LUND UNIVERSITY
School of Economics and Management

Expected Shortfall Estimation

The Marginal Gain of Increased Distributional Complexity

by

Kristina Boehm

May 2019

Master's Programme in Finance

Supervisor: Birger Nilsson

LUND UNIVERSITY
School of Economics and Management

Abstract

This thesis evaluates the performance of Expected Shortfall estimation with normal, student-t and skewed distributions. It is stylized fact that student-t distribution generally outperforms normal distribution. What is particularly peculiar is whether there is marginal gain of increased distributional complexity with combining two half normal (skewed) distributions, developed by de Roon and Karehnke (2016), in comparison with t-distribution. In the cited paper authors suggest that recent research has identified skewness as one of the most prominent features of risk. For my research I utilized daily total returns (TR) on four composite indexes: Standard & Poor 500 (S&P 500), Russell 2000, Morgan Stanley Capital International (MSCI) and Goldman Sachs Commodity Index (GSCI). The sample period used for the empirical analysis runs from January 2002 to the end of December 2018. Nonetheless, MSCI is only available starting from 2007. Once distributions are estimated, I implement back-testing methodology to evaluate which outputs pass the traffic light test developed by Costanzino and Curran (2018). From the results presented in this paper, I conclude that generally skewed and t-distributions outperform the normal distribution in fitting financial returns and forecasting Expected Shortfall. However, the winner model remains student-t distribution with fat tails.

Keywords: normal, skewed, t-distribution, Expected Shortfall, Value at Risk.

Supervisor: Birger Nilsson

Acknowledgements

I would like to thank my supervisor Birger Nilsson for his time and dedication to help me find an interesting focus for the study, as well as advising me with model implementations for this research.

Table of Contents

1. Introduction

Amidst financial crisis Fundamental Review of the Trading Book (FRTB) officially recommends a switch from Value at Risk VaR to Expected Shortfall ES, as a regulatory treatment (BIS, 2013). A risk metrics ES is not universal and frequently referred to as *conditional* VaR – cVaR (Miller, 2019) *conditional tail expectation* - *CTE*, or *expected tail loss* - *ETL* (Hull, 2018), *average* VaR – AVaR (Hult, 2012). ES similarly to VaR uses time periods and confidence levels. Yet ES forecasts losses beyond VaR by capturing tails of abnormal events. What's more, normal distribution popularity is continuing to decline. Incorporating values which enhance the probability of extreme events is a possible remedy (FT [a, b], 2012).

Recent publication in statistics *Finite Mixture of Skewed Distributions,* by Dávila, Cabral and Zeller (2018), brings forward a stylized fact that normal distribution is highly unrealistic for skewed data and heavy tails. A more adaptable class of flexible models to non-normality behavior seems more appropriate.

On a general level skewness is a concept essential to various financial models. Looking from a different dimension a paper on *Low Risk Anomalies?* demonstrates controlling for skewness lowers the impact of the alphas of betting-against-beta and volatility (Schneider & Wagner & Zechner, 2016).

One of the relevant studies on the subject matter – *A simple skewed distribution with asset pricing applications* – a cornerstone to this thesis, suggests that even modest levels of skewness have a large impact on ES estimation. The paper articulates – for some quantiles ES calculated with the smooth half-normal distribution appears to be closer to ES than, e.g., the skewed t-distribution (de Roon & Karehnke, 2016).

1.1 Objectives

Daily excess index returns of large stocks, small stocks, commodities and emerging markets are analyzed over the risk-free rate in the period for past 15 years, as well as in five year clusters in this thesis. De Roon and Karehnke (2016) suggest the portfolio selection for CARA (Constant Absolute Risk Aversion) utility functions, imply a preference for skewness and are applicable for assessment of the cost of negative skewness. The finding of the paper is that the utility gain of taking skewness into account is larger with high expected excess return, low volatility, and when investors are more risk tolerant and do not face short-sale constraints. Also the utility gains of taking skewness into account are smaller than the risk premium associated with skewness. The limitation of the cited study is that they don't actually test their model.

As the final step a Traffic Light test developed by Costanzino and Curran (2018) with ES is applied to back-test distribution scenarios of various asset classes. Critical values derived from the finite-sample distribution of ES test statistics are generated for normal, skewed and t-distributions to determine which distributional assumption has least violations.

1.2 Research Purpose

As aforesaid, it is common that financial asset returns do not follow a normal distribution, but are rather fat-tailed – leptokurtic. This fact has several implications. First, models and inference procedures should be resistant to non-normal error distributions. Second, measuring riskiness with variance alone is not sufficient. Finally, assuming normality when returns are fat tailed will result in a systematic underestimation of the riskiness of the portfolio. Thus Student-t distribution accounts for this problem allowing leptokurtosis with the usual degrees of freedom, which control the fatness of the tails fitted from the model (Brooks, 2008).

To allow for the leptokurtosis in financial data Brooks (2008) argues the simplest approach is the use of a mixture of normal distributions. It can be seen that a mixture of normal distributions with different variances will lead to an overall series that is leptokurtic. De Roon and Karehnke (2016) present a distribution which they claim skewed risks no more complex than normally distributed (symmetric) risks. Their distribution is a combination of the 'downside' and 'upside' half of two normal distributions [figures 5-6, appendix A].

The purpose of this research is to investigate the marginal gain of increased distributional complexity on ES with an Exponentially Weighted Moving Average (EWMA) volatility model with normal, student-t and skewed distribution. What is particularly interesting is whether skewed distribution will surpass the t-distribution in number of least violations or will t-distribution remain as the most feasible option.

1.3 Delimitations

A possible delimitation of this research is that one back-testing method was implemented with a single confidence level of 99%. However there is no valuable argumentation doing otherwise. De Roon and Karehnke (2016) recommend using 99% confidence (significance level). Often a risk manager who measures at 95% confidence level will experience an exceedance every 20 days, and measuring at 99.9%, once every 1,000 days. If an event occurs once every 20 days is it really significant? Thus the risk manager using the 99.9% confidence level is concerned with riskier outcomes, and therefore will achieve more sound results (Miller, 2018).

Empirical findings of Degiannakis and Potamia (2016), however, suggest risk modeling at a confidence level of 97.5%, consistent with the Basel Committee proposal to replace 99% VaR by 97.5% ES. Kellner and Rösch (2016) provide evidence that under models allowing for skewness and heavy tails the level of capitalization would be higher when using 97.5% ES instead of the 99% VaR.

2 Literature Review

The focus of the literature review is to identify relevant sources covering skewness as well as specific researches implementing skewness for ES. To see which model wins the race I overview back-testing possibilities, which will later be implemented in 'Analysis and Discussion' section.

2.1 Skewed Distribution with ES

Literature presenting skewness as an essential feature of distribution is abundant. Degiannakis and Potamia (2016) claim the use of a skewed rather than symmetrical distribution produces more precise VaR and ES forecasts. They mention that findings of Giot, and Laurent (2003); Angelidis, Benos and Degiannakis (2004) as well as Degiannakis, Dent, and Floros (2014) confirm better results for the skewed student-t distribution under GARCH specification than a symmetric one across a variety of asset classes. Accordingly, Braione and Scholtes (2016) demonstrated the significance of allowing for heavy-tails and skewness with the student-t outperforming the others across all tests and confidence levels.

Mixtures of normal or student-t distributions may capture both leptokurtosis and skewness in return distributions and they usually address various market regimes. As an example, in a mixture of two normal distributions, there are two regimes for returns: one where the return has mean μ_1 and variance σ_1^2 and another where the return has mean μ_2 and variance σ_2^2. The other parameter of the mixture is the probability p with which the first regime occurs, so the second regime occurs with probability $1-p$ (Alexander, 2008).

The mixture distribution is a probability-weighted sum of the component distribution functions. For greater flexibility to fit the empirical return distribution, more than two component may be incorporated in the distribution. As the number of distributions in the mixture increases the probability weight on these components lowers. However, in finance it is not always necessary to use more than two or three components in the mixture, since financial asset return distributions are seldom irregular to the point to have multiple modes (Alexander, 2008).

Alexander (2008) refers to the case study illustrating the application of different parametric linear models to estimate the *ETL* for an exposure to the iTraxx Europe 5-year credit spread index. The historical distribution of this risk factor is non-normal, with a noticeable negative skewness and a high excess kurtosis, and its daily changes have a significant positive autocorrelation. Hence, the assumption was that the normal i.i.d. (independently and identically distributed) model is inappropriate and would underestimate the risk of such an exposure. Thus a mixture of distributions is recommended.

The paper titled *A simple skewed distribution with asset pricing applications* depicts a skewed distribution based on the combination of the halves of two normal distributions. This distribution can be parametrized in closed-form as a function of a given mean, variance, and skewness, and its easy application for ES. The distribution has insignificant excess kurtosis, enabling it to compare statics of skewness for portfolio choice and asset pricing. With data selected as monthly index returns on different asset classes, the methodology well identifies the left tail and overall shape of the empirical distributions (de Roon & Karehnke, 2016).

De Roon and Karehnke (2016) mention numerous studies honoring skewness. They argue that Cumulative Prospect Theory (CPT) is essential for portfolio selection with a reference to a paper by Ebert and Strack (2015), which suggests that CPT is possibly the most prominent alternative to Expected Utility Theory (EUT). The study has a high preference for skewness. Other referenced study by Schneider and Spalt (2016) shows that capital expenditure increases in the expected skewness of segment returns.

2.2 ES and Backtesting

A measure with better overall capacity to generate more sound incentives for traders than VaR is ES. 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 X (the confidence level). In order to calculate ES it is necessary to calculate VaR first. ES is the expected loss during time T conditional on the loss being greater than the VaR (Hull, 2018).

ES surpasses VaR in a way that it acknowledges the benefits of diversification. Due to the lack of information about the size of the tail loss VaR does not always have such property and is not sub-additive (Hull, 2018). ES is a coherent risk metrics, whereas VaR is estimated using simulation and is not coherent as it is not sub-additive (Alexander, 2008).

Amongst disadvantages ES is very sensitive to the possibility of highly unlikely but very large losses (Miller, 2019). ES is more complex than VaR and more difficult to back-test (Hull, 2018; Miller 2019). In earlier years there had already been criticism towards the latter, as discussed in further paragraphs.

Back-testing is a test of how efficient the current procedure for calculating the measure would have worked in the past (Hull, 2018). A paper on back-testing by Kerkhof & Melenberg (2003) nonetheless concludes: opposed to general assumptions ES is not harder to backtest than VaR. In fact, the power of the test for ES is considerably higher. Those debates led Acerbi and Szekely (2014) to introduce model-free, nonparametric ES back-testing methodologies, contrary to the belief of Miller (2019) and Hull (2018). Acerbi and Szekely (2014) talk about three test alternatives: testing ES after VaR, testing ES directly and estimating ES from realized ranks.

A study named *A Simple Traffic Light Approach to Backtesting Expected Shortfall,* based on empirical results, recommends a traffic light test for ES using the finite-sample distribution of the test statistic under the null hypothesis. The test is similar to the Basel Traffic Light test for VaR. The test relies on the computation of critical values derived from the finite-sample distribution of the ES test statistic (Costanzino & Curran, 2018).

Costanzino and Curran (2018) in their work also refer to other approaches to back-testing ES. Du and Escanciano (2015) investigate an empirical application to three major stock indexes with Monte Carlo simulation. The results indicate that VaR is generally unresponsive to extreme events such as those experienced during the recent financial crisis, whereas ES provides a more accurate description of the risk involved.

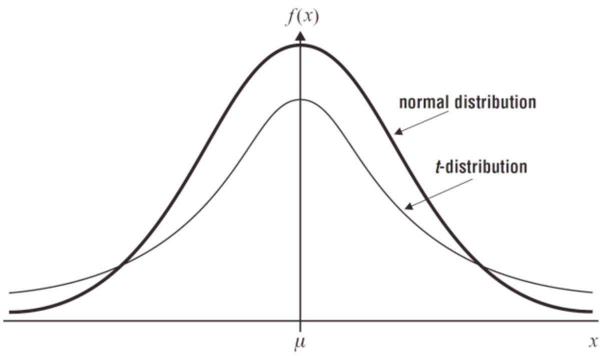
3 Methodology

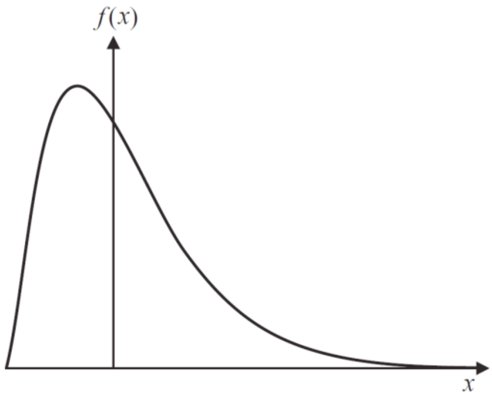
In this section I start by describing fundamental distributional assumptions relevant to my research. As I implement parametric approach I hereby define it. Normal distribution is an example of a parametric model, which is based on mathematically defined parameters the mean μ and the standard deviation σ.

Parametric models have certain limitations. Achieving a parametric model that reproduces all of the observed features of financial markets may be tricky. Yet, models based on distributions may be easier to interpret. In the case of non-parametric model, for example – the historical simulation (HS), it is difficult to say if the data used for the model are unusual because the usual is not defined (Miller, 2018).

Hence I present the collected data for specified period of time, then I transform the data into losses and gains. I address time-varying volatility as an essential feature for distributions and illustrate it graphically on losses. I compute descriptive statistics on gains to have an informative output on data characteristics. In the next section I start with EWMA (Exponentially Weighted Moving Average Volatility) and discuss model estimations.

3.1 Normal, Student-t and Skewed Distributions





The normal distribution is characterized by 'bell' shape and its symmetry around the mean. T-distribution has fatter tails and a smaller peak at the mean. Skewed distribution (on the right) will have one tail longer - in this case being positively skewed with tail extended to the right (Brooks, 2008).

First central moment of a distribution is mean μ. Variance σ^2 – the second central moment – outlines the spread of a random variable around the mean. Skewness – the third central moment tells us how symmetrical the distribution. A random variable that is symmetrical will have zero skewness.

To standardize skewness for a random variable *X*, the equation is given as

$$\frac{E[(X-\mu)^3]}{\sigma^3}$$

σ is the standard deviation of *X*
μ is the mean of *X* (Miller, 2019).

Skewness is an indispensable concept in risk valuation. If the distributions of returns of two investments are identical with the mean and standard deviation, but distinct skewness, then the investment with more negative skewness is considered to be riskier. Historical data suggests that many financial assets exhibit negative skewness. Generally, the returns of most equity indexes have negative skewness (Miller, 2019).

The fourth central moment kurtosis also tells us how spread out a random variable is by emphasizing more weight on extreme points. For a random variable *X*, the kurtosis is defined as *K*, where

$$K = \frac{E[(X-\mu)^4]}{\sigma^4}$$

σ is the standard deviation of X
μ is the mean of *X* (Miller, 2019).

A distribution with the same-size tails is defined as mesokurtic. A leptokurtic distribution has heavier tails than the normal distribution and a platykurtic one has less heavy tails (Brooks, 2008).

3.2 Data Collection

As adapted from the paper *A simple skewed distribution with asset pricing applications* following data is incorporated (de Roon & Karehnke, 2016).

Index (TR)	Abbreviation	Category
Standard & Poor 500, TR Index, Composite	S&P 500, TR Index	Large Stocks
Russell 2000, TR Index	Russel 2000, TR Index	Small Stocks
Morgan Stanley Capital International, TR Index	MSCI, TR Index	Emerging Markets
Goldman Sachs Commodity Index, TR Index	GSCI, TR Index	Commodities

Historical data on total return (TR) indices is retrieved from Datastream (developed by Thomson Reuters). Single source is preferable for precision, to avoid errors in data collection. Data is second hand, however the programme is well recognized in academic world.

Data is obtained on total return ranging from 2002-01-01 until 2018-01-31. First two years are processed as 'in sample'.

De Roon and Karehnke (2016) suggest that the given portfolio selection may be addressed with the empirical skewness measure. The distribution allows to quantify the total effect of skewness on utility. For this reason I implement the same indices. The portfolio weights of the CPT are more reactive to skewness and the utility gains of incorporating skewness are greater than for the CARA (Constant Absolute Risk Aversion) approach. The risk premium associated with skewness is larger than for CARA investors but smaller than the utility gains of allowing skewness.

3.3 Data Processing and Transformations

Daily excess returns over the risk-free rate for the full period of the past 15 years were tested on various asset classes, as well as in five year clusters. The preceding period of two years is defined as 'in sample' with no ES estimates.

$$\text{5 year test periods} \begin{cases} 2014-2018 \\ 2009-2013 \\ 2004-2008 \end{cases}$$

2002	**2003**	**5 YEARS**	**5 YEARS**	**5 YEARS**
January 1st, 2002	January 1st, 2004			December 31st, 2018

MSCI Emerging Market Total Return index is nonetheless only available from year 2007. Thus first two years are processed as 'in sample', the remaining years were estimated accordingly:

$$\text{5 year test periods} \begin{cases} 2014-2018 \\ 2009-2013 \end{cases}$$

3.3.1 Losses

In principle, the parameter estimates can be updated each holding period that each time a new loss observation is available. A 'minus sign' is placed in front of the formula to obtain the loss. Assuming $100 is invested in the beginning of each day, loss transformation is provided as:

$$L_t = -\frac{I_t - I_{t-1}}{I_{t-1}} * 100$$

I_t is defined as observation

I_{t-1} is the previous observation

Losses are illustrated on the following figures (1-4).

Figure 1. GSCI Commodity TR Index Daily Losses

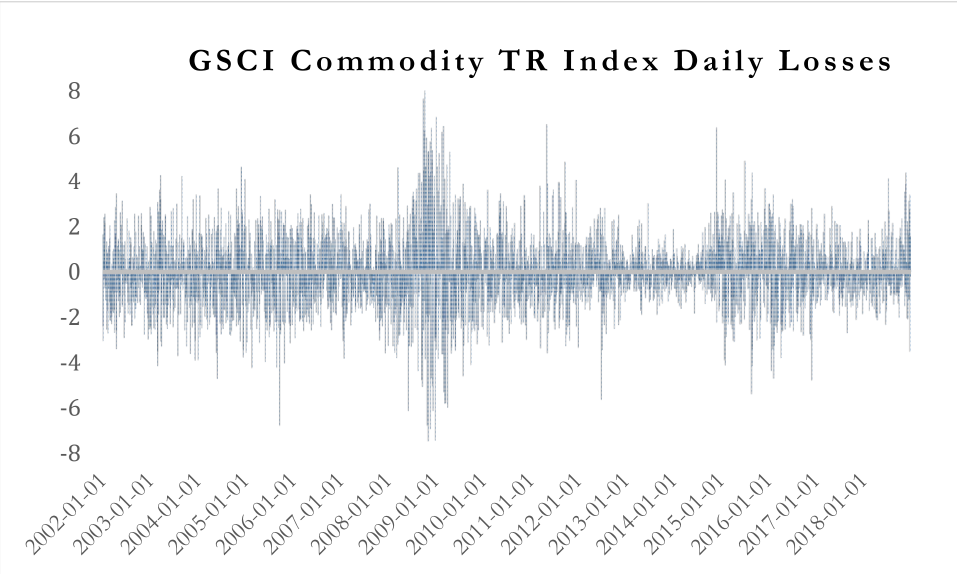


Figure 2. MSCI TR Index Daily Losses

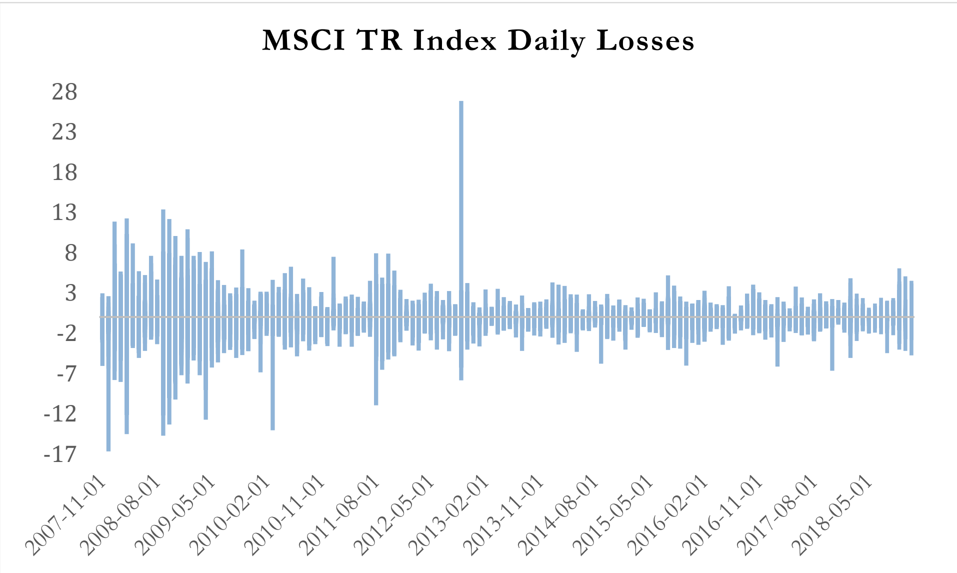


Figure 3. Russell 2000 TR Index Daily Losses

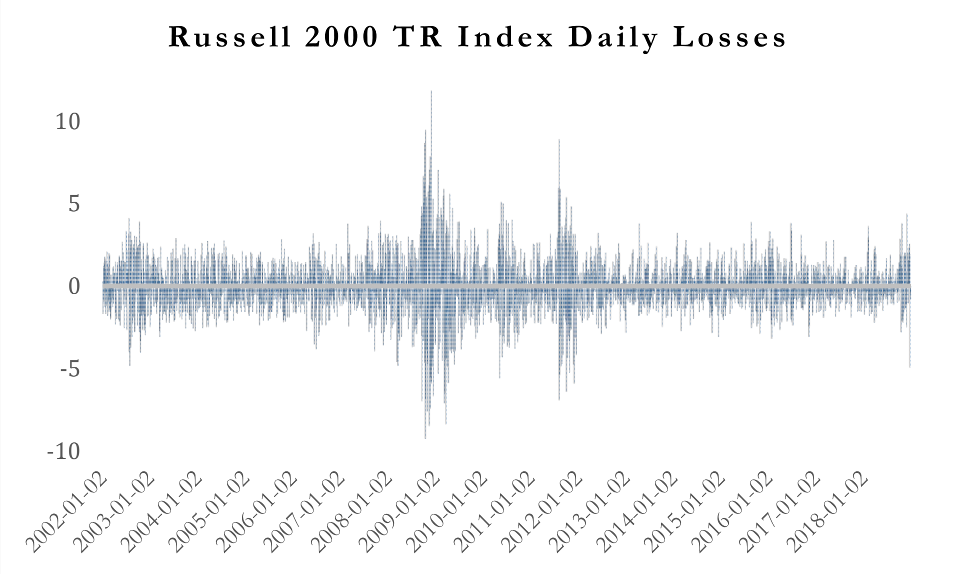
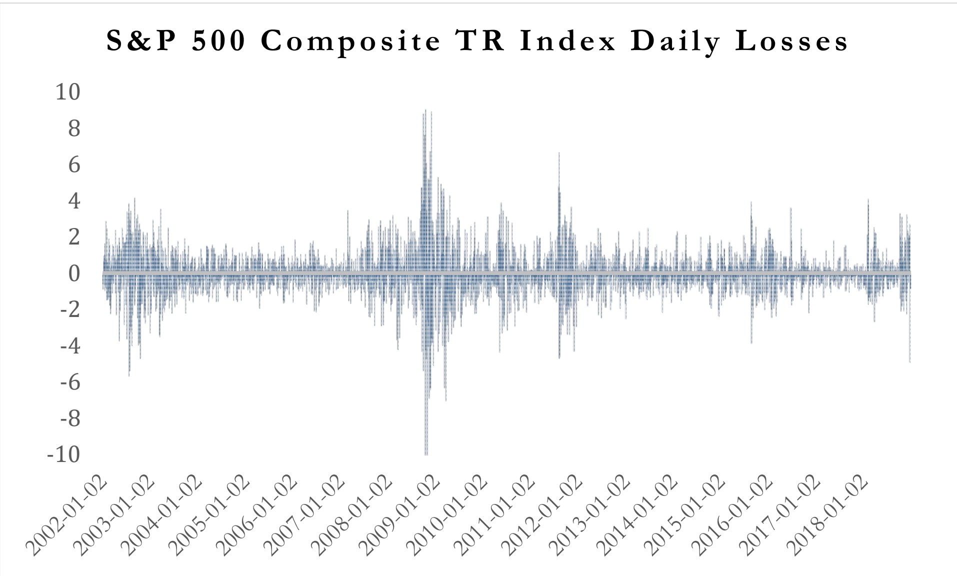


Figure 4. S&P 500 Composite TR Index Daily Losses



3.3.2 Time-varying volatility

From figures above (1-4) depicting asset losses I observe volatility clustering, when large returns are followed by large returns, and small returns are followed by small returns, appearing in clusters. As Brooks explains, this phenomenon occurs because information arrivals, which define price changes, themselves occur in clusters rather than being evenly spaced over time (2008). Volatility clustering is accounted for when selecting a model.

3.3.3 Gains

Gains, respectively, are calculated as:

$$R_t = \frac{I_t - I_{t-1}}{I_{t-1}} * 100$$

I_t is subsequently defined as observation
I_{t-1} is the previous observation

3.3.4 Descriptive Statistics on Gains

Distributional properties are introduced to have a more comprehensive overview on data by running 'Descriptive Statistics' function on gains in MS Excel. As shown in 'Descriptive Statistics', the number of observations, 4434, should be sufficient coverage. As per availability, MSCI only has 2902 observations.

On the given data set (table 1) I observe that skewness is generally time varying, not constant and is not equal to zero. A negatively skewed data set has its tail extended towards the left (Hull, 2018), which usually indicates that both the mean and the median are less than the mode of the data set.

Table 1. Descriptive Statistics on Gains

	GSCI Commodity	*RUSSELL 2000*	*S&P 500 Composite*	*MSCI*
Mean	0.0046	0.0334	0.0321	0.0884
Standard Error	0.0214	0.0217	0.0174	0.0416
Median	0	0.0368	0.0359	0.0683
Standard Deviation	1.4314	1.4503	1.1638	2.2446
Kurtosis	2.6560	5.5287	10.5617	14.2988
Skewness	-0.1028	-0.1652	-0.0336	-0.3437
Minimum	-8.2851	-11.8506	-9.0259	-26.8246
Maximum	7.4825	9.2654	11.5811	16.7156
Count	4434	4434	4434	2902

4 Analysis and Discussion

4.1 Expected Shortfall (ES)

The expected profit of a fund is a probability density function given by $f(x)$,

ES is defined as

$$ES = -\frac{1}{1-\gamma}\int_{-\infty}^{VaR} xf(x)dx$$

VaR is the VaR at the γ confidence level (Miller, 2019)

4.2 Time-varying Volatility

Hull (2018) defines a variable's volatility σ as the standard deviation of the return provided by the variable per unit of time when the return is expressed using continuous compounding. Time-varying volatility is a stylized fact of the vast majority of asset returns and it is essential to be integrated into the model. As EWMA (Exponentially Weighted Moving Average Volatility) model tracks changes in the volatility – the data is tested with the following distributional assumption:

$$\text{EWMA}\begin{cases} Normal\ Distribution \\ Student - t\ Distribution \\ Skewed\ Distribution \end{cases}$$

4.2.1 The Exponentially Weighted Moving Average Volatility Model (EWMA)

Since EWMA as a particular case of GARCH, I first define
GARCH (1,1) - Generalized Autoregressive Conditional Heteroscedasticity

$$\sigma_n^2 = \gamma V_L + \alpha u_{n-1}^2 + \beta\sigma_{n-1}^2$$

V_L long-run average variance rate
γ is the weight assigned to V_L
α is the weight assigned to u_{n-1}^2
β is the weight assigned to σ_{n-1}^2
where weights sum to one $\gamma + \alpha + \beta = 1$
if $yV_L = \omega$ GARCH (1,1) becomes

$$\sigma_n^2 = \omega + \alpha u_{n-1}^2 + \beta\sigma_{n-1}^2$$

Instead of calculating variances and covariances, I use EWMA, which tracks changes in the volatility and is a variation of GARCH (1,1); where
$\omega = 0$
$\alpha = 1 - \lambda$
$\beta = \lambda$, λ maximizes the objective function.

The equation is

$$\sigma_n^2 = \lambda\sigma_{n-1}^2 + (1-\lambda)u_{n-1}^2$$

σ_n volatility for day n made at the end of day $n-1$
σ_{n-1} the estimate that was made at the end of day $n-2$ of the volatility for day $n-1$
u_{n-1} the most recent daily percentage change.
λ defines sensitivity of the estimate of the daily volatility (Hull, 2018).

In practice I estimate EWMA, assuming the mean is zero. Sample standard deviation is taken on first two years of observations, then

$\lambda = 0.94$
$1- \lambda = 0.06$

$$\sigma_n^2 = 0.94 * \sigma_{n-1}^2 + 0.06 * u_{n-1}^2$$

The RiskMetrics database, originally developed by JPMorgan and disclosed to public in 1994, used the EWMA model with λ = 0.94 for updating daily volatility estimates. JPMorgan established that, across a range of different market variables this value of λ gives forecasts of the variance rate that come closest to the realized variance rate (Hull, 2018).

4.2.2 Volatility and Average

Volatility σ_{t+1}

As a next step volatility σ_{t+1} is calculated as the square root of EWMA.

$$\sigma_{t+1} = \sqrt{0.06 * u_t^2 + 0.94 * \sigma_t^2}$$

Mu μ

Mu μ in the context of ES estimates is defined as the average of previous losses.

4.3 Normal (Gaussian) distribution with ES

Beta β is the confidence level or significance level (de Roon & Karehnke, 2016) equal to 1%, therefore 0.99.

$$ES_{\beta}(\mathrm{L}) = \mu + \sigma \frac{\varphi\,(\varphi^{-1}(\beta))}{1-\beta}$$

φ is the density of the standard normal distribution (McNeil, Frey, Embrechts, 2005).
* Some notations may be modified for the purpose of simplicity.

Vega v

The *vega* v of a portfolio are the degrees of freedom used in student-t distribution, where v = 120, Kurt = 3. It is the rate of change of the value of the portfolio with respect to the volatility, of the underlying asset price. If v is high in absolute terms, the portfolio's value is very sensitive to small changes in volatility. If v is low in absolute terms, volatility changes have relatively little impact on the value of the portfolio (Hull, 2018).

4.4 T-distribution with ES

Direct integration may be used to calculate ES of the standard *t* distribution

$$ES\beta\ (L) = \frac{g_v({t_v}^{-1}(\beta))}{1-\beta}\left(\frac{v+({t_v}^{-1}(\beta))^2}{v-1}\right)$$

t distribution
v degrees of freedom
t_v denotes the d_f
g_v the density of standard t (McNeil, Frey, Embrechts, 2005).
* Some notations may be modified for the purpose of simplicity.

4.5 Skewed distribution with ES

Firstly, I used the Excel 'ABS' function, which produces the absolute value of a number. Negative numbers are transformed into positive, and positive numbers remain unaffected. 'SKEW' function calculates the skewness of the distribution. I execute the test with the following restrictions: -0.995, 0.001.

To arrive at a calculation of ES with skewed distribution, I started by estimating alpha I took the square root of two divided by π which numerically equals to approximately 0.80.

$$\alpha = \sqrt{\frac{2}{\pi}}(\approx 0.80)$$

To calculate Δ delta, I used following formula

$$\Delta = -\left(\frac{\alpha}{Skew}\right)^4 (2-3\alpha^2)^3 + [27(2-3\alpha^2) - 4\left(\frac{\alpha}{Skew}\right)^2].$$

Then I calculate q, as defined, I computed

$$q = \left(\frac{\alpha}{Skew}\right)^2 [(-(2-3\alpha^2)^2 + \frac{2}{3}\left(\frac{\alpha}{Skew}\right)^2 (2-3\alpha^2) - \frac{2}{27}\left(\frac{a}{Skew}\right)^4]$$

Once q is obtained I calculated a, as provided

$$a = \sqrt[3]{-\frac{q}{2} + \frac{1}{2}\sqrt{\frac{\Delta}{27}}} + \sqrt[3]{-\frac{q}{2} - \frac{1}{2}\sqrt{\frac{\Delta}{27}}} - (1-\alpha^2) + \frac{1}{3}\left(\frac{\alpha}{Skew}\right)^2$$

Z ratio, as specified

$$z = \begin{cases} 1 + \frac{1}{2a} + \frac{1}{2a}\sqrt{4a+1} \; if \; Skew > 0, \\ 1 \; if \; Skew = 0, \\ 1 + \frac{1}{2a} + \frac{1}{2a}\sqrt{4a+1} \; if \; Skew < 0 \end{cases}$$

Henceforth, S_1 and S_2 (downside and upside standard deviation) are derived as

$$s_1 = \frac{\sigma}{[(1-\alpha^2)(z-1)^2 + z]^{1/2}}$$

$s_2 = s_1 z$ (de Roon & Karehnke, 2016).

4.5.1 Combining two normal distributions

Skewed distribution is referred to as 'smooth half-normal distribution', by combining the left-hand side and the right-hand side half of two normal distributions. Each half is scaled to ensure that the Probability Density Function (PDF) is continuous and integrates to one. The smooth half-normal distribution will be skewed on a condition that its downside and upside half are from two normal distributions with different variances (de Roon & Karehnke, 2016).

Lambda λ defines sensitivity of the estimate of the daily volatility (Hull, 2018).

Two normal distributions
with means m_1 and m_2
and standard deviations S_1 and S_2, respectively,

$m_1 = m_2 = m$ are defined with:

$$g(x; m,\ s_1, s_2\ ) = \begin{cases} f(x; m, s1), for\ x\ \leq m \\ f(x; m, s1), for\ x > m \end{cases}$$

Thus, if the two parts of g with λ_1 and λ_2 are scaled respectively, such that

$$\lambda_1 f(m; m, s_1) = \lambda_2 f(m; m, s_2),$$

a continuous (differentiable) density function is obtained.

The density is at $x = m$ equals

$$\frac{1}{\sqrt{2\pi s^2}}$$

and that the total density function should integrate to one, the following restrictions are obtained:

$$\frac{\lambda_1}{\sqrt{2\pi s_1^2}} = \frac{\lambda_2}{\sqrt{2\pi s_2^2}}$$

$$\frac{1}{2}\lambda_1 + \frac{1}{2}\lambda_2 = 1$$

from which is derived:

$$\lambda_1 = \frac{2s_1}{s_1 + s_2}$$

$$\lambda_2 = 2 - \lambda_1 = \frac{2s_2}{s_1 + s_2}$$

Finally, m is given as

$$m = \mu - as_1(z - 1)$$

(de Roon & Karehnke, 2016).

φ Phi

The skew normal with location 0 and scale 1 has the PDF of $2\varphi(x)\varphi(ax)$, where φ denotes the CDF (Cumulative Distribution Function) of the standard normal and the real number α is the shape parameter, where skewness is increasing in α. The skew normal is different from the smooth half-normal for $0 < |a| < +\infty$ (de Roon & Karehnke, 2016).

4.5.2 Distribution Estimates

As I illustrate ES with normal, skewed and t-distribution estimates with figures, I observe that results are generally correlated [figures 7-10, appendix B]. To get more accurate projection I resort to back-testing.

4.6 Back-testing

Costanzino and Curran (2018) propose a Traffic Light test for ES using the finite-sample distribution of the test statistic under the null hypothesis. The test model is somewhat a replication of the Basel Traffic Light test for VaR that measures the severity of the breach.

I assume confidence level at 99%. Back-testing involves looking at how often the loss in a day would have exceeded 99%. When the actual loss exceeds the given value, I have a violation, which is referred to as breach in the afore-cited paper of Costanzino and Curran (2018). If breaches happen on, for example, 1% of the days, the current methodology is reasonable. However, if they exceed substantially, the methodology is questionable.

An approximation is defined as follows

Green: $X_{ES}^{N} < 5.4768$
Yellow: X_{ES}^{N} (in between)
Red: $X_{ES}^{N} \geq 9.2229$

Alternatively

a Green Zone $q < 0.95$
Yellow Zone $0.95 \leq q < 0.9999$
Red Zone $0.9999 \leq q$

The derivation of the ES Traffic Light test relies on the computation of the finite-sample cumulative distribution of the test statistic

$$X_{ES}^{N}$$

Where N is total number of trading days/observations

ES generalized breach indicator $X_{ES}^{(i)}$: [0,1] → [0,1] is defined as

$$X_{ES}^{(i)}(\propto) := \frac{1}{\propto}\int_{0}^{\propto} 1\{Li \leq VaRi(p)\}^{\ dp}$$

VaR needs to be estimated to be able to say if there has been a VaR breach/violation on a given day.

$$\left(1 - \frac{F_L(L_i)}{\propto}\right) 1\ \{Li\ \leq VaR\ i\ (\propto)\}$$
$$= \theta^{(i)}(\propto) * X_{VaR}^{i}$$

$X_{ES}^{(i)}$ keeps track of whether a breach happened on trading day *i* and severity and is the basis for the test. The larger the size of violation the worse it is.

$\{t_i\}_{i=0}^{N}$ is a sequence of historical trading days

$\{L_i\}_{i=1}^{N}$ is realized trading losses

$\theta^{(i)}(\propto) = 1 - \frac{F_L(L_i)}{\propto}$

F_L the cumulative distribution of the random loss variable L

$\frac{F_L\,(Li)}{\propto}$ is the severity of the breach

X_{ES}^{N}: [0, 1] ⟶ [0, N] the total severity of breaches over all *N* trading days is

$$\sum_{i=1}^{N} \theta^{(i)}(\propto) X_{VaR}^{(i)}(\propto)$$

The severity of the breaches is measured in terms of probabilities

$$\theta^{i} = \left(1 - \frac{\Pr(L > Li)}{(1-\propto)}\right)$$

If loss is very large then Pr ($L > L^{i}$) ≈ 0

If $\theta^{i} \approx 1$ – the breach is severe if theta is close to 1

If loss is close to VaR^{i} then Pr ($L > L^{i}$) ≈ 1 - $\propto$
If $\theta^{i} \approx 0$ – the breach is mild is theta is close to zero

First I estimated the average given by

$$\mu_{ES} = \frac{1}{2}(1-\propto) * N$$

Consequently, I estimated standard deviation

$$\sigma_{ES}^{2} = (1-\propto)\left(\frac{4 - 3(1-\propto)}{12}\right) * N$$

hence

$$\lim_{N\to\infty} X_{ES}^{N}(\propto) \sim N(\mu_{ES}, \sigma_{ES}^{2}).$$

(Costanzino & Curran, 2018).

For day i $\theta^{(i)}(\propto) * X_{VaR}^{(i)}$

$$X_{ES=}^{i}\ \theta^{i} * X_{VaR}^{i}$$

Values are computed for every day, if there is no violation, then X is denoted as 0, violation, on the other hand, is denoted as 1.

Finally, I sum over all days i (*N* is number of observations or trading days of the sample):

$$X_{ES=}^{N} \sum_{i=1}^{N} \theta^{i} * X_{ES}^{i}$$

The probabilities are then be calculated: Pr $(L > L^{i})$ for following distributions

$$\begin{cases} \text{Normal distributon} \\ \text{T} - \text{distribution} \\ \text{Skewed distribution} \end{cases}$$

4.6.1 Back-testing Output Summary

When analyzing back-testing summary output, I not only look at the number of exceedances but also their values. Overall conclusion is that in large samples (over 3900 values) none of the models (normal, t, skewed) pass the test on full samples. Generally t-distribution outperforms all distributions. Skewed distribution outperforms normal in 12 out of 15 trials. On the following pages I illustrate back-testing outputs in tables 2-5, which summarize the results.

To have a better retrospective on back-testing results, I refer to Carol Alexander (2008). Their results for the normal and normal GARCH models showed that the assumption of normality is unrealistic, particularly when estimating *ETL*. Capturing volatility clustering with the GARCH process the *ETL* understated the true potential for losses beyond the VaR. Their *ETL* results for the normal mixture model were also not satisfactory, even with volatility clustering. However, the student-t GARCH model produced *ETL* test results that were the best of all the risk models considered in the study. Interestingly, their results indicated that the student-t GARCH model may even appear ultra-conservative, since no exceedances were recorded for the portfolios of their study (Alexander, 2008).

Table 2. S&P 500 Index Back-testing Outputs

Critical Values	μ_{ES}, σ^2_{ES}	**Critical Values 99% Confidence Level**	**ES-N**	**ES-T**	**ES-Skew**
Full Period N = 3913	μ_{ES}=19.5650 σ^2_{ES} =3.5980	25.48 Green Yellow 32.95 Red	66.19	36.16	57.39
First 5 years = 2004-2008 N = 1305	μ_{ES}=6.5250 σ^2_{ES} =2.0778	9.94 Green Yellow 14.25 Red	20.74	10.37	19.74
Next 5 years = 2009-2013 N = 1304	μ_{ES}=6.5200 σ^2_{ES} =2.0770	9.94 Green Yellow 14.24 Red	19.65	9.02	18.30
Last 5 years = 2014-2018 N=1304	μ_{ES}=6.5200 σ^2_{ES} =2.0770	9.94 Green Yellow 14.24 Red	25.81	16.77	19.35

For S&P 500 none of the models 'pass' over the full sample period, but the t-distribution is 'green' over one subsample and 'yellow' over another subsample. The normal and the skew never pass, but the skew is better than the normal. Surprisingly, the last 'good' period on stock markets is the period when all models fail.

Table 3. MSCI Index Back-testing Outputs

Critical Values	μ_{ES}, σ^2_{ES}	**Critical Values 99% Confidence Level**	**ES-N**	**ES-T**	**ES-Skew**
Full Period N = 2381	μ_{ES}=11.9050 σ^2_{ES} =2.8066	16.52 Green Yellow 22.34 Red	29.57	15.13	45.65
First 5 years = 2009-2014 N = 1305	μ_{ES}=6.5250 σ^2_{ES} =2.0778	9.94 Green Yellow 14.25 Red	15.79	7.59	13.54
Last 5 years = 2014-2018 N = 1076	μ_{ES}=5.3800 σ^2_{ES} =1.8867	8.48 Green Yellow 12.40 Red	13.79	7.54	11.84

For MSCI t-distribution is 'green' over the full sample period and all subsamples. However in this case the sample is smaller. Skewed distribution outperforms normal on 2 occasions and is 'yellow'.

Table 4. Russell 2000 Index Back-testing Outputs

Critical Values	μ_{ES}, σ^2_{ES}	Critical Values 99% Confidence Level	ES-N	ES-T	ES-Skew
Full Period N = 3913	μ_{ES}=19.5650 σ^2_{ES} =3.5980	25.48 Green Yellow 32.95 Red	55.07	34.00	42.18
First 5 years = 2004-2008 N = 1305	μ_{ES}=6.5250 σ^2_{ES} =2.0778	9.94 Green Yellow 14.25 Red	15.05	10.54	12.83
Next 5 years = 2009-2013 N = 1304	μ_{ES}=6.5200 σ^2_{ES} =2.0770	9.94 Green Yellow 14.24 Red	16.80	7.55	14.25
Last 5 years = 2014-2018 N=1304	μ_{ES}=6.5200 σ^2_{ES} =2.0770	9.94 Green Yellow 14.24 Red	23.23	15.91	15.10

For Russell 2000 none of the models 'pass' over the full sample period. T-distribution is the only one containing 'green' in one subsample. Overall t-distribution wins, with skewed outperforming the normal distribution, even though 'yellow' only in one subsample and 'red' in the rest.

Table 5. GSCI Index Back-testing Outputs

Critical Values	μ_{ES}, σ^2_{ES}	Critical Values 99% Confidence Level	ES-N	ES-T	ES-Skew
Full Period N = 3913	μ_{ES}=19.5650 σ^2_{ES} =3.5980	25.48 Green Yellow 32.95 Red	48.77	29.96	48.68
First 5 years = 2004-2008 N = 1305	μ_{ES}=6.5250 σ^2_{ES} =2.0778	9.94 Green Yellow 14.25 Red	9.91	6.97	13.54
Next 5 years = 2009-2013 N = 1304	μ_{ES}=6.5200 σ^2_{ES} =2.0770	9.94 Green Yellow 14.24 Red	19.60	10.50	11.84
Last 5 years = 2014-2018 N=1304	μ_{ES}=6.5200 σ^2_{ES} =2.0770	9.94 Green Yellow 14.24 Red	19.26	12.49	23.30

For GSCI, 'green' is found in one t-distribution and one normal distribution subsamples. In 2 trials normal distribution actually surpasses t-distribution for GSCI, which can be defined as 'acceptable', since normal distribution is always predisposed to fail the tests.

5 Conclusion

5.1 Research Aims and Objectives

This thesis evaluates which distribution (i) normal, (ii) student-t or (iii) skewed outperforms one another, when estimating losses for ES. Daily total returns were retrieved on four composite indexes: S&P 500, Russell 2000, MSCI and GSCI. The sample period used for the empirical analysis runs from January 2002 to the end of December 2018 and MSCI starting from 2007 upon availability.

From the results of the traffic light test presented in the tables (2-5), I conclude that t-distribution generally outperforms all distributions. Skewed distribution outperforms normal in 12 out of 15 trials. The gain from going from the normal to the t-distribution therefore seems much bigger than the gain from going from the normal to the skewed distribution. Excess kurtosis seems more important to account for than non-zero skewness.

As stated in the introduction, assuming normality when returns are fat tailed will result in an underestimation of portfolio riskiness. Thus, the results of this thesis, where student-t distribution outperforms others, are consistent with stylized assumptions. It is, however, questionable whether models with abundance of parameters provide discernible advancements, or whether the additional flexibility is after all excessive. Another question is whether it is really time efficient to estimate additional parameters.

The evidence from model estimations, statistical inference and back-testing suggests a more complex approach of combining two half normal distributions is not necessarily better in forecasting power relative to a model with less specifications , i.e. t-distribution. Nonetheless, as skewness is a complex phenomenon, it might still outperform taking other factors into account or implementing other possible models of skewness, since spectrum of possibilities for modelling skewness is very broad.

5.2 Related Research

A paper *A simple skewed distribution with asset pricing applications* by de Roon and Karehnke, (2016) has undoubtedly provided valuable input for this thesis and distributional implications of ES. It is nonetheless interesting to follow up on other upcoming researches which further develop models of skewed distributions which may benefit the financial industry. It is possible that in other approaches skewed distribution may as well outperform t-distribution.

An implemented study *A Simple Traffic Light Approach to Backtesting Expected Shortfall* developed by Costanzino and Curran (2018) has been a useful resource for estimating number of violations. It would also be practical to test the results with other back-testing methods, as they arise in the future, to see how much the results of this approach deviates from other back-testing methods. I hope that by repeating methodology in this thesis other researches may replicate the given distributions and attempt back-testing using other approaches.

References

Acerbi C, Szekely B. (2014). MSCI [pdf] Available at: Backtesting Expected Shortfall [Accessed 10 April 2019]

Alexander C. (2008). Value-at-Risk Models, Chichester, West Sussex: John Wiley & Sons Ltd.

Angelidis, T., Benos, A., Degiannakis S. (2004). Available at: The use of GARCH models in VaR estimation [pdf] [Accessed 15 April 2019]

Bank for International Settlements (BIS) (2013) Available at: Consultative document: Fundamental review of the trading book [pdf] [Accessed 10 April 2019]

Braione M., Scholtes N.K. (2015). Available at: Forecasting Value-at-Risk under Different Distributional Assumptions [pdf] [Accessed 10 May 2019]

Brooks C. (2008). Introductory Econometrics for Finance, New York: Cambridge University Press.

Dávila V. H. L, Cabral C. R. B, Zeller C. B. (2018). Finite Mixture of Skewed Distributions, New York: Springer International Publishing

Du, Z., Escanciano J. C. (2015). Available at: Backtesting Expected Shortfall: Accounting for Tail Risk [pdf] [Accessed 10 May 2019]

Costanzino N., Curran M., (2019). Available at: A Simple Traffic Light Approach to Backtesting Expected Shortfall [pdf] [Accessed 15 May 2019]

Degiannakis S, Potamia A (2016). [pdf] Available at: Multiple-days-ahead value-at-risk and expected shortfall forecasting for stock indices, commodities and exchange rates: Inter-day versus intra-day data [pdf] [Accessed 10 April 2019]

Degiannakis, S., Dent, P., & Floros, C. (2014). Available at: A Monte Carlo simulation approach to forecasting multi-period VaR and ES using the FIGARCH-skT specification [pdf] The Manchester School [Accessed 10 April 2019]

Ebert S. and Strack P. (2015). Available at: Until the bitter end: on prospect theory in a dynamic context [pdf] [Accessed 10 May 2019]

Financial Times (FT) (A) (2012) Available at: Modelling: Normal distribution is not always the norm [Accessed 10 April 2019]

Financial Times (B) (FT) (2012) Available at: JPMorgan loss stokes risk model fears [Accessed 10 March 2019]

Giot, P., & Laurent, S (2003). Available at: Value-at-risk for long and short trading positions [pdf] [Accessed 15 April 2019]

Hull J. C., (2018). Risk Management and Financial Institutions, New Jersey: John Wiley & Sons, Inc.

Hult, H., Lindskog F., Hammarlid O, Rehn C. J. (2012). Risk and Portfolio Analysis: Principles and Methods, New York: Springer.

Kellner, R., & Rösch, D., (2016). Available at: Quantifying market risk with value-at-risk or expected shortfall? Consequences for capital requirements and model risk [pdf] [Accessed 15 April 2019]

Kerkhof J., Melenberg B., (2003). Available at: Backtesting for Risk-based Regulatory Capital [pdf] [Accessed 27 April 2019]

McNeil A. J., Frey R., Embrechts P. (2005). Quantitative Risk Management, New Jersey: Princeton University Press.

Miller M. B. (2019). Quantitative Financial Risk Management, New Jersey: John Wiley & Sons.

Roon de F., Paul Karehnke (2016). Available at: A simple skewed distribution with asset pricing applications [pdf] [Accessed 10 April 2019]

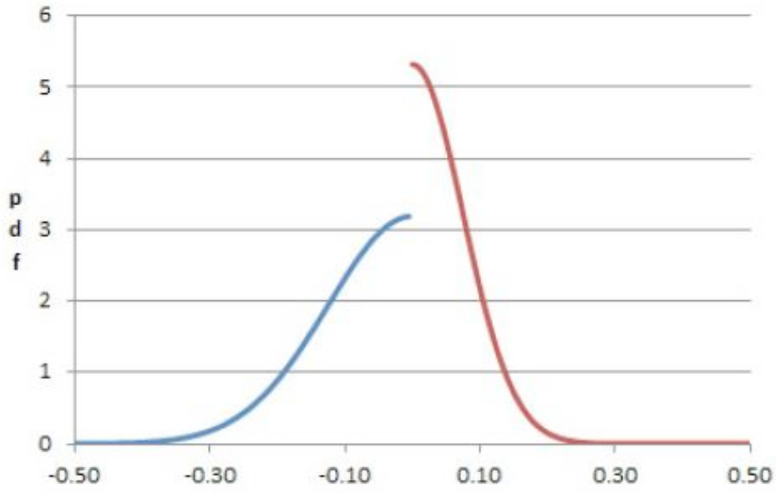
Schneider C., Spalt O. (2015). Available at: Conglomerate investment, skewness, and the CEO long-shot, [pdf] [Accessed 27 May 2019]

Schneider P., Wagner C., Zechner J. (2016) Available at: Low Risk Anomalies? [pdf] [Accessed 27 April 2019]

Thomson Reuters Datastream Application, Version 5.1. [Accessed on various dates]

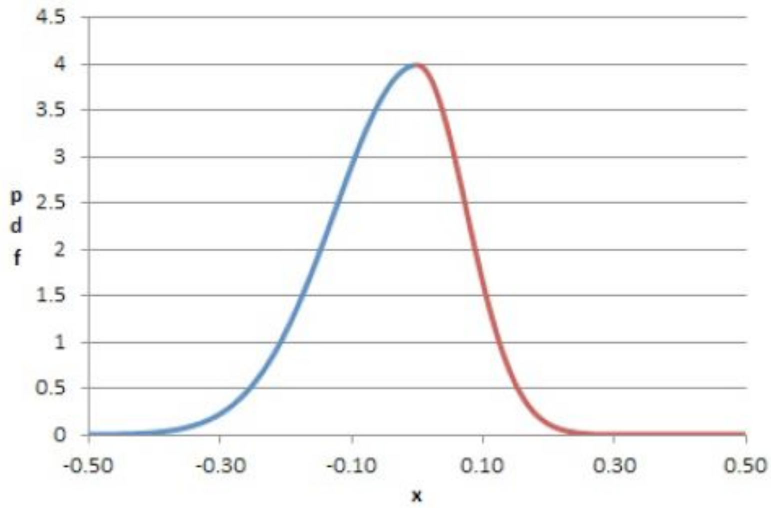
Appendix A

Figure 5



The blue distribution is the left part of N (0; 0:125) and the red distribution is the right part of N (0; 0:075), (de Roon & Karehnke, 2016).

Figure 6



The halves of the normal distribution N (0; 0:125) and N (0; 0:075) are scaled to obtain a continuous density function (de Roon & Karehnke, 2016).

Appendix B

Figure 7

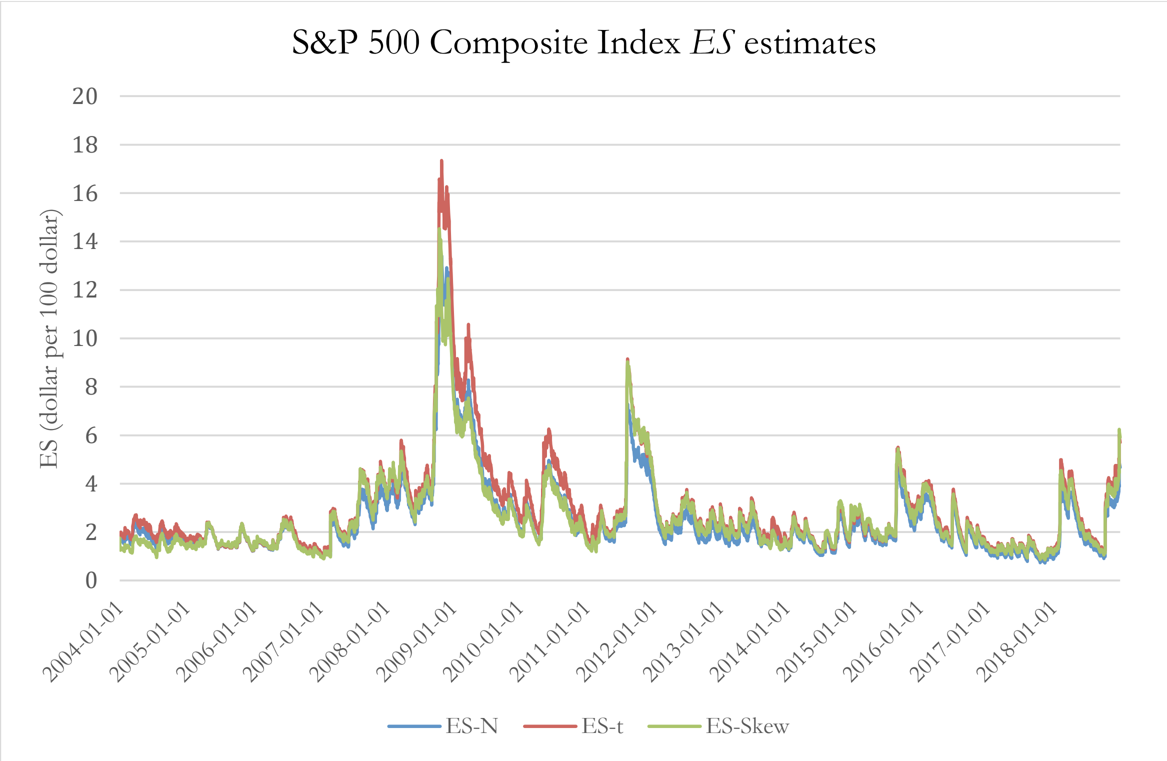


Figure 8

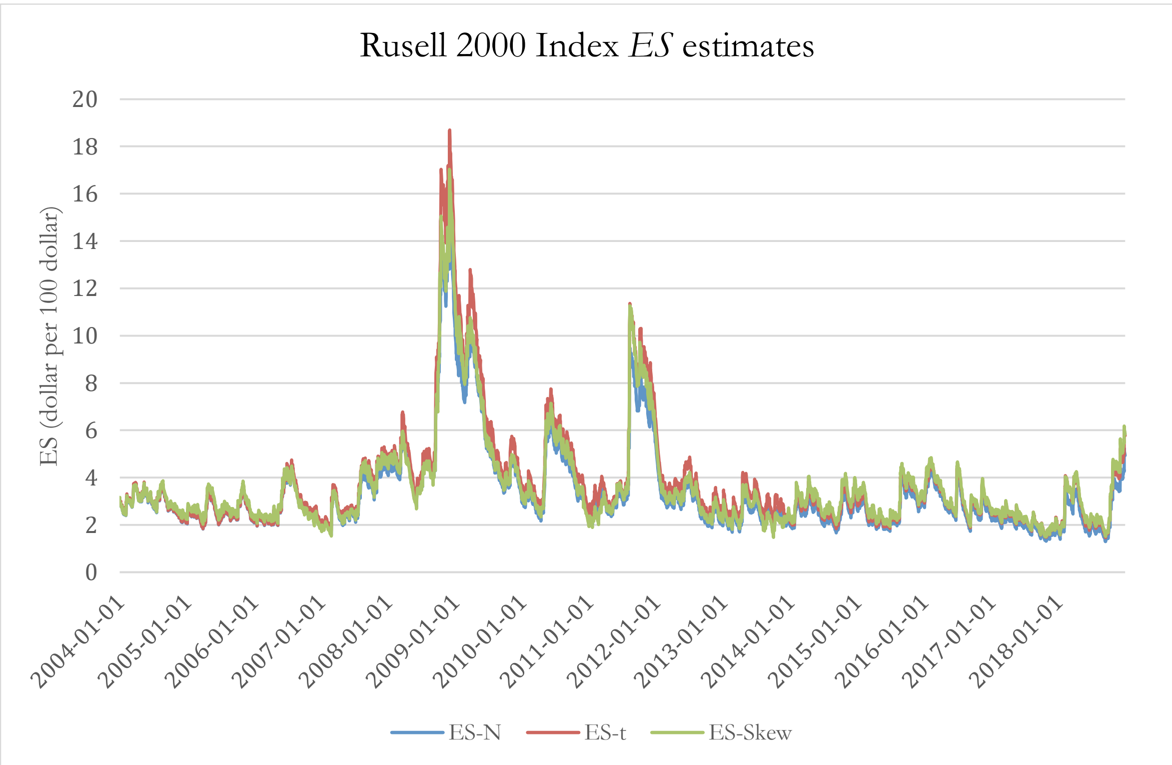


Figure 9

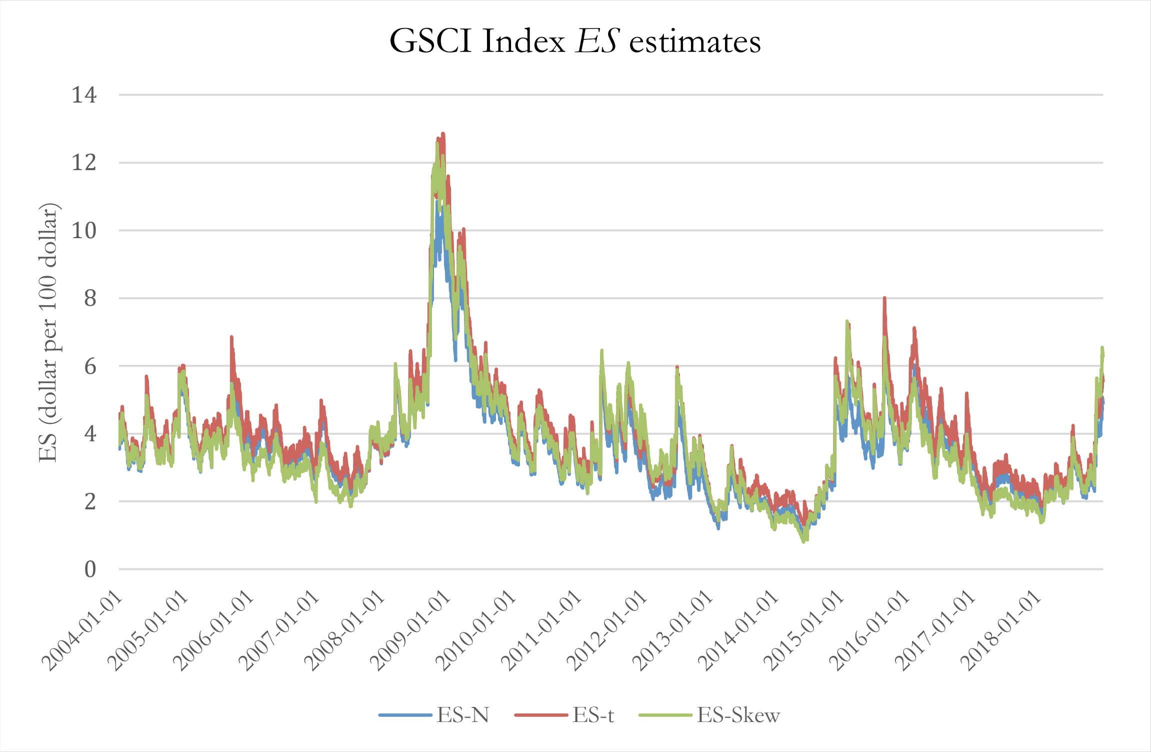


Figure 10