha}\int_\alpha^1 VaR_x\,(L)dx \qquad (2)$$

If the loss distribution is continuous, the definition of expected shortfall may be written in the somewhat more intuitive form:

$$ES_\alpha(L) = E[L\colon L > VaR_\alpha(L)] \qquad (3)$$

2.3 Extreme Value Theory

Extreme value theory is a way of smoothing the tails of the probability distribution of portfolio daily changes calculated using historical simulation. It leads to estimates of VaR and ES that reflect the whole shape of the tail of the distribution, not just the positions of a few losses in the tails. Extreme value theory can also be used to estimate VaR and ES when the confidence level is very high. For example, even if we have only 500 days of data, it could be used to come up with an estimate of VaR or ES for a confidence level of 99.9%.

The Generalized Pareto distribution (GPD) is the key distributions of Extreme Value Theory (EVT). The idea underlying EVT is to model the extreme outcomes rather than all outcomes because it is exactly these large losses that are relevant for estimating VaR and ES. However, the Peak Over Threshold model (POT) has become the preferred extreme value approach in finance. By using all losses in a sample larger than some pre-specified threshold value, POT solves the problem of information loss that happens in traditional EVT.

2.3.1 Peak Over Threshold Method

The Peak Over Threshold method (POT) has become the preferred extreme value approach in finance. Largely, the reason is the obvious and quite serious drawback of the traditional EVT that we are likely to throw away information by the block maxima method. Indeed, if there is more than one large loss in a given block, only the largest loss in the block is used in the subsequent analysis. Information loss of this kind is very likely to happen with financial data due to the well-known stylized fact of volatility clustering. This problem is “solved” in the POT model of extreme losses by using all losses in a sample larger than some pre-specified threshold value. Instead, of course, the problem of choosing the threshold is introduced. The threshold should be chosen such that all losses above the threshold are “extreme losses” in the sense of the underlying extreme value theory. This clearly leads to some arbitrariness in the choice of the threshold value and also to a non-trivial trade-off; for the underlying theory to go through we want to choose a high threshold, for the estimation of the parameters in the distribution of the extreme losses we want many observations above the threshold (i.e., we want to choose a low threshold).

The theory underlying the POT approach aims at modeling excess losses $L - u$, where u is the predetermined threshold value. Assume that L is a stochastic loss variable with cumulative density function F, i.e., $Pr(L \leq l) = F(l)$. We may then define a cumulative density function $Fu(l)$ for excess losses $L - u$ given that $L > u$:

$$F_u(l) = \Pr(L - u \leq l | L > u) = \Pr(L \leq l + u | L > u) \quad (4)$$

Defining the events $A: L \leq 1 + u$ and $B: L > u$ we can use the definition of a conditional probability $Pr(A|B)$ to obtain an explicit expression for $Fu(l)$ which is the basis for the POT extreme value theory:

$$F_u(l) = \Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} = \frac{F(l+u) - F(u)}{1 - F(u)} \quad (5)$$

Before proceeding we note that:

$$F_u(l - u) = \frac{F(l) - F(u)}{1 - F(u)} \quad (6)$$

an equality that contains $F(l)$. This is useful since by definition $F(l) = Pr(L \leq l)$ and therefore putting $l = VaR_\alpha$ and solving the equation $F\,(VaR_\alpha) = \alpha$ gives us an estimate of $VaR\alpha$ (because VaR_α is the α-quantile). From here on our focus is for this reason on the expression for $Fu(l - u)$. Solving for $F\,(l)$ we find:

$$F(l) = [1 - F(u)]F_u(l - u) + F(u) \quad (7)$$

The idea as explained above is now to put $F(VaR_\alpha) = \alpha$ and solve for the α-quantile VaR_α. The problem is that $Fu(l - u)$ is an unknown distribution and therefore this is not immediately possible. This is where the so called Pickands-Balkema-deHaan extreme value theorem comes in. Their theorem says that we can approximate $Fu(l - u)$ with a generalized Pareto distribution provided that u is a high enough threshold. Using this approximation we can explicitly and directly solve the equation $F(VaR_\alpha) = \alpha$.

The limit theorem by Pickands-Balkema-deHaan essentially states that under certain (quite weak) assumptions the limiting distribution of $Fu(l - u)$ as $u \to \infty$ is a generalized Pareto distribution (GPD):

$$G(l-u) = \begin{cases} 1-\left(1+\xi\frac{l-u}{\beta}\right)^{-\frac{1}{\xi}}, & \xi \neq 0 \\ 1-\exp\left(-\frac{l-u}{\beta}\right), & \xi = 0 \end{cases} \tag{8}$$

which from a practical perspective implies that $Fu(l - u) \approx G(l - u)$ for high values of the threshold u. The shape parameter ξ governs the tail behavior of the GPD distribution (similar, but not identical to the degrees of freedom parameter in the Student t-distribution. It has the opposite interpretation; higher values of ξ mean more probability in the right tail). The parameter β is a scale parameter (similar, but not identical to volatility). Note that the limit theorem makes no specific statement about the underlying (parent) distribution for the losses;

A natural estimate of the probability $F(u) = Pr(L \leq u)$ is the proportion of loss observationsnot exceeding u, i.e., $F(u) = (N - N_u)/N$, where N_u is the number of observations exceeding u and N is the total number of observations. Using this empirical probability as the estimate of $F(u)$, the POT estimate of VaR_α becomes ($\xi \neq 0$ case)

$$VaR_\alpha = u + \frac{\beta}{\xi}\left[\left(\frac{1-\alpha}{1-F(u)}\right)^{-\xi} - 1\right] \tag{9}$$

and again estimating $F(u)$ with $(N - N_u)/N$, the POT estimate of VaRα becomes ($\xi = 0$ case):

$$VaR_\alpha = u - \beta ln\left(\frac{N}{N_u}(1-\alpha)\right) \tag{10}$$

The parameters β and ξ can be estimated by Maximum likelihood (ML). Assume that we have observations of m losses above some threshold u. As already mentioned, the choice of threshold involves a non-trivial trade-off between "large u" and "large m". Denoting the threshold loss observations by $l_u^1, l_u^2, \ldots, l_u^m$, we can derive the log-likelihood functions as (the log of) the derivative of $G(l-u)$; denote this derivative $f(l)$. The probability density functions for the GPD distributions are therefore:

$$f(l) = \frac{1}{\beta}\left(1+\xi\frac{l-u}{\beta}\right)^{-\left(1+\frac{1}{\xi}\right)} \tag{11}$$

$$f(l) = \frac{1}{\beta}exp\left(-\frac{l-u}{\beta}\right) \tag{12}$$

Taking logs of the respective probability density functions and summing over the m observations $l_u^1, l_u^2, \ldots, l_u^m$, it follows that the corresponding log-likelihood functions are:

$$logL(\beta,\xi) = -mlg\beta - (1+\frac{1}{\xi})\sum_{i=1}^{m} ln\left(1+\xi\frac{l_u^i-u}{\beta}\right) \tag{13}$$

$$logL(\beta) = -mlg\beta - \frac{1}{\beta}\sum_{i=1}^{m}\left(l_u^i - u\right) \tag{14}$$

Maximizing the log-likelihood functions with a numerical maximizing algorithm (such as the Solver in Excel) with respect to the parameters, we obtain the ML estimates of the parameters $\beta\ and\ \xi$;

For both $\xi \neq 0\ and\ \xi = 0$, an analytical expression for ES_α can be derived using the definition of ES and the closed form expression for VaR derived:

$$ES_\alpha = \int_\alpha^1\left(u+\frac{\beta}{\xi}\left[\left(\frac{N}{N_u}(1-x)\right)^{-\xi}-1\right]\right)dx = \frac{VaR_\alpha+\beta-u\xi}{1-\xi}\ for\ \xi \neq 0 \tag{15}$$

$$ES_\alpha = \int_\alpha^1\left(u-\beta ln\left(\frac{N}{N_u}(1-x)\right)\right)dx = VaR_\alpha + \beta\ for\ \xi = 0 \tag{16}$$

2.3.2 Conditional POT

One suggestion in the literature to make the POT model more "dynamic" to take current market conditions into account is to instead apply the POT analysis to the standardized loss residuals and combine this with the usual GARCH/EWMA volatility models. Note however that often POT is used to estimate VaR_α at (very) high confidence levels α, with corresponding tail-events happening perhaps only once every 5 or every 10 years. In that situation is not obvious that making the POT more responsive to short run market conditions is logically meaningful. The α-quantile of the GPD distribution of the standardized residuals, denoted q_α, can then be used to estimate VaR_α according to the "hybrid method" (it involves the POT quantile q_α but the usual volatility σ_{T+1}):

$$VaR_\alpha = \mu + \sigma_{T+1}q_\alpha \tag{17}$$

where σ_{T+1} denotes the GARCH/EWMA volatility estimate one day out-of-sample (for the day we want to estimate VaR_{α}).

The standardized residuals to which we should apply the POT analysis are defined by:

$$\varepsilon_i^* = \frac{l_i - \bar{l}}{\sigma_i}, i = 1,2, \dots, T \qquad (18)$$

where $\bar{l}$ denotes the sample average of the observed losses $l_1, l_2, \dots, l_T$,. The α-quantiles for the distribution of standardized residuals based on the POT model are given by the expressions already derived in equation (9) and (10):

$$q_{\alpha} = u^* - \beta^* ln\left[\frac{N}{N_{u^*}}(1-\alpha)\right] \; Where \; \xi^* = 0 \qquad (19)$$

$$q_{\alpha} = u^* + \frac{\beta^*}{\xi^*}\left[\left(\frac{N}{N_u}(1-\alpha)\right)^{-\xi^*} - 1\right] \; Where \; \xi^* \neq 0 \qquad (20)$$

where the star-notation indicates that these parameter values and the threshold value are GPD parameters when the POT approach is applied to the standardized residuals $\varepsilon_1^*, \varepsilon_2^*, \dots, \varepsilon_T^*$, (not to the original loss data $l_1, l_2, \dots, l_T$). The volatility estimate for the "next day" can as usual be calculated from the GARCH or EWMA model directly:

$$\sigma_{T+1}^2 = \gamma_0 + \gamma_1 \varepsilon_T^2 + \gamma_2 \sigma_T^2 \text{ (GARCH)} \qquad (21)$$

$$\sigma_{T+1}^2 = (1-\lambda)\varepsilon_T^2 + \lambda \sigma_T^2 \text{ (EWMA)} \qquad (22)$$

Volatility Modeling GARCH(1,1)

Conditional forecasts incorporate the information available at each period of time and therefore are superior to unconditional forecasts. A common finding in empirical studies of financial markets is that the variance of the returns is not constant over time. Specifically, there are periods when volatility is relatively high, and periods when price movements are quite small. To capture this "volatility clustering" effect, one approach is to use the Generalized Autoregressive Conditionally Heteroscedastic model (GARCH) of Bollerslev (1986), in which the conditional variance depends on past values of the squared errors and on past conditional variances. Usually, it is found that a GARCH(1,1) model is sufficient to capture the volatility dynamics. (See equation 21)

2.4 Back-testing

When a risk model has been constructed and presented, it should be statistically evaluated. One important feature in this context is back-testing, which is a quantitative method of determining whether the estimated results of a model are correct given the assumptions of the model.

2.4.1 Back-testing VaR

Kupiec (1995) considered statistical techniques that can be used to quantify the accuracy of the tail values of the distribution of losses. The Kupiec frequency test is the most fundamental test, stating that the actual number of VaR violations, calls it x, is significantly different from the expected number of VaR violations, which is $(1-\alpha)N$, where N is the total number of observations.

A VaR violation is said to occur if the observed loss exceeds our VaR estimate for a given day. The actual number of violations (x) is binomially distributed:

$$\Pr(X = x) = \binom{N}{x} p^x (1-p)^{N-x} \tag{23}$$

Where $p = 1 - \alpha$

The cumulative probabilities are then:

$$\Pr(X \leq x) = \sum_{i=0}^{x} \binom{N}{i} p^i (1-p)^{N-i} \tag{24}$$

If the actual frequency of violations deviates too much from the predicted frequency of violations, the model underlying the VaR estimator is statistically rejected.

To apply the exact Kupiec test based on the binomial distribution, we proceed as follows:

1. Calculate the expected number of VaR_αviolations under the assumption of a correct VaR-model, which is $(1 - \alpha)N$.
2. Count the actual number of violations, which we call x, and assume that we observe $x \geq (1 - \alpha)N$ violations.

3. Calculate the probability of observing $X \geq x$ violations under the assumption that the underlying model is correct; this is $Pr\ (X \geq x)$ in the above notation.
4. Compare the probability calculated with the statistical significance level of interest; a standard level for statistical tests is 5%. If the probability calculated is less than the significance level of interest, the underlying VaR-model is rejected, otherwise the underlying model is ”accepted” (or, rather, not rejected).

Note that this is a one-sided test, i.e., we test if the actual frequency of violations x is ”too large” compared to the expected frequency of violations $(1 - \alpha)N$, for the underlying model to be accepted. This version of the test is based on the assumption that we found $x \geq (1 - \alpha)N$ in the sample.

If we instead observe fewer than expected violations, i.e., $x \leq (1 - \alpha)N$, we turn the test around and instead test if the actual frequency of violations is “too low” compared to the expected frequency of violations. In this case we reformulate the steps above as:

1. Calculate the expected number of VaR_α violations under the assumption of a correct VaR-model, which is $(1 - \alpha)N$.
2. Count the actual number of violations, which we call x, and assume that we observe $x \leq (1 - \alpha)N$ violations.
3. Calculate the probability of observing $X \leq x$ violations under the assumption that the underlying model is correct; this is $Pr\ (X \leq x)$ in the above notation.
4. Compare the probability calculated with the statistical significance level of interest. If the probability calculated is less than the significance level of interest, the underlying VaR-model is rejected, otherwise the underlying model is “accepted”.

Finally, a two-sided test may be implemented. This involves the construction of a confidence interval for the observed frequency of violations. If the actual number of violation falls outside the confidence interval, the underlying VaR-model is rejected. This test is slightly more complicated, since by definition, the two-sided test has two rejection regions; it rejects the underlying VaR-model if we observe ”too few” violations or if we observe ”too many” violations. We could use the above formula for the cumulative binomial probability to find the lower bound x_{low} and the upper bound x_{high} for the number of violations (by trial and error). These bounds of the confidence interval are defined by the requirement that there should be an equal probability of observing fewer than x_{low} violations and observing more than x_{high} violations (for example 2.5% probabilities in each tail, at the standard statistical level of 5%).

2.4.2 Back-testing ES

ES measures something else than VaR; as we know it measures the average loss beyond VaR or the expected loss beyond VaR. The area of Back-testing ES is both empirically and theoretically much less developed than the area of Back-testing VaR, and has only very recently started to attract more attention from researchers. Presumably this increased interest is triggered by the proposed changes in the Basel regulation; the so called Fundamental Review of the Trading Book (FRTB) suggests a switch from VaR to ES for measuring market risk. Acerbi and Szekely (2015) proposed three back-testing methods for expected shortfall. Amongst these three, the second method—test statistic Z_2, which is referred to as "Testing ES Directly", is considered to be the most applicable one because the other two tests require Monte Carlo simulation of the distribution of the test statistic to compute the $p - value$ and therefore need to store predictive distributions. Moreover, based on fixed significance thresholds: Z_2 can be treated as a traffic-light system, in which it shows a remarkable stability of the significance thresholds across a wide range of tail index values, which spans over all financially realistic cases. Besides, calculating Z_2 only requires recording two numbers per day: one is the estimated $ES_{\alpha,t}$ and the other one is the magnitude $L_t I_t$ of an $VaR_{\alpha,t}$ exception, where L_t is the loss at time t and I_t is an indicator variable:

$$I_t = \begin{cases} 1, & when\ L_t > VaR_{\alpha,t} \\ 0, & otherwise \end{cases} \qquad (25)$$

In reality, it is sufficient to only record the size of the α−tail of the model distribution, since $L_t I_t$ can be simulated because of $I_t \sim Bernoulli$.

Similar to the previous notations, the independent profit loss is denoted as L_t, the true but unknown distribution of profit losses is F_t, and the model distribution is represented by P_t. Assume that the losses follow a continuous distribution. Then, the definition of ES can be written as:

$$ES_{\alpha,t} = E\left[L_t | L_t > VaR_{\alpha,t}\right] \qquad (26)$$

where α is the confidence level and the Basel governed $\alpha = 0{,}975$.

We may rewrite the formula assuming that the model is correct:

$$ES_{\alpha,t} = E\left[\frac{L_t l_t}{1-\alpha}\right] \qquad (27)$$

Now, define the test statistic:

$$Z_2(L) = -\sum_{t=1}^{T} \frac{L_t l_t}{T(1-\alpha)ES_{\alpha,t}} + 1 \qquad (28)$$

The null hypothesis states that the model is exact in the tail, while the alternative hypothesis states that ES is underestimated. They are mathematically written as:

$$H_0: p_t^{[\alpha]} = F_t^{[\alpha]}\ for\ all\ t$$

$$H_1: ES_{\alpha,t}^{p} \leq ES_{\alpha,t}^{F}\ for\ all\ t\ and\ < for\ some\ t$$

$$VaR_{\alpha,t}^{p} \leq VaR_{\alpha,t}^{F}\ for\ all\ t \qquad (29)$$

Under the null hypothesis:

$$E_{H_0}[Z_2(L)] = E_{H_0}\left[-\sum_{t=1}^{T} \frac{L_t l_t}{T(1-\alpha)ES_{\alpha,t}} + 1\right]$$

$$= -\frac{1}{T}\sum_{t=1}^{T} E\left[\frac{L_t l_t}{(1-\alpha)}\right]\frac{1}{ES_{\alpha,t}} + 1$$

$$= -\frac{1}{T}\sum_{t=1}^{T} ES_{\alpha,t}\frac{1}{ES_{\alpha,t}} + 1$$

$$= -\frac{1}{T}\sum_{t=1}^{T} 1 + 1 = 0 \qquad (30)$$

Then, we can conclude:

$$E_{H_0}[Z_2] = 0\ \ and\ \ E_{H_1}[Z_2] < 0 \qquad (31)$$

Provided the critical values for different significance levels and different degree of freedom (see Table 1), which was published by Acerbi and Szekely (2015), it is clear that the thresholds deviate significantly from $-0{,}7$ only for dramatically heavy tailed distribution, with v closing to 3. As a result, Z_2 with fixed levels $Z_2 = -0.7$ (when significance level is 5%) and $Z_2 = -1.8$ (when significance level is 0,01%) would perfectly be the traffic-light in all occasions, which implies that Z_2 lends itself to implementations that do not require the recording of the predictive distributions. The ± 1 location shifts across an unrealistically large area for a real loss distribution, which is expected to converge around zero.

Finally, compare the actual value of Z_2 to the critical value. If Z_2 is smaller than the corresponding critical value, then the underlying model is rejected. According to a "Basel type" traffic light system: Z_2 above -0.70 gives green light; Z_2 between -0.70 and -1.80 gives yellow light; Z_2 below -1.80 gives red light.

Table 1*: Acerbi and Szekely's recommendation for potential future Basel regulation*

	5%			0,01%		
	Location (mean)			Location (mean)		
dgf(v)	-1%	0%	1%	-1%	0%	1%
3	-0,78	-0,82	-0,88	-3,9	-4,4	-5,5
5	-0,72	-0,74	-0,78	-1,9	-2	-2,3
10	-0,7	-0,71	-0,74	-1,8	-1,9	-1,9
100	-0,7	-0,7	-0,72	-1,8	-1,8	-1,9
∞ (Normal)	-0,7	-0,7	-0,72	-1,8	-1,8	-1,9

2.5 Previous research

McNeil (1997), Jondeau and Rockinger (1999) and Da Silva and Mendez (2003) reported that during an extreme event, returns do not follow the normal distribution because its empirical distribution has heavier tails. Therefore, classical parametric approach that is based on the assumption of normal distribution is not suitable to estimate VaR during an extreme event like major financial crisis. Bollerslev (1986) proposed a conditional heteroscedasticity model, the GARCH model. From that on, the dynamic methods have gained much popularity, since they can capture the stylized facts, volatility clustering and leptokurtosis, within the financial data. However, according to a research conducted by Danielsson and de Vries (2000), financial institutions preferred unconditional models due to their simplicity, although these models are based on false assumption of independence and equal distribution of returns. Gilli and Këllezi

(2006) admitted that the choice between conditional and unconditional model should depend on the period for the analysis and risk measures, of which the researcher or manager wants to use. Echaust and Just (2013) found that conditional models perform better than unconditional models, and, the GARCH-EVT model enables to estimate the VaR correctly regardless of the considered assets.

The article *The Peaks over Thresholds Method for Estimating High Quantiles of Loss Distributions*, by Alexander J. McNeil & Thomas Saladin (1997), reviewed the POT method for modelling tails of loss severity distributions and discussed the use of this technique for estimating high quantiles and the possible relevance of this to excess of loss insurance in high layers. The authors concluded that the POT method is a theoretically well supported technique for fitting a parametric distribution to the tail of an unknown underlying distribution and reading off quantile estimates from the fitted curve.

The article *An Application of Extreme Value Theory for Measuring Financial Risk,* by Manfred Gilli and Evis Kèllezi (2006), applying EVT to compute tail risk measures and the related confidence intervals on six major stock market indices. They concluded that EVT can be useful for assessing the size of extreme events and the POT method proved superior as it better exploits the information in the data sample. Being interested in long term behavior rather than in short term forecasting, they favored an unconditional approach.

The paper *Extreme Value at Risk and Expected Shortfall during Financial Crisis*, by Lanciné Kourouma, Denis Dupre, Gilles Sanfilippo and Ollivier Taramasco (2011), investigated VaR and ES based on EVT and historical simulation (HS) approach for CAC 40, S&P 500, Wheat and Crude Oil indexes during the 2008 financial crisis. The authors concluded an underestimation of the risk of loss for the unconditional VaR models as compared with the conditional models. The underestimation is stronger using the historical VaR approach than using the extreme values theory VaR model for both normal periods and 2008 financial crisis period.

The study *Extreme Value Theory and Peaks Over Threshold Model in the Russian Stock Market,* by Vladimir O. Andreev, Sergey E. Tinykov, Oksana P. Ovchinnikova and Gennady P. Parahin (2012), applied EVT and POT techniques to a series of daily losses of the RTS index (RTSI) over a 15-year period (1995-2009). They also concluded that utilized POT model of EVT, and GPD distribution which can give more accurate description on tail distribution of financial returns or losses.

The article *Extreme Value Theory for Time Series using Peak-Over-Threshold method,* by Gianluca Rosso (2015), summarized the chances offered by the Peak-Over-Threshold method, related with analysis of extremes. Identification of appropriate Value at Risk can be solved by fitting data with a Generalized Pareto Distribution. Also an estimation of value for the Expected Shortfall can be useful, and the application of these few concepts is valid for the widest range of risk analysis.

3 Methodology

3.1 Data

Two sets of data, daily stock price of Ford Motor Company (Ford) and index of the Standard & Poor's 500 (the S&P 500) in USD, are collected from Yahoo Finance, covers from January 1st 2002 to December 31st 2011. The S&P 500 is an American stock market index based on the market capitalizations of 500 large companies having common stock listed on the NYSE or NASDAQ. The S&P 500 index components and their weightings are determined by S&P Dow Jones Indices. It differs from other U.S. stock market indices, such as the Dow Jones Industrial Average or the Nasdaq Composite index, because of its diverse constituency and weighting methodology. It is one of the most commonly followed equity indices, and many consider it one of the best representations of the U.S. stock market, and a bellwether for the U.S. economy. The National Bureau of Economic Research has classified common stocks as a leading indicator of business cycles. Ford is an American multinational automotive company and one of the S&P 500 components.

3.2 Method

The losses are calculated as 100 times the daily log-return of each series. Five-year estimation windows are sampled for the next one-year VaR and ES forecast using POT model and evaluation using Kupiec test and Acebi & Szekely test at 5% critical value. Final, a total five-year overall back-testing was performed to test whether the behavior of the models differs. The GARCH(1,1) model incorporates mean reversion whereas the EWMA model does not. GARCH(1,1) is, therefore, theoretically more appealing than the EWMA model and being used in the paper. In POT models, there are four scenarios set for the threshold u (1%, 2%, 3% and 4%), three scenarios set for the probability α (95%, 99% and 99.9%) and one critical value set for the back-testing at 5%. Maximum Likelihood (ML) method is used to obtain the estimates of parameters ξ and β for both $\xi = 0$ and $\xi > 0$ situations. Parameters are estimated “before" every new evaluation period starts. Considering the fact that EVT only focus on the large losses and these large losses are essentially the same before and after the estimation window is moved one day or one month forward, the estimates of EVT parameters would be almost the same, and consequently there is no reason to update parameters too often. EViews has been used for GARCH(1,1) parameters estimation, Solver in Excel has been used in ML method to estimate β, ξ and verify GARCH(1,1) parameters’ estimation.

To give a brief summary of the methodology as follows (repeating the following process for each threshold u - 1%, 2%, 3% and 4%)

- Unconditional POT
  1. Five-year daily log returns are calculated at percentage level
  2. Count the losses that over the threshold u
  3. Estimate the parameters ξ and β for both $\xi = 0$ and $\xi > 0$ situations using Excel Solver function
  4. Compute the VaR at 95%, 99% and 99.9% confidence level
  5. Back-testing the VaR by both Kupiec one-sided and two-sided tests at 5% critical value
  6. Compute the ES at 95%, 99% and 99.9% confidence level
  7. Back-testing the ES by Acebi & Szekely test at 5% critical value
  8. When done the above process for each year, then total the five years violations and do the back-testing for five-year-overall for both VaR and ES at 5% critical value

- Conditional POT
  1. Five-year daily log returns are calculated at percentage level
  2. Estimate the parameters of GARCH(1,1) model in EView and check in Excel using Solver function
  3. Compute the residual value for each t by extracting the five-year mean of the daily log returns and standardized the residual for each t by GARCH(1,1) model
  4. Count the standardized residuals that over the threshold u
  5. Estimate the parameters ξ and β for both $\xi = 0$ and $\xi > 0$ situations using Excel Solver function
  6. Calculate the quantile estimates at 95%, 99% and 99.9% confidence level

7. Compute the conditional daily VaR at 95%, 99% and 99.9% confidence levels
8. Back-testing the VaR by both Kupiec one-sided and two-sided tests at 5% critical value
9. Compute the conditional daily ES at 95%, 99% and 99.9% confidence levels
10. Back-testing the ES by Acebi & Szekely test at 5% critical value
11. When done the above process for each year, then total the five years violations and do the back-testing for five-year-overall for both VaR and ES at 5% critical value

4 Empirical Results

In this chapter, the results are presented in the order of data analysis, back-testing results of Ford and of the S&P 500 separately. Furthermore, a comparison amongst different scenarios for each data stream is analyzed.

4.1 Data Analysis

Figure 1 & 2 present the loss observations from 2002 to 2011 for both datasets during the entire sample period. Huge fluctuations are both noticeable from mid-2008 to 2009, during which the financial crisis period was, such jumps increase the market risk of the portfolios. However, the S&P 500 is a diverse and weighted index, its losses obviously surge less than Ford's. Such lower ups and downs impact the choices of thresholds for POT models. Thresholds 1%, 2%, 3% and 4% are chosen for Ford analysis, only 1% and 2% are chosen for the S&P 500 due to higher thresholds will cause zero observation over threshold under conditional POT model for most of the years in-sample. (Note: here the threshold is a fixed number used to compare with the losses or the standardized residuals, rather than a percentage of the total observations.) Otherwise, POT model could not be processed. However, there is one zero violation (N_u) observed in Ford 2010 when threshold equal to 4%. In order to have a 5-year overall picture, zero violations for final VaR back-testing and $Z_2 = 1$ for ES back-testing are manually set for year 2010.

Figure 1*: Ford daily losses during 2002 to 2011*

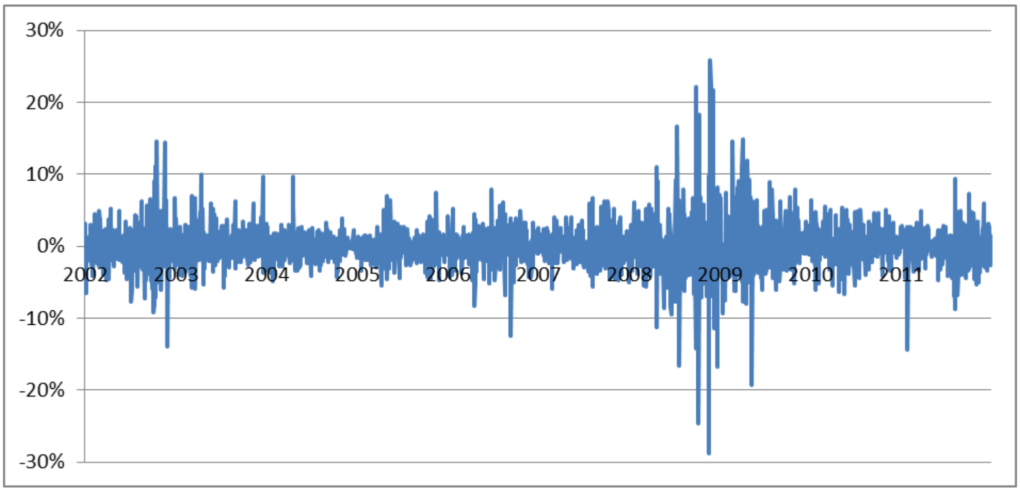

Figure 2*: S&P 500 daily losses during 2002 to 2011*

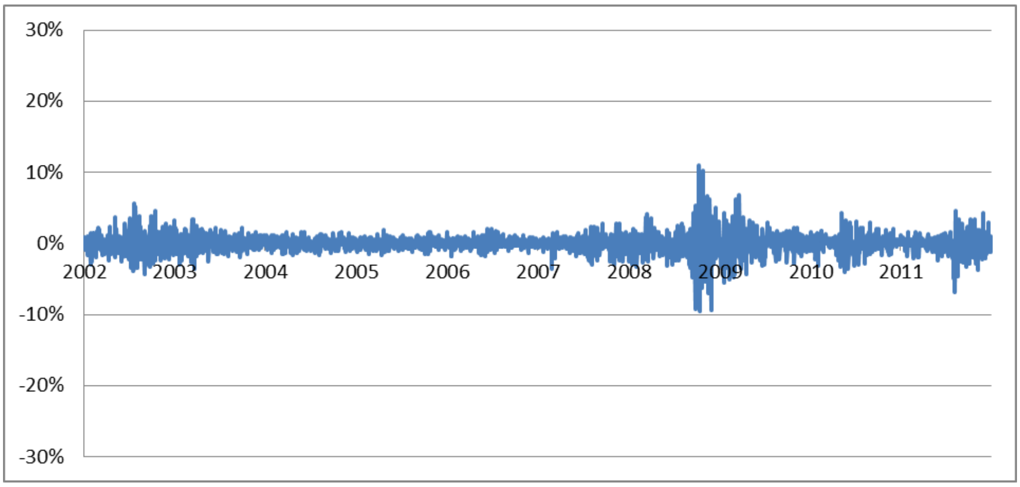

Table 2 & 3 present the summary statistics for both daily losses over the entire sample period. These tables include five testing periods – 5-year-window for each, and the whole five-year evaluation period. Each distribution includes 2519 observations during the entire sample period. Statistics results show the same shape with line chart illustrated in Figure 1 & 2. Parameters β and ξ are estimated once for each testing periods under different thresholds.

Table 2: *Ford daily loss observations descriptive statistics*

Ford	*Testing Periods*					*Evaluation Period*
	2002-2006	2003-2007	2004-2008	2005-2009	2006-2010	2007-2011
Mean	-0,0587	-0,0257	-0,1544	-0,0303	0,0617	0,0285
Standard Error	0,0678	0,0600	0,0934	0,1055	0,1073	0,1078
Median	-0,1226	-0,1169	-0,1404	-0,1226	0,0000	0,0000
Standard Deviation	2,4051	2,1295	3,3150	3,7420	3,8055	3,8280
Sample Variance	5,7844	4,5349	10,9892	14,0024	14,4819	14,6536
Kurtosis	4,7546	2,5883	18,4578	11,9225	10,9992	10,7677
Skewness	0,3504	0,2650	-0,1351	-0,0113	-0,0772	-0,0681
Range	28,4529	22,5434	54,6332	54,6332	54,6332	54,6332
Minimum	-13,9413	-12,5236	-28,7682	-28,7682	-28,7682	-28,7682
Maximum	14,5117	10,0198	25,8650	25,8650	25,8650	25,8650
Sum	-73,8698	-32,3439	-194,4037	-38,1172	77,6969	35,9600
Observations	1259	1258	1259	1259	1259	1260

Table 3*: S&P 500 daily loss observations descriptive statistics*

S&P500	*Testing Periods*					*Evaluation Period*
	2002-2006	2003-2007	2004-2008	2005-2009	2006-2010	2007-2011
Mean	0,0168	0,0407	-0,0165	-0,0066	0,0006	-0,0095
Standard Error	0,0286	0,0235	0,0379	0,0427	0,0443	0,0474
Median	0,0564	0,0813	0,0699	0,0817	0,0842	0,0840
Standard Deviation	1,0141	0,8332	1,3447	1,5168	1,5735	1,6810
Sample Variance	1,0284	0,6942	1,8082	2,3008	2,4758	2,8257
Kurtosis	3,1137	1,7745	15,7336	10,1431	8,6064	6,5646
Skewness	0,2385	-0,1350	-0,3581	-0,2379	-0,2330	-0,2454
Range	9,8168	7,0681	20,4267	20,4267	20,4267	20,4267
Minimum	-4,2423	-3,5867	-9,4695	-9,4695	-9,4695	-9,4695
Maximum	5,5744	3,4814	10,9572	10,9572	10,9572	10,9572
Sum	21,1368	51,2184	-20,7844	-8,3262	0,7462	-12,0254
Observations	1259	1258	1259	1259	1259	1260

Different threshold does not impact on GARCH(1,1) method, thus Appendix A present the EView estimates for 5 testing periods of Ford and Appendix B present for the S&P 500. With all GARCH items' parameters close to 0,94, which is usually the number of λ used to captured the volatility under EWMA method, this study was thus continued with GARCH(1,1) method only. ML method is also used to estimate the parameters for GARCH(1,1) using Solver in Excel, results see Appendix C. There are slightly differences between EViews and Solver on parameters, mostly less than 1%, but no gap on sum maximum. Therefore, EViews results are captured for further calculation.

4.2 Results - Ford

Table 4,5,6,7 show both one-sided and two-sided Kupiec test results for VaR back-testing at 5% critical value under four different threshold scenarios in the order of 1%, 2%, 3% and 4%. The one-sided Kupiec test is used to check if the actual frequency of violations is “too large” or “too low” compared to the expected frequency of violations and the two-sided one focuses on an confidence interval, is used to check if the actual frequency of violations is “too many” or “too few” compared to the expected frequency of violations, so, the obvious difference between two tests are under extreme high confidence level. When we observe zero violations, mostly under $\alpha = 99.9\%$, the models are always rejected by one-sided Kupiec test but accepted by two-sided one. Therefore, the two-sided Kupiec test results are adopted to further discussion as the study focuses on the EVT, especially when there is an extreme high confidence α.

When the threshold $u = 1\%$, all models are accepted in 2007, only unconditional models at 95% confidence level are rejected in 2010 and 2011, and only conditional GARCH(1,1) models are accepted for 2008, 2009 and five-year overall. When the threshold $u = 2\%$, more models are rejected comparing to the results when $u = 1\%$. GARCH(1,1) models at 95% confidence level are rejected in 2008, no matter ξ equals to zero or not. And GARCH(1,1) models at 95% confidence level with $\xi = 0$ is rejected for five-year overall. When increasing the threshold to 3% and 4%, fewer and fewer ‘accept’ observed at 95% confidence level, models’ performances are still worse in 2008 and 2009, during which were the financial crisis period with higher cluttering, than the other years. Moreover, unconditional models are all rejected for five-year overall.

***Table 4:** Kupiec test results under threshold u = 1%*

α		95%				99%				99,9%			
		Unconditional		Conditional		Unconditional		Conditional		Unconditional		Conditional	
		ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0
	Yearly Confidence Interval	*X_low* = 6		*X_high* = 20		*X_low* = 0		*X_high* = 6		*X_low* = 0		*X_high* = 2	
2007	Nr. of Violations	12	12	15	15	1	0	4	3	0	0	0	0
	Prob.	51,17%	51,17%	27,61%	27,61%	0,00%	0,00%	0,44%	0,12%	0,00%	0,00%	0,00%	0,00%
	One-sided Test Result	Accept	Accept	Accept	Accept	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject
	Two-sided Test Result	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept
2008	Nr. of Violations	45	45	20	20	21	19	5	5	8	7	2	0
	Prob.	0,00%	0,00%	3,03%	3,03%	1,68%	5,22%	1,18%	1,18%	11,07%	6,00%	0,02%	0,00%
	One-sided Test Result	Reject	Reject	Reject	Reject	Reject	Accept	Reject	Reject	Accept	Accept	Reject	Reject
	Two-sided Test Result	Reject	Reject	Accept	Accept	Reject	Reject	Accept	Accept	Reject	Reject	Accept	Accept
2009	Nr. of Violations	39	40	17	17	14	10	4	4	3	0	0	0
	Prob.	0,00%	0,00%	13,14%	13,14%	38,19%	28,13%	0,42%	0,42%	0,12%	0,00%	0,00%	0,00%
	One-sided Test Result	Reject	Reject	Accept	Accept	Accept	Accept	Reject	Reject	Reject	Reject	Reject	Reject
	Two-sided Test Result	Reject	Reject	Accept	Accept	Reject	Reject	Accept	Accept	Reject	Accept	Accept	Accept
2010	Nr. of Violations	1	2	16	16	0	0	3	3	0	0	0	0
	Prob.	0,00%	0,02%	19,70%	19,70%	0,00%	0,00%	0,12%	0,12%	0,00%	0,00%	0,00%	0,00%
	One-sided Test Result	Reject	Reject	Accept	Accept	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject
	Two-sided Test Result	Reject	Reject	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept
2011	Nr. of Violations	3	3	6	6	0	0	0	0	0	0	0	0
	Prob.	0,12%	0,12%	2,96%	2,96%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
	One-sided Test Result	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject
	Two-sided Test Result	Reject	Reject	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept
5-Year	Confidence Interval	*X_low* = 48		*X_high* = 79		*X_low* = 6		*X_high* = 20		*X_low* = 0		*X_high* = 4	
Overall	Nr. of Violations	100	102	74	74	36	29	16	15	11	7	2	0
	Prob.	0,00%	0,00%	8,97%	8,97%	0,01%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
	One-sided Test Result	Reject	Reject	Accept	Accept	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject
	Two-sided Test Result	Reject	Reject	Accept	Accept	Reject	Reject	Accept	Accept	Reject	Reject	Accept	Accept

***Table 5:** Kupiec test results under threshold u = 2%*

α		95%				99%				99,9%			
		Unconditional		Conditional		Unconditional		Conditional		Unconditional		Conditional	
		ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0
	Yearly Confidence Interval	*X_low* = 6		*X_high* = 20		*X_low* = 0		*X_high* = 6		*X_low* = 0		*X_high* = 2	
2007	Nr. of Violations	12	12	15	15	1	0	2	2	0	0	0	0
	Prob.	51,17%	51,17%	27,61%	27,61%	0,00%	0,00%	0,03%	0,03%	0,00%	0,00%	0,00%	0,00%
	One-sided Test Result	Accept	Accept	Accept	Accept	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject
	Two-sided Test Result	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept
2008	Nr. of Violations	45	47	22	22	20	19	5	5	8	7	1	1
	Prob.	0,00%	0,00%	0,89%	0,89%	3,03%	5,22%	1,18%	1,18%	11,07%	6,00%	0,00%	0,00%
	One-sided Test Result	Reject	Reject	Reject	Reject	Reject	Accept	Reject	Reject	Accept	Accept	Reject	Reject
	Two-sided Test Result	Reject	Reject	Reject	Reject	Reject	Reject	Accept	Accept	Reject	Reject	Accept	Accept
2009	Nr. of Violations	38	41	19	17	10	10	3	3	2	0	0	0
	Prob.	0,00%	0,00%	5,05%	13,14%	28,13%	28,13%	0,12%	0,12%	0,02%	0,00%	0,00%	0,00%
	One-sided Test Result	Reject	Reject	Accept	Accept	Accept	Accept	Reject	Reject	Reject	Reject	Reject	Reject
	Two-sided Test Result	Reject	Reject	Accept	Accept	Reject	Reject	Accept	Accept	Accept	Accept	Accept	Accept
2010	Nr. of Violations	1	2	17	17	0	0	3	3	0	0	0	0
	Prob.	0,00%	0,02%	13,14%	13,14%	0,00%	0,00%	0,12%	0,12%	0,00%	0,00%	0,00%	0,00%
	One-sided Test Result	Reject	Reject	Accept	Accept	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject
	Two-sided Test Result	Reject	Reject	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept
2011	Nr. of Violations	3	3	7	7	0	0	2	2	0	0	0	0
	Prob.	0,12%	0,12%	6,16%	6,16%	0,00%	0,00%	0,02%	0,02%	0,00%	0,00%	0,00%	0,00%
	One-sided Test Result	Reject	Reject	Accept	Accept	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject
	Two-sided Test Result	Reject	Reject	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept
5-Year	Confidence Interval	*X_low* = 48		*X_high* = 79		*X_low* = 6		*X_high* = 20		*X_low* = 0		*X_high* = 4	
Overall	Nr. of Violations	99	105	80	78	31	29	15	15	10	7	1	1
	Prob.	0,00%	0,00%	1,91%	3,36%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
	One-sided Test Result	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject
	Two-sided Test Result	Reject	Reject	Reject	Accept	Reject	Reject	Accept	Accept	Reject	Reject	Accept	Accept

Table 6: *Kupiec test results under threshold u = 3%*

α		95%				99%				99,9%			
		Unconditional		Conditional		Unconditional		Conditional		Unconditional		Conditional	
		ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0
	Yearly Confidence Interval	*X_low* = 6		*X_high* = 20		*X_low* = 0		*X_high* = 6		*X_low* = 0		*X_high* = 2	
2007	Nr. of Violations	12	12	32	32	0	0	3	3	0	0	0	0
	Prob.	51,17%	51,17%	0,00%	0,00%	0,00%	0,00%	0,12%	0,12%	0,00%	0,00%	0,00%	0,00%
	One-sided Test Result	Accept	Accept	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject
	Two-sided Test Result	Accept	Accept	Reject	Reject	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept
2008	Nr. of Violations	47	47	25	25	18	18	4	4	7	7	0	0
	Prob.	0,00%	0,00%	0,10%	0,10%	8,59%	8,59%	0,41%	0,41%	6,00%	6,00%	0,00%	0,00%
	One-sided Test Result	Reject	Reject	Reject	Reject	Accept	Accept	Reject	Reject	Accept	Accept	Reject	Reject
	Two-sided Test Result	Reject	Reject	Reject	Reject	Reject	Reject	Accept	Accept	Reject	Reject	Accept	Accept
2009	Nr. of Violations	38	42	25	25	9	10	4	4	0	0	0	0
	Prob.	0,00%	0,00%	0,10%	0,10%	18,70%	28,13%	0,42%	0,42%	0,00%	0,00%	0,00%	0,00%
	One-sided Test Result	Reject	Reject	Reject	Reject	Accept	Accept	Reject	Reject	Reject	Reject	Reject	Reject
	Two-sided Test Result	Reject	Reject	Reject	Reject	Reject	Reject	Accept	Accept	Accept	Accept	Accept	Accept
2010	Nr. of Violations	1	2	11	11	0	0	3	3	0	0	0	0
	Prob.	0,00%	0,02%	39,05%	39,05%	0,00%	0,00%	0,12%	0,12%	0,00%	0,00%	0,00%	0,00%
	One-sided Test Result	Reject	Reject	Accept	Accept	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject
	Two-sided Test Result	Reject	Reject	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept
2011	Nr. of Violations	3	3	4	4	0	0	0	0	0	0	0	0
	Prob.	0,12%	0,12%	0,42%	0,42%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
	One-sided Test Result	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject
	Two-sided Test Result	Reject	Reject	Reject	Reject	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept
5-Year	Confidence Interval	*X_low* = 48		*X_high* = 79		*X_low* = 6		*X_high* = 20		*X_low* = 0		*X_high* = 4	
Overall	Nr. of Violations	101	106	97	97	27	28	14	14	7	7	0	0
	Prob.	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
	One-sided Test Result	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject
	Two-sided Test Result	Reject	Reject	Reject	Reject	Reject	Reject	Accept	Accept	Reject	Reject	Accept	Accept

Table 7: *Kupiec test results under threshold u = 4%*

α		95%				99%				99,9%			
		Unconditional		Conditional		Unconditional		Conditional		Unconditional		Conditional	
		ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0
	Yearly Confidence Interval	*X_low* = 6		*X_high* = 20		*X_low* = 0		*X_high* = 6		*X_low* = 0		*X_high* = 2	
2007	Nr. of Violations	13	13	125	125	0	0	10	10	0	0	0	0
	Prob.	48,83%	48,83%	0,00%	0,00%	0,00%	0,00%	28,61%	28,61%	0,00%	0,00%	0,00%	0,00%
	One-sided Test Result	Accept	Accept	Reject	Reject	Reject	Reject	Accept	Accept	Reject	Reject	Reject	Reject
	Two-sided Test Result	Accept	Accept	Reject	Reject	Accept	Accept	Reject	Reject	Accept	Accept	Accept	Accept
2008	Nr. of Violations	49	49	22	22	18	18	4	4	7	7	0	0
	Prob.	0,00%	0,00%	0,89%	0,89%	8,59%	8,59%	0,41%	0,41%	6,00%	6,00%	0,00%	0,00%
	One-sided Test Result	Reject	Reject	Reject	Reject	Accept	Accept	Reject	Reject	Accept	Accept	Reject	Reject
	Two-sided Test Result	Reject	Reject	Reject	Reject	Reject	Reject	Accept	Accept	Reject	Reject	Accept	Accept
2009	Nr. of Violations	41	42	55	55	6	10	5	5	0	0	0	0
	Prob.	0,00%	0,00%	0,00%	0,00%	2,96%	28,13%	1,22%	1,22%	0,00%	0,00%	0,00%	0,00%
	One-sided Test Result	Reject	Reject	Reject	Reject	Reject	Accept	Reject	Reject	Reject	Reject	Reject	Reject
	Two-sided Test Result	Reject	Reject	Reject	Reject	Accept	Reject	Accept	Accept	Accept	Accept	Accept	Accept
2010	Nr. of Violations	1	2	0	0	0	0	0	0	0	0	0	0
	Prob.	0,00%	0,02%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
	One-sided Test Result	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject
	Two-sided Test Result	Reject	Reject	Reject	Reject	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept
2011	Nr. of Violations	3	3	0	0	0	0	0	0	0	0	0	0
	Prob.	0,12%	0,12%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
	One-sided Test Result	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject
	Two-sided Test Result	Reject	Reject	Reject	Reject	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept
5-Year	Confidence Interval	*X_low* = 48		*X_high* = 79		*X_low* = 6		*X_high* = 20		*X_low* = 0		*X_high* = 4	
Overall	Nr. of Violations	107	109	202	202	24	28	19	19	7	7	0	0
	Prob.	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
	One-sided Test Result	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject
	Two-sided Test Result	Reject	Reject	Reject	Reject	Reject	Reject	Accept	Accept	Reject	Reject	Accept	Accept

Table 8,9,10,11 illustrate the Acerbi and Szekely's ES back-testing results and traffic light results at 5% and 0,01% critical level under four different threshold scenarios in the same order of 1%, 2%, 3% and 4%. According to ES back-testing formula (28), when $-\sum_{t=1}^{T}\frac{L_t l_t}{T(1-\alpha)ES_{\alpha,t}}$ equals to zero, Z_2 statistic test results will exactly equal to one. Only Green in the traffic light results will be considered as accept, thus the traffic light results will be discussed. For year 2010 and 2011, all models are Green, whatever assumptions are. When the threshold $u = 1\%\ and\ 2\%,$ all models are Green for year 2007. For 2008, all unconditional models are 'Red'. One conditional model with $\xi > 0$ at 99.9% confidence level turns 'Green', the other one with $\xi = 0$ is 'Red', the rest conditional model are all 'Yellow', better than the unconditional ones. For 2009, most conditional models are 'Red' except the one at 99,9% confidence level, conditional models are either 'Yellow' or 'Green', still better than the unconditional ones. For five-years-overall, all conditional models are 'Green' while all unconditional ones are either 'Yellow' or 'Red'. When increasing the threshold $u = 3\%$, for 2007, conditional models at 95% confidence level turns to 'Yellow' from 'Green'. For 2008, conditional models at 99,9% confidence level turns to 'Green' from 'Red'. For 2009, conditional model at 95% confidence level with $\xi > 0$ turns to 'Yellow' from 'Green' while unconditional model at 99,9% confidence level with $\xi = 0$ turns to 'Green' from 'Red'. Overall, conditional models at 95% confidence level turn to 'Yellow' from 'Green'. Continue to increase the threshold $u = 4\%$, for 2007, conditional models at both 95% and 99% change to 'Red'. No changes happened in 2008. For 2009, only conditional models at 99,9% confidence level are still 'Green'. Five-years-overall, only conditional models at 99% and 99.9% levels are remained as 'Green'. Under ES back-testing, when increasing the threshold, the 'Accept' level is decreased, but the models are still doing well at 99% and 99.9% GARCH methods.

***Table 8**: ES Back-testing results under threshold u = 1%*

	α	95%				99%				99,9%			
		Unconditional		Conditional		Unconditional		Conditional		Unconditional		Conditional	
		ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0
2007	Z (test statistic)	0,117	0,157	-0,372	-0,285	0,675	1,000	-0,534	-0,016	1,000	1,000	1,000	1,000
	Test Result	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept
	Traffic light Result	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
2008	Z (test statistic)	-4,671	-4,564	-1,162	-0,973	-12,073	-10,839	-1,671	-1,279	-53,808	-45,430	-7,238	1,000
	Test Result	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Accept
	Traffic light Result	Red	Red	Yellow	Yellow	Red	Red	Yellow	Yellow	Red	Red	Red	Green
2009	Z (test statistic)	-2,579	-2,177	-0,609	-0,522	-4,889	-2,199	-0,581	-0,438	-10,555	1,000	1,000	1,000
	Test Result	Reject	Reject	Accept	Accept	Reject	Reject	Accept	Accept	Reject	Accept	Accept	Accept
	Traffic light Result	Red	Red	Green	Green	Red	Red	Green	Green	Red	Green	Green	Green
2010	Z (test statistic)	0,938	0,896	-0,406	-0,406	1,000	1,000	-0,175	-0,175	1,000	1,000	1,000	1,000
	Test Result	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept
	Traffic light Result	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
2011	Z (test statistic)	0,784	0,801	0,448	0,448	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
	Test Result	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept
	Traffic light Result	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
5-Year Overall	Z (test statistic)	-1,086	-0,981	-0,421	-0,348	-2,868	-2,017	-0,393	-0,182	-12,316	-8,323	-0,654	1,000
	Test Result	Reject	Reject	Accept	Accept	Reject	Reject	Accept	Accept	Reject	Reject	Accept	Accept
	Traffic light Result	Yellow	Yellow	Green	Green	Red	Red	Green	Green	Red	Red	Green	Green

***Table 9**: ES Back-testing results under threshold u = 2%*

	α	95%				99%				99,9%			
		Unconditional		Conditional		Unconditional		Conditional		Unconditional		Conditional	
		ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0
2007	Z (test statistic)	0,125	0,158	-0,371	-0,301	0,679	1,000	0,257	0,287	1,000	1,000	1,000	1,000
	Test Result	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept
	Traffic light Result	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
2008	Z (test statistic)	-4,642	-4,670	-1,409	-1,409	-11,641	-10,562	-1,519	-1,519	-53,179	-42,270	-2,854	-2,854
	Test Result	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject
	Traffic light Result	Red	Red	Yellow	Yellow	Red	Red	Yellow	Yellow	Red	Red	Red	Red
2009	Z (test statistic)	-2,271	-2,195	-0,791	-0,553	-3,117	-2,085	-0,138	-0,097	-6,170	1,000	1,000	1,000
	Test Result	Reject	Reject	Reject	Accept	Reject	Reject	Accept	Accept	Reject	Accept	Accept	Accept
	Traffic light Result	Red	Red	Yellow	Green	Red	Red	Green	Green	Red	Green	Green	Green
2010	Z (test statistic)	0,941	0,895	-0,492	-0,492	1,000	1,000	-0,163	-0,163	1,000	1,000	1,000	1,000
	Test Result	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept
	Traffic light Result	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
2011	Z (test statistic)	0,788	0,801	0,363	0,363	1,000	1,000	0,263	0,263	1,000	1,000	1,000	1,000
	Test Result	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept
	Traffic light Result	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
5-Year Overall	Z (test statistic)	-1,015	-1,006	-0,541	-0,479	-2,425	-1,939	-0,261	-0,247	-11,313	-7,688	0,226	0,226
	Test Result	Reject	Reject	Accept	Accept	Reject	Reject	Accept	Accept	Reject	Reject	Accept	Accept
	Traffic light Result	Yellow	Yellow	Green	Green	Red	Red	Green	Green	Red	Red	Green	Green

***Table 10**: ES Back-testing results under threshold u = 3%*

	α	95%				99%				99,9%			
		Unconditional		Conditional		Unconditional		Conditional		Unconditional		Conditional	
		ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0
2007	Z (test statistic)	0,150	0,158	-1,719	-1,719	1,000	1,000	-0,077	-0,077	1,000	1,000	1,000	1,000
	Test Result	Accept	Accept	Reject	Reject	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept
	Traffic light Result	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green	Green	Green	Green
2008	Z (test statistic)	-4,613	-4,613	-1,650	-1,650	-10,253	-10,253	-1,149	-1,149	-45,255	-45,255	1,000	1,000
	Test Result	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Accept	Accept
	Traffic light Result	Red	Red	Yellow	Yellow	Red	Red	Yellow	Yellow	Red	Red	Green	Green
2009	Z (test statistic)	-2,122	-2,216	-1,627	-1,627	-2,459	-1,963	-0,549	-0,549	1,000	1,000	1,000	1,000
	Test Result	Reject	Reject	Reject	Reject	Reject	Reject	Accept	Accept	Accept	Accept	Accept	Accept
	Traffic light Result	Red	Red	Yellow	Yellow	Red	Red	Green	Green	Green	Green	Green	Green
2010	Z (test statistic)	0,942	0,895	-0,009	-0,009	1,000	1,000	-0,192	-0,192	1,000	1,000	1,000	1,000
	Test Result	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept
	Traffic light Result	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
2011	Z (test statistic)	0,792	0,801	0,616	0,616	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
	Test Result	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept
	Traffic light Result	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
5-Year Overall	Z (test statistic)	-0,974	-0,999	-0,878	-0,878	-1,951	-1,852	-0,194	-0,194	-8,288	-8,288	1,000	1,000
	Test Result	Reject	Reject	Reject	Reject	Reject	Reject	Accept	Accept	Reject	Reject	Accept	Accept
	Traffic light Result	Yellow	Yellow	Yellow	Yellow	Red	Red	Green	Green	Red	Red	Green	Green

Table 11*: ES Back-testing results under threshold u = 4%*

α		95%				99%				99,9%			
		Unconditional		Conditional		Unconditional		Conditional		Unconditional		Conditional	
		ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0
2007	Z (test statistic)	0,106	0,107	-15,382	-15,385	1,000	1,000	-3,077	-3,077	1,000	1,000	1,000	1,000
	Test Result	Accept	Accept	Reject	Reject	Accept	Accept	Reject	Reject	Accept	Accept	Accept	Accept
	Traffic light Result	Green	Green	Red	Red	Green	Green	Red	Red	Green	Green	Green	Green
2008	Z (test statistic)	-4,758	-4,758	-1,406	-1,406	-10,135	-10,135	-1,139	-1,139	-44,273	-44,273	1,000	1,000
	Test Result	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Accept	Accept
	Traffic light Result	Red	Red	Yellow	Yellow	Red	Red	Yellow	Yellow	Red	Red	Green	Green
2009	Z (test statistic)	-2,197	-2,218	-5,125	-5,125	-1,331	-1,967	-1,146	-1,146	1,000	1,000	1,000	1,000
	Test Result	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Accept	Accept	Accept	Accept
	Traffic light Result	Red	Red	Red	Red	Yellow	Red	Yellow	Yellow	Green	Green	Green	Green
2010	Z (test statistic)	0,943	0,895	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
	Test Result	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept
	Traffic light Result	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
2011	Z (test statistic)	0,796	0,801	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
	Test Result	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept
	Traffic light Result	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
5-Year Overall	Z (test statistic)	-1,026	-1,038	-3,972	-3,972	-1,702	-1,829	-0,671	-0,671	-8,091	-8,091	1,000	1,000
	Test Result	Reject	Reject	Reject	Reject	Reject	Reject	Accept	Accept	Reject	Reject	Accept	Accept
	Traffic light Result	Yellow	Yellow	Red	Red	Yellow	Red	Green	Green	Red	Red	Green	Green

4.3 Results - the S&P 500

Table 12 and 13 show the Kupiec back-testing results for the S&P 500 under 1% and 2% threshold. There have been the same results for both thresholds under two-sided Kupiec test. All models are accepted for 2007, 2010 and 2011, only unconditional models are rejected at 95% confidence level and year 2008 has the same result as five-year-overal with all conditional models accept but unconditional ones reject.

Table 12: *the S&P500 Kupiec test results under threshold u = 1%*

α		95%				99%				99,9%			
		Unconditional		Conditional		Unconditional		Conditional		Unconditional		Conditional	
		ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0
	Yearly Confidence Interval	X_low = 6		X_high = 20		X_low = 0		X_high = 6		X_low = 0		X_high = 2	
2007	Nr. of Violations	10	10	10	10	1	0	0	0	0	0	0	0
	Prob.	28,61%	28,61%	28,61%	28,61%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
	One-sided Test Result	Accept	Accept	Accept	Accept	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject
	Two-sided Test Result	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept
2008	Nr. of Violations	47	47	19	19	29	29	5	5	19	19	0	0
	Prob.	0,00%	0,00%	5,22%	5,22%	0,00%	0,00%	1,18%	1,18%	5,22%	5,22%	0,00%	0,00%
	One-sided Test Result	Reject	Reject	Accept	Accept	Reject	Reject	Reject	Reject	Accept	Accept	Reject	Reject
	Two-sided Test Result	Reject	Reject	Accept	Accept	Reject	Reject	Accept	Accept	Reject	Reject	Accept	Accept
2009	Nr. of Violations	29	35	18	18	6	6	2	2	2	0	0	0
	Prob.	0,00%	0,00%	8,35%	8,35%	2,96%	2,96%	0,02%	0,02%	0,02%	0,00%	0,00%	0,00%
	One-sided Test Result	Reject	Reject	Accept	Accept	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject
	Two-sided Test Result	Reject	Reject	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept
2010	Nr. of Violations	8	12	18	18	1	1	3	3	0	0	0	0
	Prob.	11,33%	50,59%	8,35%	8,35%	0,00%	0,00%	0,12%	0,12%	0,00%	0,00%	0,00%	0,00%
	One-sided Test Result	Accept	Accept	Accept	Accept	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject
	Two-sided Test Result	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept
2011	Nr. of Violations	10	12	11	11	3	2	1	1	0	0	0	0
	Prob.	28,13%	50,59%	39,05%	39,05%	0,12%	0,02%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
	One-sided Test Result	Accept	Accept	Accept	Accept	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject
	Two-sided Test Result	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept
5-Year	Confidence Interval	X_low = 48		X_high = 79		X_low = 6		X_high = 20		X_low = 0		X_high = 4	
Overall	Nr. of Violations	104	116	76	76	40	38	11	11	21	19	0	0
	Prob.	0,00%	0,00%	5,62%	5,62%	0,10%	0,04%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
	One-sided Test Result	Reject	Reject	Accept	Accept	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject
	Two-sided Test Result	Reject	Reject	Accept	Accept	Reject	Reject	Accept	Accept	Reject	Reject	Accept	Accept

Table 13: *the S&P500 Kupiec test results under threshold u = 2%*

α		95%				99%				99,9%			
		Unconditional		Conditional		Unconditional		Conditional		Unconditional		Conditional	
		ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0
	Yearly Confidence Interval	X_low = 6		X_high = 20		X_low = 0		X_high = 6		X_low = 0		X_high = 2	
2007	Nr. of Violations	15	15	6	6	0	0	3	3	0	0	0	0
	Prob.	27,61%	27,61%	3,05%	3,05%	0,00%	0,00%	0,12%	0,12%	0,00%	0,00%	0,00%	0,00%
	One-sided Test Result	Accept	Accept	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject
	Two-sided Test Result	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept
2008	Nr. of Violations	50	49	17	17	29	29	6	6	19	19	1	1
	Prob.	0,00%	0,00%	13,47%	13,47%	0,00%	0,00%	2,87%	2,87%	5,22%	5,22%	0,00%	0,00%
	One-sided Test Result	Reject	Reject	Accept	Accept	Reject	Reject	Reject	Reject	Accept	Accept	Reject	Reject
	Two-sided Test Result	Reject	Reject	Accept	Accept	Reject	Reject	Accept	Accept	Reject	Reject	Accept	Accept
2009	Nr. of Violations	52	46	10	10	4	5	2	2	0	0	0	0
	Prob.	0,00%	0,00%	28,13%	28,13%	0,42%	1,22%	0,02%	0,02%	0,00%	0,00%	0,00%	0,00%
	One-sided Test Result	Reject	Reject	Accept	Accept	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject
	Two-sided Test Result	Reject	Reject	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept
2010	Nr. of Violations	11	12	12	12	0	0	4	4	0	0	0	0
	Prob.	39,05%	50,59%	50,59%	50,59%	0,00%	0,00%	0,42%	0,42%	0,00%	0,00%	0,00%	0,00%
	One-sided Test Result	Accept	Accept	Accept	Accept	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject
	Two-sided Test Result	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept
2011	Nr. of Violations	11	11	11	11	2	2	1	1	0	0	0	0
	Prob.	39,05%	39,05%	39,05%	39,05%	0,02%	0,02%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
	One-sided Test Result	Accept	Accept	Accept	Accept	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject
	Two-sided Test Result	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept
5-Year	Confidence Interval	X_low = 48		X_high = 79		X_low = 6		X_high = 20		X_low = 0		X_high = 4	
Overall	Nr. of Violations	139	133	56	56	35	36	16	16	19	19	1	1
	Prob.	0,00%	0,00%	20,19%	20,19%	0,01%	0,01%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
	One-sided Test Result	Reject	Reject	Accept	Accept	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject
	Two-sided Test Result	Reject	Reject	Accept	Accept	Reject	Reject	Accept	Accept	Reject	Reject	Accept	Accept

Table 14 and 15 show the ES back-testing results for both thresholds. Like two-sided Kupiec test results, ES back-testing results shows all 'Green' for 2007, 2010 and 2011 and only conditional models are 'Green' for five-year-overal. There is some difference for 2008 and 2009 between two thresholds. For 2008, only conditional models at 99.9% are 'Green' when $u = 1\%$, but only conditional models at 95% confidence level are 'Green' when $u = 2\%$. For 2009, conditional models are all 'Green' for both thresholds, when increasing the threshold from 1% to 2%, ES back-testing has the better results for unconditional models, besides the one at 95% confidence level.

Table 14:*the S&P500 ES Back-testing results under threshold u = 1%*

α		95%				99%				99,9%			
		Unconditional		Conditional		Unconditional		Conditional		Unconditional		Conditional	
		ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0
2007	Z (test statistic)	0,257	0,263	0,268	0,268	0,685	1,000	1,000	1,000	1,000	1,000	1,000	1,000
	Test Result	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept
	Traffic light Result	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
2008	Z (test statistic)	-5,402	-5,401	-0,705	-0,705	-16,682	-16,678	-1,117	-1,117	-96,415	-96,392	1,000	1,000
	Test Result	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Accept	Accept
	Traffic light Result	Red	Red	Yellow	Yellow	Red	Red	Yellow	Yellow	Red	Red	Green	Green
2009	Z (test statistic)	-1,412	-1,513	-0,406	-0,406	-1,535	-0,804	0,184	0,184	-6,459	1,000	1,000	1,000
	Test Result	Reject	Reject	Accept	Accept	Reject	Reject	Accept	Accept	Reject	Accept	Accept	Accept
	Traffic light Result	Yellow	Yellow	Green	Green	Yellow	Yellow	Green	Green	Red	Green	Green	Green
2010	Z (test statistic)	0,442	0,287	-0,410	-0,410	0,671	0,743	-0,152	-0,152	1,000	1,000	1,000	1,000
	Test Result	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept
	Traffic light Result	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
2011	Z (test statistic)	0,196	0,144	0,143	0,143	-0,013	0,442	0,623	0,623	1,000	1,000	1,000	1,000
	Test Result	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept
	Traffic light Result	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
5-Year Overall	Z (test statistic)	-1,188	-1,249	-0,223	-0,223	-3,388	-3,073	0,106	0,106	-20,052	-18,556	1,000	1,000
	Test Result	Reject	Reject	Accept	Accept	Reject	Reject	Accept	Accept	Reject	Reject	Accept	Accept
	Traffic light Result	Yellow	Yellow	Green	Green	Red	Red	Green	Green	Red	Red	Green	Green

Table 15: *ES Back-testing results under threshold u = 2%*

α		95%				99%				99,9%			
		Unconditional		Conditional		Unconditional		Conditional		Unconditional		Conditional	
		ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0	ξ = 0	ξ > 0
2007	Z (test statistic)	-0,010	-0,010	0,494	0,494	1,000	1,000	-0,159	-0,159	1,000	1,000	1,000	1,000
	Test Result	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept
	Traffic light Result	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
2008	Z (test statistic)	-5,647	-5,581	-0,590	-0,590	-16,749	-16,849	-1,662	-1,662	-96,274	-97,301	-3,019	-3,019
	Test Result	Reject	Reject	Accept	Accept	Reject	Reject	Reject	Reject	Reject	Reject	Reject	Reject
	Traffic light Result	Red	Red	Green	Green	Red	Red	Yellow	Yellow	Red	Red	Red	Red
2009	Z (test statistic)	-2,275	-2,039	0,173	0,173	-0,469	-0,682	0,130	0,125	1,000	1,000	1,000	1,000
	Test Result	Reject	Reject	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept
	Traffic light Result	Red	Red	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
2010	Z (test statistic)	0,330	0,285	-0,032	-0,032	1,000	1,000	-0,611	-0,611	1,000	1,000	1,000	1,000
	Test Result	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept
	Traffic light Result	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
2011	Z (test statistic)	0,189	0,192	0,149	0,149	0,376	0,429	0,598	0,598	1,000	1,000	1,000	1,000
	Test Result	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept	Accept
	Traffic light Result	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
5-Year Overall	Z (test statistic)	-1,487	-1,435	0,038	0,038	-2,982	-3,035	-0,342	-0,343	-18,532	-18,738	0,193	0,193
	Test Result	Reject	Reject	Accept	Accept	Reject	Reject	Accept	Accept	Reject	Reject	Accept	Accept
	Traffic light Result	Yellow	Yellow	Green	Green	Red	Red	Green	Green	Red	Red	Green	Green

4.4 Comparison

Back-testing acceptance ratios are calculated and summarized in four categories: all models, models at 99% & 99.9% confidence levels, all GARCH models and GARCH models at 99% & 99.9% confidence levels. Results are showed in table 16, 17, 18 & 19. The reason of no comparison for parametric ξ in these tables is because only several differences figured in the results and both situations - $\xi > 0$ is better or worse - are noticeable.

Comparing table 16 with table 17 and table 18 with table 19, I find that at lower thresholds - 1%,2% - there is no big difference between VaR models or ES models for both Ford and the S&P 500. However, when increasing the threshold to 3%, 4%, there is a higher rejection rate in both VaR and ES models when $\alpha = 95\%$. Comparing table 16 with table 18 and table 17 with table 19, I discover that conditional GARCH(1,1) models, captured the current market condition, perform better than unconditional ones.

Furthermore, both ES and VaR models on average did worse in 2008 and 2009, the years with higher volatilities. For five-year-overall, the acceptance ratios are quite consistent when $\alpha = 99\% \ or \ \alpha = 99.9\%$. Also, in GARCH(1,1) models at 99% and 99,9% confidence levels, nearly all models are passed back-testing, besides a few models: all ES models in 2008 and both VaR and ES models when $u = 4\%$ for Ford in 2007 and 2009.

***Table 16**: Back-testing Acceptance Ratios – All models*

	VaR				ES					VaR		ES	
Ford	1%	2%	3%	4%	1%	2%	3%	4%	***S&P 500***	1%	2%	1%	2%
2007	100%	100%	83%	67%	100%	100%	83%	67%	2007	100%	100%	100%	100%
2008	50%	33%	33%	33%	8%	0%	17%	17%	2008	50%	50%	17%	17%
2009	58%	67%	50%	58%	58%	50%	50%	33%	2009	83%	83%	58%	83%
2010	83%	83%	83%	67%	100%	100%	100%	100%	2010	100%	100%	100%	100%
2011	83%	83%	67%	67%	100%	100%	100%	100%	2011	100%	100%	100%	100%
5-Year	50%	42%	33%	33%	50%	50%	33%	33%	5-Year	50%	50%	50%	50%
Total	**71%**	**68%**	**58%**	**54%**	**69%**	**67%**	**64%**	**58%**	**Total**	**81%**	**81%**	**71%**	**75%**

***Table 17**: Back-testing Acceptance Ratios – Models at 99% and 99,9% confidence levels*

	VaR				ES					VaR		ES	
Ford	1%	2%	3%	4%	1%	2%	3%	4%	***S&P 500***	1%	2%	1%	2%
2007	100%	100%	100%	75%	100%	100%	100%	75%	2007	100%	100%	100%	100%
2008	50%	50%	50%	50%	13%	0%	25%	25%	2008	50%	50%	25%	0%
2009	63%	75%	75%	88%	63%	63%	75%	50%	2009	100%	100%	63%	100%
2010	100%	100%	100%	100%	100%	100%	100%	100%	2010	100%	100%	100%	100%
2011	100%	100%	100%	100%	100%	100%	100%	100%	2011	100%	100%	100%	100%
5-Year	50%	50%	50%	50%	50%	50%	50%	50%	5-Year	50%	50%	50%	50%
Total	**77%**	**79%**	**79%**	**77%**	**71%**	**69%**	**75%**	**67%**	**Total**	**83%**	**83%**	**73%**	**75%**

***Table 18**: Back-testing Acceptance Ratios – All GARCH models*

	VaR				ES					VaR		ES	
Ford	1%	2%	3%	4%	1%	2%	3%	4%	***S&P 500***	1%	2%	1%	2%
2007	100%	100%	67%	33%	100%	100%	67%	33%	2007	100%	100%	100%	100%
2008	100%	67%	67%	67%	17%	0%	33%	33%	2008	100%	100%	33%	33%
2009	100%	100%	67%	67%	100%	83%	67%	33%	2009	100%	100%	100%	100%
2010	100%	100%	100%	67%	100%	100%	100%	100%	2010	100%	100%	100%	100%
2011	100%	100%	67%	67%	100%	100%	100%	100%	2011	100%	100%	100%	100%
5-Year	100%	83%	67%	67%	100%	100%	67%	67%	5-Year	100%	100%	100%	100%
Total	**100%**	**92%**	**72%**	**61%**	**86%**	**81%**	**72%**	**61%**	**Total**	**100%**	**100%**	**89%**	**89%**

***Table 19**: Back-testing Acceptance Ratios – GARCH models at 99% and 99,9% confidence levels*

	VaR				ES					VaR		ES	
Ford	1%	2%	3%	4%	1%	2%	3%	4%	***S&P 500***	1%	2%	1%	2%
2007	100%	100%	100%	50%	100%	100%	100%	50%	2007	100%	100%	100%	100%
2008	100%	100%	100%	100%	25%	0%	50%	50%	2008	100%	100%	50%	0%
2009	100%	100%	100%	100%	100%	100%	100%	50%	2009	100%	100%	100%	100%
2010	100%	100%	100%	100%	100%	100%	100%	100%	2010	100%	100%	100%	100%
2011	100%	100%	100%	100%	100%	100%	100%	100%	2011	100%	100%	100%	100%
5-Year	100%	100%	100%	100%	100%	100%	100%	100%	5-Year	100%	100%	100%	100%
Total	**100%**	**100%**	**100%**	**92%**	**88%**	**83%**	**92%**	**75%**	**Total**	**100%**	**100%**	**92%**	**83%**

5 Conclusion

Extreme value theory is a way of smoothing the tails of the probability distribution of portfolio daily changes calculated using historical simulation. It leads to estimates of VaR and ES that reflect the whole shape of the tail of the distribution, not just the positions of a few losses in the tails. In addition, EVT could focus on directly modeling the right tail of the loss distribution, in contrast to other parametric methods that models the whole distribution. As stated by previous researchers that EVT can be used to improve VaR and ES estimates and to help in situations where analysts want to estimate VaR or ES with a very high confidence level, my study shows the same result by both two distributions - models perform better at 99% or 99.9% confidence levels. But according to the higher back-testing acceptance ratios in VaR than in ES, it should be noted that the estimation error is larger for ES than for VaR when fat-tailed distributions are used.

In the GARCH(1,1) model, the weights assigned to observations decrease exponentially as the observations become older, the more recent an observation, the greater the weight assigned to it. The GARCH(1,1) model keeps track of the current level of volatility, it improve the test results, especially during financial crisis periods when there are more fluctuation log-returns. Thus, to capture the market volatilities is one of the most important methods in financial risk management.

Moreover, the threshold u does impact on the results, but not significant. We want u to be sufficiently high that we are truly investigating the shape of the tail of the distribution, but sufficiently low that the number of data items included in the maximum likelihood calculation is not too low. It is worthwhile computing the models with different thresholds as I find that there must be at least one observation larger than u in the conditional model in addition to the rule of thumb that u should be approximately equal to the 95th percentile of the empirical distribution suggested by the book.

Final, due to the limitation of the paper, suggestion for future studies would be to compute the left tail risk and apply to more portfolios. Also, more sophisticated ES back-testing processes could be applied. (See, Paul Embrechts, Roger Kaufmann & Pierre Patie (2005) Strategic Long-Term Financial Risks: Single Risk Factors, Chapter 7,1 Backtesting)

References

Acerbi, C. & Szekely, B. (2015). Backtesting Expected Shortfall [pdf] Available at: https://workspace.imperial.ac.uk/mathfin/Public/Seminars%202014-2015/Acerbi_January2015_Slides.pdf

Alexander J. McNeil & Thomas Saladin (1997). The Peaks over Thresholds Method for Estimating High Quantiles of Loss Distributions [pdf] Availabel at http://www.macs.hw.ac.uk/~mcneil/ftp/cairns.pdf

Bollerslev, T. "Generalized Autoregressive Conditional Heteroscedasticity." Journal of Econometrics 31 (1986): 307–327.

Chris Brooks, 2014, Introductory Econometrics for Finance 9.8, 9.9 pp428-438 pp484-489

Da Silva, A.C., and Mendez, B. V. D. M. (2003), Value at Risk and extreme Returns in Asian Stock Markets, International Journal of Business, 8, 17-40.

Danielsson, J. & de Vries C.G. (2000). Value-at-Risk and Extreme Returns, Annales d'Economie et de Statistique, ENSAE, 60, pp. 239–270.

Diebold, F. X., Schuermann, T., and Stroughair, J. D. (1998). Pitfalls and opportunities in the use of extreme value theory in risk management. In Refenes, A.-P., Burgess, A., and Moody, J., editors, Decision Technologies for Computational Finance, pages 3–12. Kluwer Academic Publishers.

Duffie, D., and J. Pan. "An Overview of Value at Risk." Journal of Derivatives 4, no. 3 (Spring 1997): 7–49.

Echaust, K. & Just, M. (2013). Conditional Versus Unconditional Models For VaR Measurement. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2365588

Embrechts, P., C. Kluppelberg, and T. Mikosch. Modeling Extremal Events for Insurance and Finance (New York: Springer, 1997).

Gerstädt, Filippa and Olander, Maria. Expected Shortfall as a Complement to Value at Risk - A study applied to commodities. [pdf] Available at:https://lup.lub.lu.se/student-papers/search/publication/1613445

Gianluca Rosso (2015). Extreme Value Theory for Time Series using Peak-Over-Threshold method. [pdf] Available at https://arxiv.org/ftp/arxiv/papers/1509/1509.01051.pdf

Huan Liu & Stacy Kuntjoro (2015). Measuring Risk with Expected Shortfall [pdf] Available at: https://lup.lub.lu.se/student-papers/search/publication/5470705

Hull, John C.(2015). Risk Management and Financial Institutions, 4th edition, Wiley Finance C10, C12, C14,C17 pp201-324

Jondeau, E., & Rockinger, M. (1999). The tail behavior of stock returns: Emerging versus mature markets. Documents de Travail 66: Banque de France.

Lanciné Kourouma, Denis Dupre, Gilles Sanfilippo and Ollivier Taramasco (2011). Extreme Value at Risk and Expected Shortfall during Financial Crisis. [pdf] Available at https://halshs.archives-ouvertes.fr/halshs-00658495/document

Manfred Gilli & Evis Këllezi (2006). An Application of Extreme Value Theory for Measuring Financial Risk [pdf] Available at: http://www.unige.ch/ses/dsec/static/gilli/evtrm/GilliKelleziCE.pdf

Manfred Gilli and Evis Kèllezi (2006). An Application of Extreme Value Theory for Measuring Financial Risk. [pdf] Availabel at http://www.unige.ch/ses/dsec/static/gilli/evtrm/GilliKelleziCE.pdf

Marshall, C., and M. Siegel. “Value at Risk: Implementing a Risk Measurement Standard.” Journal of Derivatives 4, no. 3 (Spring 1997): 91–111.

McNeil, A. J. “Extreme Value Theory for Risk Managers.” In Internal Modeling and CAD II (London: Risk Books, 1999). See also www.macs.hw.ac.uk/∼mcneil/ftp/cad.pdf.

Neftci, S. N. “Value at Risk Calculations, Extreme Events and Tail Estimation.” Journal of Derivatives 7, no. 3 (Spring 2000): 23–38.

"P. Kupiec, “Techniques for Verifying the Accuracy of Risk Management Models,” Journal of Derivatives 3 (1995): 73–84."

Reiss, R. D. and Thomas, M. (1997). Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields. Birkhäuser Verlag, Basel.

Vladimir O. Andreev, Sergey E. Tinykov, Oksana P. Ovchinnikova and Gennady P. Parahin (2012). Extreme Value Theory and Peaks Over Threshold Model in the Russian Stock Market. [pdf] Available at https://core.ac.uk/download/files/960/38634342.pdf

Appendix A. GARCH(1,1) EView Results – Ford

2007 Forecast

Dependent Variable: LNR
Method: ML - ARCH (Marquardt) - Normal distribution
Date: 07/04/16 Time: 13:53
Sample: 1/02/2002 12/29/2006
Included observations: 1259
Convergence achieved after 10 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(1) + C(2)*RESID(-1)^2 + C(3)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
Variance Equation				
C	0.074705	0.019483	3.834356	0.0001
RESID(-1)^2	0.060202	0.009161	6.571695	0.0000
GARCH(-1)	0.927650	0.009715	95.48844	0.0000

R-squared	-0.000596	Mean dependent var	-0.058673
Adjusted R-squared	0.000199	S.D. dependent var	2.405074
S.E. of regression	2.404835	Akaike info criterion	4.443283
Sum squared resid	7281.087	Schwarz criterion	4.455526
Log likelihood	-2794.047	Hannan-Quinn criter.	4.447884
Durbin-Watson stat	2.029513		

2008 Forecast

Dependent Variable: LNR
Method: ML - ARCH (Marquardt) - Normal distribution
Date: 07/04/16 Time: 13:56
Sample: 1/02/2003 12/31/2007
Included observations: 1258
Convergence achieved after 11 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(1) + C(2)*RESID(-1)^2 + C(3)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
Variance Equation				
C	0.111359	0.027337	4.073649	0.0000
RESID(-1)^2	0.056259	0.010293	5.465581	0.0000
GARCH(-1)	0.919346	0.012717	72.29416	0.0000

R-squared	-0.000146	Mean dependent var	-0.025711
Adjusted R-squared	0.000649	S.D. dependent var	2.129530
S.E. of regression	2.128839	Akaike info criterion	4.287078
Sum squared resid	5701.198	Schwarz criterion	4.299329
Log likelihood	-2693.572	Hannan-Quinn criter.	4.291682
Durbin-Watson stat	1.989758		

2009 Forecast

Dependent Variable: LNR
Method: ML - ARCH (Marquardt) - Normal distribution
Date: 07/04/16 Time: 13:57
Sample: 1/02/2004 12/31/2008
Included observations: 1259
Convergence achieved after 13 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(1) + C(2)*RESID(-1)^2 + C(3)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
Variance Equation				
C	0.028253	0.011336	2.492258	0.0127
RESID(-1)^2	0.065611	0.007494	8.755319	0.0000
GARCH(-1)	0.936986	0.006647	140.9615	0.0000

R-squared	-0.002171	Mean dependent var	-0.154411
Adjusted R-squared	-0.001375	S.D. dependent var	3.314999
S.E. of regression	3.317278	Akaike info criterion	4.597236
Sum squared resid	13854.46	Schwarz criterion	4.609479
Log likelihood	-2890.960	Hannan-Quinn criter.	4.601837
Durbin-Watson stat	1.892671		

2010 Forecast

Dependent Variable: LNR
Method: ML - ARCH (Marquardt) - Normal distribution
Date: 07/04/16 Time: 13:59
Sample: 1/03/2005 12/31/2009
Included observations: 1259
Convergence achieved after 15 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(1) + C(2)*RESID(-1)^2 + C(3)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
Variance Equation				
C	0.074204	0.018046	4.111916	0.0000
RESID(-1)^2	0.078590	0.008293	9.476345	0.0000
GARCH(-1)	0.918004	0.006523	140.7397	0.0000

R-squared	-0.000066	Mean dependent var	-0.030276
Adjusted R-squared	0.000729	S.D. dependent var	3.741972
S.E. of regression	3.740609	Akaike info criterion	4.923025
Sum squared resid	17616.12	Schwarz criterion	4.935269
Log likelihood	-3096.044	Hannan-Quinn criter.	4.927626
Durbin-Watson stat	1.834269		

2011 Forecast

Dependent Variable: LNR
Method: ML - ARCH (Marquardt) - Normal distribution
Date: 07/04/16 Time: 14:00
Sample: 1/03/2006 12/31/2010
Included observations: 1259
Convergence achieved after 13 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(1) + C(2)*RESID(-1)^2 + C(3)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
Variance Equation				
C	0.074642	0.021148	3.529537	0.0004
RESID(-1)^2	0.071419	0.007707	9.267414	0.0000
GARCH(-1)	0.924172	0.006443	143.4441	0.0000

R-squared	-0.000263	Mean dependent var	0.061713
Adjusted R-squared	0.000531	S.D. dependent var	3.805507
S.E. of regression	3.804496	Akaike info criterion	5.038570
Sum squared resid	18223.01	Schwarz criterion	5.050813
Log likelihood	-3168.780	Hannan-Quinn criter.	5.043171
Durbin-Watson stat	1.839042		

Appendix B. GARCH(1,1) EView Results – the S&P 500

2007 Forecast

Dependent Variable: LNR
Method: ML - ARCH (Marquardt) - Normal distribution
Date: 07/07/16 Time: 11:49
Sample: 1/02/2002 12/29/2006
Included observations: 1259
Convergence achieved after 8 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(1) + C(2)*RESID(-1)^2 + C(3)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
		Variance Equation		
C	0.005038	0.002760	1.825081	0.0680
RESID(-1)^2	0.055580	0.010752	5.169513	0.0000
GARCH(-1)	0.938106	0.012171	77.07786	0.0000
R-squared	-0.000274	Mean dependent var		0.016789
Adjusted R-squared	0.000520	S.D. dependent var		1.014088
S.E. of regression	1.013824	Akaike info criterion		2.519449
Sum squared resid	1294.050	Schwarz criterion		2.531693
Log likelihood	-1582.993	Hannan-Quinn criter.		2.524050
Durbin-Watson stat	2.094485			

2008 Forecast

Dependent Variable: LNR
Method: ML - ARCH (Marquardt) - Normal distribution
Date: 07/07/16 Time: 11:50
Sample: 1/02/2003 12/31/2007
Included observations: 1258
Convergence achieved after 10 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(1) + C(2)*RESID(-1)^2 + C(3)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
		Variance Equation		
C	0.011951	0.003022	3.954962	0.0001
RESID(-1)^2	0.043311	0.009007	4.808403	0.0000
GARCH(-1)	0.936172	0.012533	74.69925	0.0000
R-squared	-0.002390	Mean dependent var		0.040714
Adjusted R-squared	-0.001593	S.D. dependent var		0.833167
S.E. of regression	0.833830	Akaike info criterion		2.330668
Sum squared resid	874.6536	Schwarz criterion		2.342919
Log likelihood	-1462.990	Hannan-Quinn criter.		2.335272
Durbin-Watson stat	2.195567			

2009 Forecast

Dependent Variable: LNR
Method: ML - ARCH (Marquardt) - Normal distribution
Date: 07/07/16 Time: 11:51
Sample: 1/02/2004 12/31/2008
Included observations: 1259
Convergence achieved after 11 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(1) + C(2)*RESID(-1)^2 + C(3)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
		Variance Equation		
C	0.011042	0.003033	3.640220	0.0003
RESID(-1)^2	0.073777	0.011547	6.389409	0.0000
GARCH(-1)	0.917045	0.013743	66.72749	0.0000
R-squared	-0.000151	Mean dependent var		-0.016509
Adjusted R-squared	0.000644	S.D. dependent var		1.344694
S.E. of regression	1.344261	Akaike info criterion		2.614676
Sum squared resid	2275.060	Schwarz criterion		2.626919
Log likelihood	-1642.938	Hannan-Quinn criter.		2.619277
Durbin-Watson stat	2.265543			

2010 Forecast

Dependent Variable: LNR
Method: ML - ARCH (Marquardt) - Normal distribution
Date: 07/07/16 Time: 11:51
Sample: 1/03/2005 12/31/2009
Included observations: 1259
Convergence achieved after 15 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(1) + C(2)*RESID(-1)^2 + C(3)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
		Variance Equation		
C	0.010690	0.002397	4.459029	0.0000
RESID(-1)^2	0.073311	0.010533	6.960206	0.0000
GARCH(-1)	0.919036	0.011220	81.91211	0.0000
R-squared	-0.000019	Mean dependent var		-0.006613
Adjusted R-squared	0.000775	S.D. dependent var		1.516831
S.E. of regression	1.516243	Akaike info criterion		2.915851
Sum squared resid	2894.430	Schwarz criterion		2.928094
Log likelihood	-1832.528	Hannan-Quinn criter.		2.920452
Durbin-Watson stat	2.264642			

2011 Forecast

Dependent Variable: LNR
Method: ML - ARCH (Marquardt) - Normal distribution
Date: 07/07/16 Time: 11:52
Sample: 1/03/2006 12/31/2010
Included observations: 1259
Convergence achieved after 15 iterations
Presample variance: backcast (parameter = 0.7)
GARCH = C(1) + C(2)*RESID(-1)^2 + C(3)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
		Variance Equation		
C	0.015066	0.003029	4.973626	0.0000
RESID(-1)^2	0.085517	0.011441	7.474842	0.0000
GARCH(-1)	0.906577	0.011770	77.02497	0.0000
R-squared	-0.000000	Mean dependent var		0.000593
Adjusted R-squared	0.000794	S.D. dependent var		1.573476
S.E. of regression	1.572851	Akaike info criterion		3.111894
Sum squared resid	3114.591	Schwarz criterion		3.124137
Log likelihood	-1955.937	Hannan-Quinn criter.		3.116495
Durbin-Watson stat	2.248794			

Appendix C. GARCH(1,1) Solver Results comparing to EViews'

Ford	2007		2008		2009		2010		2011	
	Solver	Eviews	Solver	Eviews	Solver	Eviews	Solver	Eviews	Solver	Eviews
w	0,075	0,075	0,094	0,111	0,033	0,028	0,033	0,084	0,089	0,075
α	0,061	0,060	0,053	0,056	0,072	0,066	0,072	0,096	0,080	0,071
β	0,927	0,928	0,927	0,919	0,931	0,937	0,931	0,902	0,915	0,924
Sum LnL	-2793	-2793	-2694	-2694	-2893	-2893	-3105	-3105	-3170	-3170

S&P 500	2007		2008		2009		2010		2011	
	Solver	Eviews	Solver	Eviews	Solver	Eviews	Solver	Eviews	Solver	Eviews
w	0,004	0,005	0,012	0,012	0,013	0,011	0,012	0,011	0,017	0,015
α	0,050	0,056	0,051	0,043	0,079	0,074	0,081	0,073	0,093	0,086
β	0,944	0,938	0,929	0,936	0,910	0,917	0,910	0,919	0,899	0,907
Sum LnL	-1580	-1580	-1467	-1467	-1648	-1648	-1838	-1838	-1959	-1959