Lund University - Department of economics

Estimating Expected Shortfall Using Parametric and Non-Parametric Approaches

Emil Håkansson
Fredrik Hedman

Supervisor: Birger Nilsson

NEKN02 - Master Thesis

Abstract

With the implementation of the Fundamental Review of the Trading Book in January of 2022, financial institutions will be obligated to implement Expected Shortfall as a means of determining market risk capital. With the transition from Value at Risk to Expected Shortfall, the question of how to accurately forecast Expected Shortfall arises. This paper investigates the forecasting ability of non-parametric and parametric approaches used for estimating Expected Shortfall. More specifically, the paper considers, Basic Historical Simulation, Age-Weighted Historical Simulation, Volatility-Weighted Historical Simulation as well as parametric models based on a Normal distribution, t-distribution and on Extreme Value Theory. As a number of previous studies have investigated the ability of various estimation approaches' ability not to underestimate market risk, the concept of overestimation of risk is introduced. The empirical results indicate that while the conditional Peaks Over Threshold approach yields the most satisfactory results when only underestimation is of a concern, the Volatility-Weighted Historical Simulation most accurately forecasts Expected shortfall when the concept of overestimation is introduced.

Keywords: Value at Risk, Expected Shortfall, Normal distribution, t-distribution, Historical Simulation, Extreme Value Theory, Peaks Over Threshold

Acknowledgements

We would like to express our deepest gratitude to our supervisor Associate Professor Birger Nilsson, without whom this paper would not have been possible. Thank you for your support and advice.

Contents

List of Abbreviations

AWHS - Age-Weighted Historical Simulation

BHS - Basic Historical Simulation

ES - Expected Shortfall

EVT - Extreme Value Theory

EWMA - Exponentially-Weighted Moving Average

GARCH - Generalized AutoRegressive Conditional Heteroskedasticity

GPD - Generalized Pareto Distribution

MLE - Maximum Likelihood Estimation

N-dist - Normal Distribution

POT - Peaks Over Threshold

t-dist - Student's t-Distribution

VaR - Value at Risk

VWHS - Volatility-Weighted Historical Simulation

1 Introduction

In the context of financial institutions, the concept of risk can be divided into two main categories: business- and non-business risk (Hull, 2018). Business risk can be defined as a firm's exposure to strategic risks such as entering new markets and/or producing new products (Hull, 2018). Non-business risk can further be broken down into three categories: credit risk, being the risk that the counterparty fails to meet their financial obligations, operational risk, or the failure of internal systems and controls, and market risk, defined as the risk associated with changes in market conditions (Hull, 2018). The subject of this paper is centered around the concept of market risk.

The 1997 U.S. Securities and Exchange Commission ruling, requiring U.S public companies to disclose information concerning their derivatives trading activities, along with the Basel II Accord of 1999 issued by the Basel Committee on Banking Supervision, led to the widespread usage of Value at Risk (VaR) as a means of measuring market risk (Jorion, 2007). VaR is a measure of the minimum loss ℓ, such that the probability of a loss L larger than ℓ is less than a certain predetermined probability (Jorion, 2007). One of the critiques of VaR however is that it fails to account for large losses that have a very small probability of occurring (Jorion, 2007). The limitations of VaR as a measure of market risk became apparent during the financial crisis of 2008. As the risk measures employed leading up to the financially turbulent year of 2008 failed to adequately reflect market risk, the years following the crisis gave rise to a revision of risk measures (BIS, 2013).

In order to augment the apparent shortcomings of VaR and in order to capture tail-risk events, a new market risk measure known as Expected Shortfall (ES) is currently being implemented. With the implementation of the Fundamental Review of the Trading Book in January of 2022, banks will be obligated to implement ES in order to account for tail-risk events (BIS, 2013). ES is the average VaR for all confidence levels greater than or equal to confidence level α (Hull, 2018). Thus ES takes into account extreme events that have a very low probability of occurring, which VaR fails to account for (Hull, 2018).

The implementation of ES as a risk measure gives rise to the question of how the underlying loss distribution should be modelled and of how ES should be estimated. This paper is concerned with estimating ES using both parametric approaches (which assumes that the loss distribution can be modelled by a probability distribution such as the Normal distribution) and non-parametric approaches (which relies on the empirical distribution). The paper will thus offer a complement to the current contemporary literature on how well, and at which times different approaches to estimating ES perform optimally.

1.1 Purpose

The purpose of the paper is to contribute to the current literature on estimating ES by implementing a series of parametric and non-parametric approaches to estimating ES and to evaluate which model best captures market risk. Three non-parametric approaches are considered: Basic Historical Simulation (BHS), Age-weighted Historical Simulation (AWHS) and Volatility-Weighted Historical Simulation (VWHS). The parametric approaches evaluated are: ES under a Normal distribution and under a Student's t-distribution, using both constant volatility and an Exponentially-Weighted Moving Average (EWMA) model for the two respective distributional assumptions. Furthermore, ES estimates based on Extreme Value Theory (EVT), more specifically using Peaks Over Threshold (POT) and Conditional POT, are evaluated. Each approach is evaluated based on the daily returns of the S&P 500 index from 1962-2019. As ES is a relatively new risk measure, that is currently being implemented, finding a model that produces accurate ES forecasts is of high relevance. Furthermore, while previous research on the topic of ES estimation approaches' ability not to *underestimate* risk has been extensive, research that takes the idea of *overestimation* of risk into account is very scarce. As the paper takes both under- and overestimation of risk into account, the paper will offer a complement to the current literature on estimating ES.

1.2 Delimitation

The main delimitation when attempting to find the optimal method for estimating ES is that time restricts one from implementing a greater number of approaches. Several other approaches, such as modelling the loss distribution using a Stable Paretian distribution or a skewed distribution have for example been carried out in previous research. Furthermore, when carrying out the VWHS, the paper is restricted to forecasting volatility using an EWMA model. Volatility can be modelled in many different ways. Various GARCH models have for example been implemented in previous research. Ideally, the number of investigated approaches would be increased. Nevertheless, the paper considers eleven different approaches which should serve the purpose of this paper well.

One delimitation when implementing EVT is concerned with the degree of subjectivity when deciding on which threshold value to be used (McNeil, 1997). The problem can be formulated as: Which threshold value should be selected such that it is low enough that sufficient data for the estimation of parameters can be collected, and high enough that losses larger than the threshold value are considered "extreme"? As there is no clear answer to this question, a "reasonable" threshold has to be decided on. In this paper the 95-quantile of the previous five years' losses is used. This is however by no means the only approach.

When carrying out the non-parametric approaches (BHS, AWHS and VWHS), a rolling window in used. As with the threshold value in the EVT approach, there is no clear answer as to how long the rolling window should be. For the purpose of this paper, a rolling window of the previous 250 trading days in used. Taking only the last 250 trading days into account should arguably be sufficient in order to have enough data for estimation as well as to avoid using old data that does not reflect current market conditions. If time were not of the essence, many different window-lengths would be used.

1.3 Outline

The first section of the paper provides a theoretical background of the properties and definitions of the risk measures VaR and ES. Furthermore, the section provides an introduction to the various methods implemented in the methodology of the research. A review and discussion of previous research and literature on the topic is presented in the subsequent section. Following the literature review is a section on the methodology used to implement the various estimation approaches. The methods used for conducting the various parametric, non-parametric and EVT approaches are introduced in this section. In the subsequent section, the results from testing the various approaches are presented. The final two sections, Conclusion and Further Research, compare and contrast the various approaches as well as offer suggestions for conducting further research on the topic.

2 Theoretical Background

2.1 Coherency

Before providing an explanation of VaR and ES, a discussion of the desirable properties of a risk measure (R) is warranted. Whether or not a risk measure is coherent is based on if it satisfies four criteria (Artzner, Delbaen, Eber & Heath, 1999), the first being *monotonicity.* In mathematical terms, monotonicity is defined as:

$$L_A \leq L_B \Rightarrow R(L_A) \leq R(L_B) \tag{1}$$

The intuition behind this argument is that if the loss of asset A is less than or equal to asset B in every future state, the risk of asset A should be less than or equal to asset B in order to satisfy the condition of monotonicity (Hull, 2018).

The second criterion of coherency is *subadditivity.* Expressed in mathematical terms the argument states that:

$$R(L_{A+B}) \leq R(L_A) + R(L_B) \tag{2}$$

The intuition behind this property is that the risk measure should encourage diversification, in other words the risk of a portfolio comprised of assets A and B should be less than or equal to the risk of asset A plus the risk of asset B (Hull, 2018).

The third criterion, determining if a risk measure is *positively homogeneous* states that:

$$h > 0 \Rightarrow R(hL) = h \cdot R(L) \tag{3}$$

The equation states that if the portfolio is scaled by a factor h, then the risk of the portfolio should increase/decrease by the scaling factor h (Hull, 2018).

Lastly, the finial criterion for a coherent risk measure concerns whether or not the risk measure is *translation invariant.* This property is satisfied if:

$$R(L - a) = R(L) - a \tag{4}$$

In this equation, a is an amount invested in the risk-free asset. The equation states that if a is invested in the risk-free asset, the loss distribution of the new portfolio should shift to the left and should thus decrease the risk of the new portfolio by the same amount (Hull, 2018).

2.2 Value at Risk

Although the main risk measure discussed in this paper is ES, as VaR is a part of the definition of ES, and the implementation of ES is motivated by the drawbacks of VaR, an explanation of the term is warranted. Mathematically, VaR is defined as:

$$VaR_a(L) = min\{\ell : Pr(L > \ell) \leq 1 - \alpha\} \tag{5}$$

The equation states that the VaR at confidence level a is equal to the smallest loss ℓ, such that the probability of a loss (L) greater than ℓ is less than or equal to 1 - α for some predetermined holding period (usually 1 or 10 days) (Jorion, 2007). Although, VaR has become a prevalent risk measure among financial institutions it does suffer from a major drawback, that limits its suitability as a risk measure. VaR is only a quantile and is completely blind to the size of losses greater than VaR (Jorion, 2007). In other words, if an asset A can take on two future states, a loss of 10 with 99% probability or a loss x with 1% probability, $VaR_{0.975}$ is unaffected by the size of x. This apparent drawback necessitates the implementation of a risk measure that does account for the magnitude of improbable losses.

2.3 Expected Shortfall

Due to the drawbacks of VaR, in particular its failure to account for extreme events, the Basel Committee on Banking Supervision has proposed the implementation of ES as a substitute to VaR (BIS, 2013). ES is defined as:

$$ES_\alpha(L) = \frac{1}{1-\alpha}\int_\alpha^1 VaR_x(L)dx \tag{6}$$

The equation can be interpreted as the average VaR for all confidence levels greater than or equal to a (Hull, 2018). Hence, ES is arguably a superior risk measure as it does satisfy all four criteria of a coherent risk measure as well as takes high and unlikely losses into account (Hull, 2018).

2.4 Maximum Likelihood Estimation

In probability theory, one attempts to carry out mathematical modelling by allowing for randomness. When carrying out inference theory, one essentially moves in the opposite direction, by attempting to draw conclusions about the mathematical model from a set of observations (Anevski, 2017). In other words, one assumes a known distribution with unknown parameters and attempts to answer the question: Given a set of observations, what are the parameter values that would most likely yield the given observations. One method for obtaining estimators is Maximum Likelihood Estimation (MLE). Using MLE one assumes that a set of independently and identically distributed random variables $(x_1, x_2, ..., x_n)$ are distributed according to a density function $f \in \mathcal{F} = \{f_\theta : \theta \in \Theta\}$ (Anevski, 2017). The likelihood function can thus be defined as:

$$G(\theta) = \prod_{i=1}^{n} f_\theta(x_i) \tag{7}$$

As the logarithm is a monotonic transformation, the log-likelihood function can be maximized instead, defining the log-likelihood function by:

$$g(\theta) = \sum_{i=1}^{n} \log f_\theta(x_i) \tag{8}$$

The maximum likelihood estimator of the parameter (θ) can thus be computed by solving:

$$\hat{\theta} = \text{argmax}_{\theta\in\Theta} g(\theta) \tag{9}$$

3 Previous Research

As VaR has been the dominant risk measure and has permeated through risk management departments and academic curricula for several decades, research conducted on estimation approaches of VaR has been extensive. ES has however only recently come into widespread usage. Thus, previous research on optimal approaches to estimating ES is scarcer. The following section is concerned with providing an overview of some of the previous research which has been conducted in the field.

Baran and Witzany (2011) compare and contrast the ES estimates of an Historical Simulation and EVT approach. Baran and Witzany (2011) consider an EVT approach, where a Generalized Pareto Distribution is used to model the tails of the distribution. The EVT approach used by Baran and Witzany (2011) attempts to account for time-varying volatility by implementing an AR(1)-GARCH(1,1) process. The data considered is a hypothetical trading book from 2003 to 2011. The portfolio is constructed using the Euro STOXX 50 index and the PX index, as well as an FX option with the Czech Koruna as underlying asset. Baran and Witzany (2011) find that all three approaches (Historical Simulation, EVT, and EVT-GARCH) manage to provide relatively accurate forecasts, and that using EVT-GARCH produces the most accurate estimates.

Sobreira and Louro (2020) examine a series of parametric-, non-parametric- and EVT methods to estimate VaR and ES using data from stocks traded on the Lisbon stock exchange. Sobreira and Louro (2020) implement a Normal distribution, t-distribution, Generalized Error distribution (GED), skewed Normal, skewed t-distribution and skewed GED as parametric distributions and HS, RiskMetrics and EVT as non-parametric methods, where different GARCH models are used as volatility estimates. In order to evaluate the VaR and ES estimates Sobreira and Louro (2020) conduct several backtesting procedures, including "Testing ES Directly" by Acerbi and Szekely. The results show that an EVT model with asymmetric GARCH as volatility is the best performing model, Sobreira and Louro (2020) also suggest that larger sample sizes provide better results.

Harmantzis, Linyan and Yifan (2006) investigate the performance of different models in measuring VaR and ES. The authors are primarily concerned with investigating if models that capture rare events more accurately predict risk than models that do not. The data used is the daily returns of six major stock indices (S&P500, DAX, CAC, Nikkei, TSE and FTSE) and four currency pairs (USD/EUR, USD/JPY, USD/GBP and USD/CAD) for a ten-year period. The daily returns are modelled using the Empirical distribution, Gaussian distribution, Peaks Over Threshold, and Stable Paretian distribution. The results of Harmantzis, Linyan and Yifan (2006) suggest that models that capture rare events more accurately predict risk. Furthermore, Harmantzis, Linyan and Yifan (2006) provide empirical evidence that POT and the Empirical distribution give accurate estimations, while the Gaussian model tends to underestimate ES.

Similar research has also been conducted by Marinelli, D'Addona and Rachev (2007), who investigate the predictive accuracy of VaR and ES models based on the assumption of Stable Paretian returns, Gaussian returns and on EVT. Marinelli, D'Addona and Rachev (2007) consider the daily returns of two stock indices (S&P500 and NASDAQ) and two stocks (Amazon and Microsoft). Marinelli, D'Addona and Rachev (2007) further suggest that stable models tend to outperform the models based on EVT when estimating VaR and that POT methods tend to give more accurate forecasts when estimating ES.

As with the aforementioned papers, Jadhav, Ramanathan and Naik-Nimbalkar (2009) consider both parametric and non-parametric approaches to estimating ES. The parametric approaches include: a Gaussian approach, EVT approach as well as a Stable Paretian approach. Jadhav, Ramanathan and Naik-Nimbalkar (2009) also consider the non-parametric, Historical method. The methods are carried out on the daily returns of the Indian BSE and NSE as well as on the NYSE and LSE. The authors find that the Historical method produces more accurate forecasts of risk than the parametric methods. Furthermore, Jadhav, Ramanathan and Naik-Nimbalkar (2009) is one of few articles that makes reference to overestimation of ES. The authors argue that due to the presence of outliers in the data, the Historical method often leads to overestimation. In order to augment the Historical method, the authors propose a new non-parametric method which attempts to get rid of outliers in the data. Empirical evidence is provided for the fact the proposed new non-parametric method does not underestimate ES and that it solves the problem of overestimation (Jadhav, Ramanathan & Naik-Nimbalkar, 2009).

LU and Huang (2007) compare three different models for estimating VaR. The first model, referred to by LU and Huang (2007) as "standard-EWMA" is derived from JPMorgan's RiskMetrics and uses a Gaussian distribution with EWMA forecasting. The second method evaluated is the "robust-EWMA" procedure proposed by Guermat and Harris (2002). The procedure proposed by Guermat and Harris (2002) attempts to augment the standard-EWMA approach used in RiskMetrics by modelling the loss distribution using a Laplace distribution. Furthermore, LU and Huang (2007) attempt to take both skewness and heavy tails of financial data into account by deriving a third method that they refer to as "skewed-EWMA". Instead of using a regular Laplace distribution as used in the research conducted by Guermat and Harris (2002), LU and Huang (2007) implement an asymmetric Laplace distribution. Through backtesting the three proceedures, LU and Huang (2007) find that the skewed-EWMA approach produces more accurate forecasts than the standard-EWMA and robust-EWMA approaches.

In Bredin and Hyde (2004), ES is not considered, but the performance of two different VaR methods are evaluated on the foreign exchange market (the Irish Punt against six other currencies). The first model evaluated is based on an orthogonal GARCH approach and the later on an EWMA approach. The empirical evidence gathered by Bredin and Hyde (2004) suggests that the orthogonal GARCH approach is more accurate but that the EWMA approach is more conservative.

Degiannakis and Potamia (2017) attempt to provide VaR and ES forecasts using a AR(1)-GARCH(1,1) and AR(1)-HAR-RV-skT approach. The approaches are carried out on various markets including: equity-, commodity- and foreign exchange markets. Using a 95%, 97.5% and 99% confidence interval, the empirical results provided by Degiannakis and Potamia (2017) suggest that the AR(1)-GARCH(1,1) yields the most accurate forecasts. Furthermore, Degiannakis and Potamia (2017) suggest implementing a 97.5% confidence interval, in line with the Basel Committees proposal to replace 99% VaR with 97.5% ES.

A test of the parametric approaches to estimating ES is also indirectly a test of the normality of returns. Mandelbrot (1963) and Fama (1965) reject the idea that stock returns follow a Normal distribution. Mandelbrot (1963) and Fama (1965) suggest instead that stock returns have leptokurtic properties, that is, the tails of the distribution are fatter. This would suggest that ES models that implement a Student's t-distribution should provide more accurate estimations of ES than their normally distributed counterparts.

4 Methodology

4.1 Data Description and Data Processing

In order to evaluate the ES estimation approaches, the adjusted closing prices of the S&P 500 are used. The adjusted closing price is the closing price of the index, adjusted for both dividends and stock splits. The adjusted closing price of the S&P500 from 1957 to 2019 are collected from the Bloomberg database. The reason for including 1957-1961 in the collected data is that, even though only the period 1962 to 2019 is evaluated, the five years leading up to 1962 are used as sample data in the EVT approach.

With the adjusted closing prices collected, the loss scenarios for all trading days between 1957 and 2019 are calculated. The formula for calculating the loss scenario is expressed as:

$$L_t = -K \frac{P_t - P_{t-1}}{P_{t-1}} \tag{10}$$

Where L_t is the loss scenario at time t, P_t is the adjusted closing price at time t, P_{t-1} is the adjusted closing price at time $t-1$ and K is the invested capital. The value of K is arbitrary and does not affect the results of the backtests of the various approaches. A value of 1000 is used for K throughout the paper. With the inclusion of a minus sign on the right-hand side of the equation, a loss (gain) is characterized by a positive (negative) value of L.

4.2 Holding Period and Confidence Level

When estimating VaR and ES, both the holding period (h) and the confidence level (α) affect the estimate (Hull, 2018). The holding period is defined as the total number of days that the losses are measured over. In order to get as many loss observations as possible, the holding period is set to one day for the remainder of this paper. As for the confidence level (α), 97.5% is used in accordance with BIS (2013). The confidence level is defined as the probability of a loss exceeding VaR occurring over the course of the holding period being equal to $1-\alpha$ (Hull, 2018).

4.3 Non-Parametric Approaches

The approaches used in this paper for estimating VaR and ES can be divided into parametric and non-parametric approaches. What separates the two approaches is the way in which assumptions are made about the underlying loss distribution (Geisser, Johnson & Wiley InterScience, 2006). The parametric approaches make assumptions about the underlying distributional and attempt to model the observed losses accordingly. As far as non-parametric approaches are concerned, no assumptions about the underlying distributional have to be made. Instead, the empirical loss distribution is used. The non-parametric approaches investigated in this paper are BHS, AWHS and VWHS.

4.3.1 Basic Historical Simulation

The most straight-forward approach of the investigated non-parametric approaches is BHS. This approach uses the observed losses from the previous year to estimate VaR and ES (Hull, 2018). Using the previous year's loss observations directly, the ES for a particular day can be estimated using *Equation 6*. Theoretically, any length of sample data can be used for BHS. However, in order to account for changing market conditions, and due to the fact that old loss observations are deemed less likely to reflect current conditions accurately, only the last year's loss observations are used as sample data in this paper. Furthermore, for the BHS as well as for the other approaches, an α of 0.975 is used.

4.3.2 Age-Weighted Historical Simulation

AWHS is analogous to BHS in the sense that it relies on the empirical loss distribution. The way in which it deviates from regular BHS is that it attempts to take current market conditions into account by placing a larger weight on more recent observations. In the aforementioned BHS, all historical losses are assigned the same weight ($1/N$). When implementing AWHS, older loss observations are assigned a lower weight. The weights are assigned accordingly (newest to oldest) (Richardson, Boudoukh & Whitelaw, 1997):

$$
\begin{aligned}
\omega_N &= \tfrac{1-\lambda}{1-\lambda^N} \\
\omega_{N-1} &= \lambda\omega_N \\
\omega_{N-2} &= \lambda^2\omega_N \\
&\vdots \\
\omega_2 &= \lambda^{N-2}\omega_N \\
\omega_1 &= \lambda^{N-1}\omega_N
\end{aligned}
$$

When determining the weight of each observed loss, the term λ is used as a decay factor (Richardson, Boudoukh & Whitelaw, 1997). A value of λ is selected between 0 and 1 depending on how much importance older observations are deemed to have (Richardson, Boudoukh & Whitelaw, 1997). Throughout this paper a decay rate of $\lambda = 0.99$ is used and the previous year's loss observations are taken into account. This implies that the weight is halved approximately every 70 days (the loss at $T = t - 70$ is given approximately half the weight of the loss at time $T = t$). The advantage of using AWHS as opposed to BHS is that current market conditions are taken into account to a greater extent. Furthermore, using AWHS, ghost effects, that is a sudden large shift in VaR and ES estimates as a result of a large loss observation falling out off sample, are reduced (Dowd, 2005).

4.3.3 Volatility-Weighted Historical Simulation

With the empirical finding that volatility tends to cluster into times of high volatility and times of lower volatility (Mandelbrot, 1963) arises the need to create an approach for estimating ES that takes this phenomena into account. One such method is VWHS as suggested by Hull and White (1999). The rationale behind using VWHS is that if volatility in the current holding period is below average, then volatility is likely to be below average in the next holding period as well (Hull & White, 1999). Thus, if volatility in the current holding period is above (bellow) average, the estimates for VaR and ES are adjusted upwards (downwards). As in the case of BHS, the losses $(\ell_1, \ell_2, \ell_3, ..., \ell_T)$ of the previous year are used. The loss observations are rescaled accordingly:

$$
\begin{aligned}
\ell_1^* &= \tfrac{\sigma_{T+1}}{\sigma_1}\ell_1 \\
\ell_2^* &= \tfrac{\sigma_{T+1}}{\sigma_2}\ell_2 \\
&\vdots \\
\ell_{T-1}^* &= \tfrac{\sigma_{T+1}}{\sigma_{T-1}}\ell_{T-1} \\
\ell_T^* &= \tfrac{\sigma_{T+1}}{\sigma_T}\ell_T
\end{aligned}
$$

In the above expression, $\sigma_1, \sigma_2, ..., \sigma_{T-1}, \sigma_T$ are the volatilities associated with each respective loss observation. Furthermore, the term σ_{T+1} is a forecast of the next holding period's volatility. As the next period's volatility is not directly observable, an estimation of σ_{T+1} is required. For this purpose, an EWMA model is implemented accordingly:

$$
\sigma_{T+1}^2 = \frac{1-\lambda}{1-\lambda^T} \sum_{t=1}^{T} \lambda^{T-t} \varepsilon_t^2 \tag{11}
$$

When dealing with large values of T, the expression can be rewritten as:

$$
\sigma_{T+1}^2 \approx (1-\lambda)\varepsilon_T^2 + \lambda\sigma_T^2 \tag{12}
$$

Where ε_T is the unexpected return at time T, σ_T is the volatility at time T and λ is a fixed constant. The value of λ can be changed in order to alter the effect of deviating volatilities. However, for the purpose of this paper, and in line with Longerstaey and Spencer (1996), a fixed value of λ of 0.94 is used.

4.4 Parametric Approaches

While non-parametric approaches to estimating ES make use of the empirical loss distribution, parametric approaches instead assume that the loss distribution can be modelled by a probability distribution such as the Normal- or t-distribution (Geisser, Johnson & Wiley InterScience, 2006). One advantage of using a parametric approach to estimate ES is that while non-parametric approaches to a large extent are dependent on the largest loss in the sample period, parametric approaches are not and are thus more suitable for managing extreme values. The parametric models in this paper are based on the Normal- and Student's t-distribution.

4.4.1 Normal Distribution

The first distributional assumption made is that of a Normal distribution. A Normal distribution is expressed by the probability density function:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \text{for } x \in (-\infty, \infty) \tag{13}$$

Where μ and σ are the mean and standard deviation respectively (Anevski, 2017). Assuming that losses follow a normal distribution ($L \sim \Phi(\mu, \sigma)$) and due to the fact that the Normal distribution is a continuous distribution, the following definition of VaR is used to make the estimation:

$$\Pr(L_t > \text{VaR}_\alpha) = 1 - \alpha \tag{14}$$

The equation can be interpreted as, the probability of a loss larger than the VaR estimate occurring, should be equal to $1-\alpha$ (Hull, 2018). As the Normal distribution is a continuous distribution, a value for VaR can always be found. Thus, at the 0.975 confidence interval, the VaR estimate can be expressed by:

$$VaR_{0.975}(L) = \mu + \sigma z_{0.975} \tag{15}$$

Where $z_{0.975}$ is the 0.975-quantile for the Normal distribution (Norton, Khokhlov & Uryasev, 2018). Similarly, as the Normal distribution is a continuous distribution the following definition of ES can be implemented in order to derive the necessary estimates:

$$ES_\alpha(L) = E\left[L : L > VaR_\alpha(L)\right] \tag{16}$$

The intuition behind the equation is that the ES estimate should be equal to the expected loss, given that the occurred loss is larger than the VaR estimate. Given *Equation 16*, and assuming that losses follow a Normal distribution ($L \sim \Phi(\mu, \sigma)$), the estimate for ES at the 0.975 confidence interval is given by:

$$ES_{0.975}(L) = \mu + \sigma \frac{f_{std}(z_{0.975})}{1 - 0.975} \tag{17}$$

Where f_{std} is the probability density function of a $\Phi(0,1)$-distributed normal variable (Norton, Khokhlov & Uryasev, 2018). The final problem at hand when implementing the parametric approach is how to estimate the mean and volatility parameters. For simplicity and due to the fact that the holding period is relatively short, a μ (mean) of zero is used in this paper. As for the volatility estimates, two different approaches are implemented. The most straightforward approach is to use the sample variance of the previous year's loss observations. However, this approach arguably suffers from the disadvantage that it fails to adequately account for current market conditions (Hull, 2018). In order to account for changing market conditions, a model that takes time-varying volatility into account is also implemented.

4.4.2 EWMA-N

The documented phenomena of volatility clustering gives rise to the need for a model that is able to account for time-varying volatility (Mandelbrot, 1963). In the previous parametric model, the sample variance is used in order to derive an estimate for ES. The EWMA-N-approach uses an EWMA model to estimate volatility for the next holding period (σ_{t+1}). Thus, the same approach as in the aforementioned N-dist-approach is implemented, with the distinction that the sample variance is substituted for the EWMA-estimate for volatility ($\sigma^2_{T+1} \approx (1-\lambda)\varepsilon^2_T + \lambda\sigma^2_T$). The models for estimating VaR and ES at the 0.975 confidence level can thus be stated:

$$VaR_{0.975}(L) = \mu + \sigma_{t+1} z_{0.975} \tag{18}$$

$$ES_{0.975}(L) = \mu + \sigma_{t+1} \frac{f_{std}(z_{0.975})}{1 - 0.975} \tag{19}$$

4.4.3 Student's t-distribution

One liability of modeling ES according to a Normal distribution is the empirical finding that stock returns typically do not follow a Normal distribution (Fama, 1965). The finding that financial returns are characterized by excess kurtosis, that is, the distribution is heavy-tailed in comparison to the Normal distribution, gives rise to the need to model ES according to a distribution that accounts for "fat tails" (Hull, 2018). By introducing the Student's t-distribution a parameter v (degrees of freedom) is implemented in order to control for excess kurtosis. In this regard, the Student's t-distribution differs from the regular Normal distribution, giving rise to the following probability density function:

$$f(x) = \frac{\Gamma\left[(v+1)/2\right]}{\sigma\sqrt{(v-2)\pi}\Gamma(v/2)}\left[1+\frac{1}{v-2}\left(\frac{x-\mu}{\sigma}\right)^2\right]^{-(v+1)/2} \text{for } x \in (-\infty, \infty) \tag{20}$$

Where v is the degrees of freedom, σ is volatility, μ is the mean and Γ is the Gamma-function (Ahsanullah, Shakil & Golam Kibria, 2014). The parameter estimates for μ and σ can be derived with the same method as implemented for the Normal distribution. As for the degrees of freedom (v), the following relationship between sample kurtosis and degrees of freedom is utilized:

$$k = 3 + \frac{6}{v-4} \Leftrightarrow v = \frac{4k-6}{k-3} \tag{21}$$

Where k is the sample kurtosis and v the degrees of freedom (Ahsanullah, Shakil & Golam Kibria, 2014). As for the for the estimation of VaR, *Equation 14* is again implemented, yielding an expression for the VaR estimate at the 0.975 confidence level of:

$$VaR_{0.975}(L) = \mu + \sqrt{\frac{v-2}{v}}\sigma t_{0.975,v} \tag{22}$$

Where $t_{0.975,v}$ is the 0.975-quantile of the standard loss distribution (Norton, Khokhlov & Uryasev, 2018). The ES estimate at the 0.975 confidence level when losses follow a Student's t-distribution is expressed as suggested by Norton, Khokhlov and Uryasev (2018):

$$ES_{0.975}(L) = \mu + \sqrt{\frac{v-2}{v}}\sigma\frac{f^*_{std}(t_{0.975,v})}{1-0.975}\left(\frac{v+t^2_{0.975,v}}{v-1}\right) \tag{23}$$

As in the case of the Normal distribution approach, parameter estimates of μ and σ have to be obtained. As in the previous example, μ (mean) is assumed to be zero. As for σ (volatility), the sample variance is used. However, the aforementioned empirical finding that volatility tends to cluster into times of high volatility and of low volatility, warrants the implementation of a model that takes time-varying volatility into account.

4.4.4 EWMA-t

In order to account for time-varying volatility and to estimate volatility for the next holding period (σ_{t+1}) an EWMA model is implemented. In the previous parametric model, the sample variance is used in order to derive an estimate for ES. The same approach as in the aforementioned Student's t-distribution approach is implemented, with the exception that the sample variance is substituted for the EWMA estimate for volatility ($\sigma^2_{T+1} \approx (1-\lambda)\varepsilon^2_T + \lambda\sigma^2_T$). The models for estimating VaR and ES at the 0.975 confidence level can thus be stated:

$$VaR_{0.975}(L) = \mu + \sqrt{\frac{v-2}{v}}\sigma_{t+1}t_{0.975,v} \tag{24}$$

$$ES_{0.975}(L) = \mu + \sqrt{\frac{v-2}{v}}\sigma_{t+1}\frac{f^*_{std}(t_{0.975,v})}{1-0.975}\left(\frac{v+t^2_{0.975,v}}{v-1}\right) \tag{25}$$

4.5 Extreme Value Theory

EVT is a method used for modelling the tail of a distribution (Hull, 2018). Gnedenko (1943) proves the fact that the tails of many different probability distributions share common features. In the context of financial economics, this finding can be used to model a loss distribution and hence estimate VaR and ES using the empirical distribution. The two extreme value approaches investigated in this paper are POT and Conditional POT.

4.5.1 Peaks over Threshold

A critique of regular EVT is that it only uses the largest loss in the sample for the analysis of VaR and ES (Hull, 2018). POT circumvents this issue, by investigating all losses that are larger than a predetermined threshold value. Thus, the method is not as susceptible to information loss as regular EVT. However, the problem that arises when implementing POT is the need to define a threshold value such that all losses above are deemed "extreme". The problem with this rather subjective threshold value is that while the threshold value should be sufficiently high in order to consider excess losses "extreme", the threshold value also needs to be low enough in order to gather sufficient data for the estimation of parameters (Hull, 2018). As a result, the decision of which threshold value to be be implemented is not entirely objective. For the purpose of this paper, the 95$^{\text{th}}$ quantile of the previous five years' loss observations is used as a threshold value. This is arguably a value that is low enough in order to estimate parameters and high enough to deem losses in excess as "extreme".

The losses in excess of the predetermined threshold value $(L - u)$ are modelled. In doing so, one assumes that L follows a stochastic process and that it can be defined by the cumulative density function F (that is, $Pr(L \leq \ell) = F(\ell)$) (Hull, 2018). Through the definition of a conditional probability, a definition of the cumulative density function for excess losses can be derived (Hull, 2018):

$$F_u(\ell) = \frac{F(\ell + u) - F(u)}{1 - F(u)} \tag{26}$$

The corollary states that:

$$F_u(\ell - u) = \frac{F(\ell) - F(u)}{1 - F(u)} \tag{27}$$

By definition, $F(\ell) = \Pr(L \leq \ell)$. One can derive an estimate of VaR by defining $\ell = VaR_\alpha$ and solving the equation $F(VaR_\alpha) = \alpha$ for VaR_α. However, when solving this equation, the fact that $F_u(\ell - u)$ follows an unknown distribution becomes apparent. This issue can however be circumvented through the use of the Pickands-Balkema-de Haan Extreme Value Theorem (Balkema & de Haan, 1974). The theorem states that as the threshold value u tends towards infinity, the tail to the right of u can be approximated by a GPD:

$$G(\ell - u) = \begin{cases} 1 - (1 + \xi \frac{\ell - u}{\beta})^{-1/\xi}, & \text{if } \xi \neq 0 \\ 1 - \exp\left(-\frac{\ell - u}{\beta}\right), & \text{if } \xi = 0 \end{cases}$$

Hence, by selecting a large enough u, the unknown distribution F_u can be substituted for

G. Thus, through the use of the Pickands-Balkema-de Haan Extreme Value Theorem, one can derive an expression for VaR:

$$VaR_\alpha = \begin{cases} u + \frac{\beta}{\xi}\left[\left(\frac{1-\alpha}{1-F(u)}\right)^{-\xi} - 1\right], & \text{when } \xi \neq 0 \\ u - \beta \ln \frac{1-\alpha}{1-F(u)}, & \text{when } \xi = 0 \end{cases}$$

The parameter values β and ξ are estimated using MLE. Through the definitions of VaR and the definition of ES, an expression for ES is stated as (Hull, 2018):

$$ES_\alpha = \begin{cases} \frac{VaR_\alpha + \beta - u\xi}{1-\xi}, & \text{for } \xi \neq 0 \\ VaR_\alpha + \beta, & \text{for } \xi = 0 \end{cases}$$

4.5.2 Conditional Peaks over Threshold

In the case of the aforementioned parametric approaches, if one attempts to take current market conditions into account by placing more emphasis on more recent loss observations an approach such as AWHS can be used. The corollary in the case of POT is Conditional POT. This is implemented through carrying out the standard Peaks over Threshold analysis but using GARCH/EWMA volatility models (McNeil & Frey, 2000). In this paper an EWMA volatility model is used. Using Unconditional POT, the estimate of VaR can be defined as:

$$VaR_\alpha = \mu + \sigma_{T+1} q_\alpha \tag{28}$$

Where σ_{T+1} is the volatility derived from the EWMA model for the first day outside the sample period, and q_α is the VaR estimate derived from the regular POT model. In order to carry out the analysis, the loss observations are standardised accordingly:

$$\begin{aligned} \varepsilon_1^* &= \frac{\ell_1 - \bar{\ell}}{\sigma_1} \\ \varepsilon_2^* &= \frac{\ell_2 - \bar{\ell}}{\sigma_2} \\ &\vdots \\ \varepsilon_{T-1}^* &= \frac{\ell_{T-1} - \bar{\ell}}{\sigma_{T-1}} \\ \varepsilon_T^* &= \frac{\ell_T - \bar{\ell}}{\sigma_T} \end{aligned}$$

Where ℓ is the average of the sample loss observation and σ_t is the volatility associated with each respective loss. The estimates for VaR and ES under Conditional POT can thus be expressed by:

$$VaR_\alpha = \mu + \sigma_{T+1} VaR(\varepsilon^*) \tag{29}$$

$$ES_\alpha = \mu + \sigma_{T+1} ES(\varepsilon^*) \tag{30}$$

Where $VaR(\varepsilon^*)$ and $ES(\varepsilon^*)$ are the VaR and ES estimates from the standardized residuals under a GPD.

4.6 Backtesting Expected Shortfall

In order to evaluate the performance of each approach, a backtest outlined by Acerbi and Szekely (2014) is carried out. Acerbi and Szekely (2014) propose three different backtests in their 2014 paper. The test implemented in this paper is the second test, referred to by Acerbi and Szekely (2014) as "testing ES directly". Acerbi and Szekely (2014) assume that each day's profit or loss (L_t where $t = 1, 2, ..., T$) follows what they refer to as a *real* distribution F_t and is forecasted by a *predictive* distribution P_t. Assuming that the distribution is continuous and strictly increasing, and following the aforementioned definitions of VaR and ES, the ES at time t can be expressed by the following equation:

$$ES_{\alpha,t} = \mathbf{E}[L_t | L_t + VaR_{\alpha,t} < 0] \tag{31}$$

The test thus follows from the unconditional expectation:

$$ES_{a,t} = \mathbf{E}\left[\frac{L_t I_t}{1-\alpha}\right] \tag{32}$$

Where I_t is an indicator function defined as:

$$I_t = \begin{cases} 1, & \text{if } L_t > VaR_{a,t}(L_t) \\ 0, & \text{if } L_t \leq VaR_{a,t}(L_t) \end{cases}$$

The test statistic can thus be defined as:

$$Z = -\sum_{t=1}^{T} \frac{L_t I_t}{T(1-\alpha)ES_{\alpha,t}} + 1 \tag{33}$$

The null- and alternative hypothesis state that:

$$\begin{aligned} H_0 &: ES^P_{\alpha,t} = ES^F_{\alpha,t} \text{ for all } t \\ H_1 &: ES^P_{\alpha,t} < ES^F_{\alpha,t} \text{ for at least one } t \end{aligned} \tag{34}$$

The intuition behind the test is that the null hypothesis states that the model correctly estimates ES for all days t and that the alternative hypothesis states that the selected model underestimates ES at least one day t (Acerbi & Szekely, 2014). In layman's terms this means that the test-statistic calculates the average ratio between the loss when the observed loss exceeds the VaR estimate for that particular day, and the ES estimate for that particular day (Acerbi & Szekely, 2014). Thus, if the model is correctly specified, the expected value of the test-statistic is zero under the null hypothesis (Acerbi & Szekely, 2014).

When evaluating the test-statistic from the Acerbi and Szekely test, the natural question is: What values of the test-statistic constitute over- and underestimation of Expected Shortfall? As one does not know the underlying distribution of the test-statistic, it is not directly evident which critical value to use. Acerbi and Szekely (2014) argue that the critical value does not deviate much between different predictive distributions. Furthermore, Acerbi and Szekely (2014) suggest using -0.70 as a critical value at the 5% level and -1.8 at the 1% level. On the other end of the spectrum, overestimation of ES will arguably lead to an inefficient use of capital. Thus, models that systematically overestimate ES need to be penalized accordingly. As the distribution of the test-statistic is not symmetric, using a critical value of +0.70 would not be prudent. Andersson (2020) suggests using a critical value of +0.59.

Another approach that can be used to evaluate ES estimates is a Basel type "Traffic Light System". This approach was first implemented by the Basel Committee on Banking Supervision in 1996 in order to evaluate VaR estimates. Costanzino and Curran (2018) argue that this approach can also be used to evaluate ES estimates. The approach consists of a confidence interval marked with critical values. A model is given a green, yellow or red light depending on how accurate its ES estimate is (Costanzino & Curran, 2018). A green light, indicating that the model is "accurate", is given when the test statistic of the ES estimate is below the 95^{th} quantile (Costanzino & Curran, 2018). A yellow light, indicating that the model underestimates ES, is given when the test statistic falls between the 95^{th} and 99^{th} quantile (Costanzino & Curran, 2018). A red light, indicating a gross underestimation of the ES estimate, is given when the test statistic lays above the 99^{th} quantile (Costanzino & Curran, 2018). To implement this approach in accordance with Acerbi and Szekely (2014) the 95^{th} and 99^{th} quantiles are equal to -0.7 and -1.8 respectively.

The evaluation of each of the eleven approaches can thus be interpreted as two different tests, based on the underlying objective of the forecasting approach. The objective of the financial institution can be expressed as optimizing the amount of reserve capital in order to have enough capital in the books to weather through financially turbulent years while at the same time minimizing an inefficient use of capital by having too high capital reserves. However, from the perspective of regulators such as the Basel Committee on Banking Supervision, overestimation of ES is of little concern. Therefore, if the objective is simply to apply estimation approaches that do not underestimate risk, the Basel type "Traffic Light System" offers an indication as to which model best captures risk.

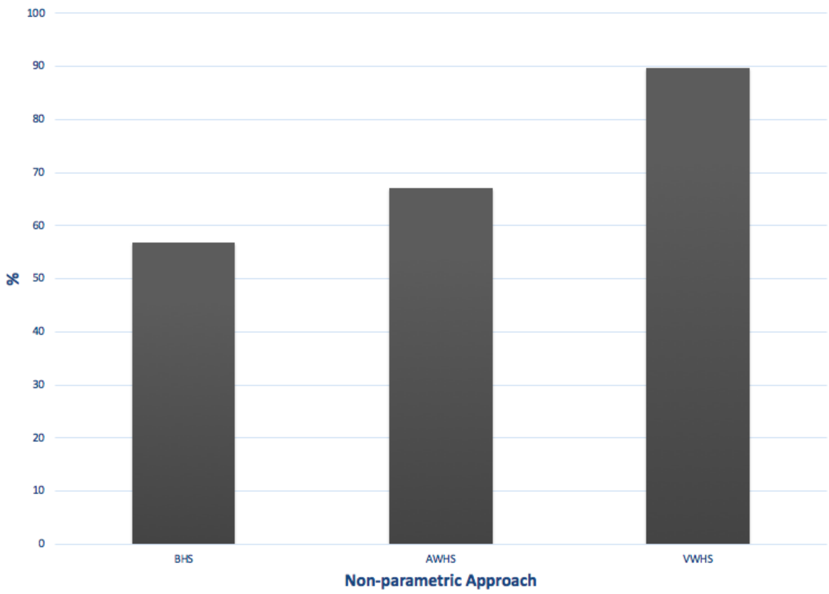
5 Results and Analysis

Table 1: The percentage of over- and underestimations for each of the eleven approaches (1962-2019)

Approach	Underestimations (%)	Overestimations (%)	Correct estimation (%)
BHS	32.8	10.3	56.9
AWHS	10.3	22.4	67.2
VWHS	10.3	0.0	89.7
N-dist	39.7	31.0	29.3
t-dist	31.0	34.5	34.5
N-dist-EWMA	36.2	0.0	63.8
t-dist-EWMA	20.7	0.0	79.3
POT ($\xi \neq 0$)	44.8	24.1	31.0
POT ($\xi = 0$)	44.8	24.1	31.0
Conditional POT ($\xi \neq 0$)	1.7	31.0	67.2
Conditional POT ($\xi = 0$)	3.4	27.6	69.0

5.1 Backtest of Non-Parametric Approaches

Figure 1: The percentage of correct ES estimations for the non-parametric approaches: Basic Historical Simulation, Age-Weighted Historical Simulation and Volatility-Weighted Historical Simulation for the years 1962-2019



Examining the 58 investigated years (1962-2019) the BHS results in a total of 19 underestimations of ES. The most severe underestimations occurred, perhaps unsurprisingly, during the years 1973, 1987 and 2008. As the oil embargo of 1973, the stock market crash of 1987 and the recession of 2008 triggered a sharp rise in volatility, the BHS model failed to account for this sudden rise in volatility. Even though the BHS model only takes into account the latest year's loss observations it nevertheless evidently fails in adapting quickly enough to prevailing market conditions.

On the other end of the spectrum is the desire not to overestimate ES. Investigating the six years when overestimation of ES occurred, further evidence for the inability of BHS to adapt to current market is presented. Three out of the six overestimations occurred in the years following economic crises (1975, 1988 and 2009). The data thus suggests that the practical implications for a financial institution implementing BHS is that too much capital is kept in the books during the years following a financial crisis.

Additionally, while an examination of whether or not a model under- or overestimates ES during a particular year is necessary, the degree to which it underestimates (overestimates) ES given that underestimation (overestimation) has occurred is also worth examining. An examination of *Table 2* (B Appendix) suggests that while BHS underestimates ES in 32.8% of the investigated years, it also severely underestimates ES in the three most turbulent years (1973, 1987 and 2008). The results form the Traffic Light System, presented in *Table 3* of the B Appendix, display that the BHS results in twelve and seven yellow and red lights respectively. In other words, BHS severely underestimates risk in seven out of the 58 investigated years. Thus from the perspective of regulatory authorities, the data suggests that BHS is not a preferable approach to evaluating risk. BHS is quite intuitive and relatively straightforward to implement, but due to the fact that it places the same weight on all loss observations that took place in the most recent year, it is slow to react to current market conditions. Thus, the findings suggest that using BHS will likely lead to too little capital being put aside during turbulent years, and too much capital in the books during recovery periods.

Moving up in model complexity, and addressing the apparent shortcomings of BHS is AWHS. As with the aforementioned BHS, AWHS also fails to accurately predict the turbulent years of 1973, 1987 and 2008. The model does however display an improvement, as the z-statistics from the Acerbi and Szekely test for the three years is closer to the critical value of -0.70. Furthermore, as the dynamic properties of the AWHS augment the apparent shortcomings of BHS, the overestimations in the years following 1973, 1987 and 2008 are not as dramatic as in the case of BHS. The results from the backtests suggest that the dynamic properties of the AWHS improve the forecasting ability of the historical simulation. As indicated in *Table 1*, augmenting the regular BHS by assigning different weights to observations based on their age, results in a decrease in underestimations by 22.5 percentage points. Furthermore, using AWHS results in correct estimations in 67.2

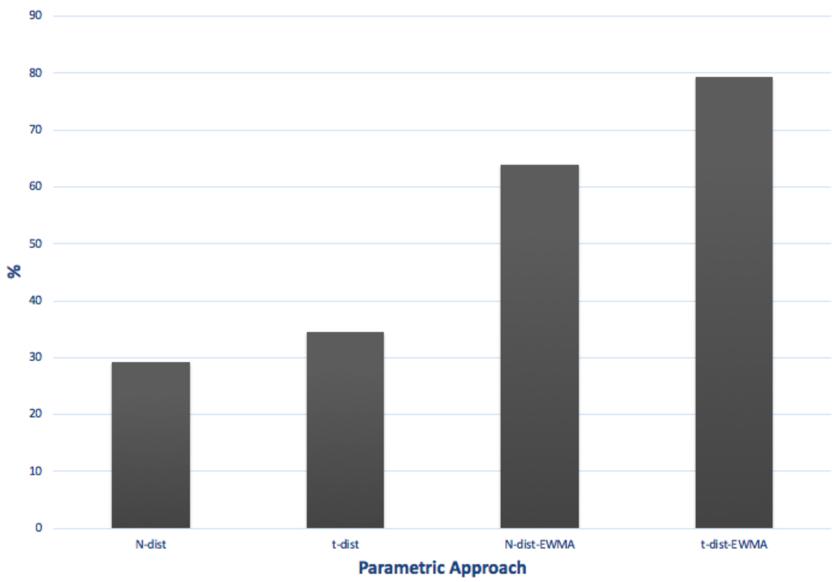
of the years as opposed to 56.9% of years for the regular BHS. Further evidence for the superiority of AWHS to BHS is provided by the results of the Traffic Light System. Although, AWHS results in six yellow lights, the red lights, or severe underestimations, that the BHS result in are completely eliminated. Thus from the perspective of financial regulators, the results suggest that AWHS is preferable to BHS.

The third non-parametric approach, VWHS, also attempts to augment the BHS by taking volatility clustering into account. The rationale behind VWHS is that if volatility is higher than average in the current holding period, the forecasts of VaR and ES should be rescaled accordingly. The results indicate that carrying out VWHS yields accurate forecasts of ES even for the turbulent years of 1973, 1987 and 2008, suggesting that the dynamic properties of the VWHS more reliably take the empirical finding of volatility clustering into account than do both BHS and AWHS. Carrying out VWHS results in underestimations of ES in 10.30% of the years 1962-2019 and in zero overestimations of ES. As in the case of AWHS, implementing VWHS also eliminates all red lights form the Traffic Light System.

A comparison of the accuracy of the three non-parametric approaches is summarized in *Figure 1*. Both AWHS and VWHS attempt to augment the shortcomings of BHS by, in the case of AWHS assign a greater weight to more recent observations, and in the case of VWHS, by rescaling the ES forecasts upwards (downwards) based on if volatility is higher (lower) than average in the current holding period. *Figure 1* suggests that accounting for volatility clustering through a VWHS produces more accurate forecasts of ES than does the methodology implemented in the AWHS. An ocular inspection of *Figure 3*, *Figure 4* and *Figure 5* (A Appendix), suggests greater dynamic properties of VWHS compared to both BHS and AWHS, further augmenting this conclusion.

5.2 Backtest of Parametric Approaches

Figure 2: The percentage of correct ES estimations for the parametric approaches for the years 1962-2019



Examining the results from the backtests of the N-dist parametric approach, the drawbacks of modelling the loss distribution based on a Normal distribution are evident. Using the N-dist approach only results in correct ES estimations in 29.3% of the investigated years. Furthermore, in the years when the model does underestimate ES, the z-statistics from the Acerbi and Szekely test indicate that the degree to which the approach underestimates ES is quite severe in comparison to most of the other parametric approaches. Examining the results from the Traffic Light System (*Table 3* of the B Appendix) shows that following the N-dist approach results in twelve red lights, or severe underestimations. Fama (1965) and Mandelbrot (1963) argue in their respective papers that financial data is not typically characterized by a Normal distribution, but instead by a distribution with leptokurtic properties. The results yielded from carrying out the backtests (presented in *Table 1*, in unison with the evidence presented in Fama (1965) and Mandelbrot (1963) suggest that basing an estimation approach on the assumption of a Normal distribution does not produce accurate forecasts of ES and will likely lead to too low capital reserves.

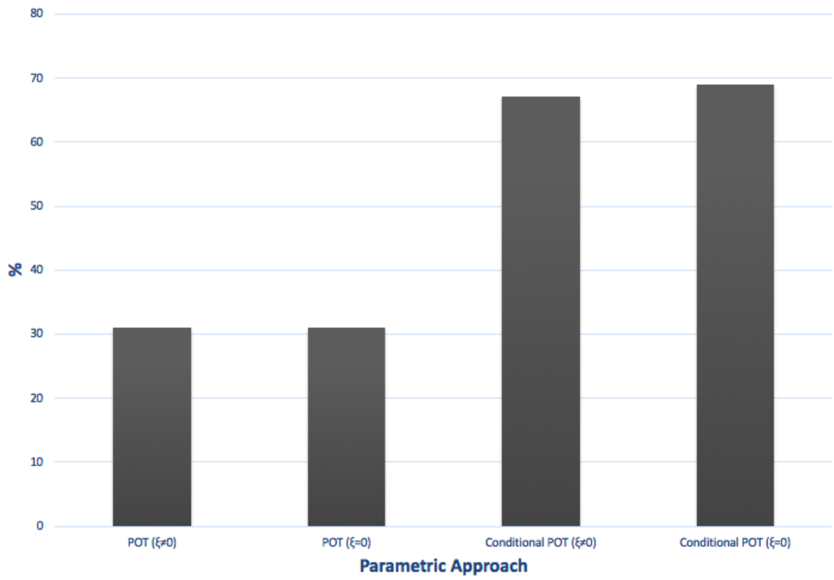
Volatility can be estimated either simply by a rolling window of the standard deviations of previous losses or by an EWMA approach where the volatility is estimated according to the prevailing unexpected losses and recent volatility in a GARCH process. As in the case of the aforementioned VWHS, the N-dist-EWMA approach attempts to augment the shortcomings of the regular N-dist approach by accounting for volatility clustering. The data suggests that estimating volatility through the EWMA approach as opposed to a rolling window, improves the N-dist approach significantly. The number of correct estimations is increased by 34.5% percentage points and the degree to which it underestimates ES in the years that it does in fact underestimate is reduced considerably. While the N-dist-EWMA does produce 21 yellow lights, indicating that the model has a tendency to slightly underestimate risk, the presence of red lights is completely eliminated, suggesting that from a financial regulator's perspective, implementing an EWMA model to the regular N-dist approach is prudent.

With the aforementioned leptokurtic properties of financial data outlined by Fama (1965) and Mandelbrot (1963), arises the need to take "heavy tails" into account. Examining the results from the t-dist approach, where the assumptions regarding the underlying loss distribution are based on the loss distribution following a t-distribution, it is evident that the forecasts of ES are more accurate when accounting for Fama's and Mandelbrot's observations. Compared to the 29.3% correct estimations produced by the N-dist approach, the t-dist approach yields a total of 34.5% correct estimations for the years 1962-2019. Furthermore, the underestimations are not as severe for the t-dist approach as compared to the approach based on the assumption of a Normal distribution. Nevertheless, the t-dist approach results in underestimation in 31.0% of the investigated years and in 34.5% overestimations. The results thus suggest that basing capital requirements on the assumption of a t-distributed loss distribution is not ideal.

In an attempt to address the failure of the regular t-dist approach to produce accurate forecasts of ES, instead of estimating volatility by a rolling window of the standard deviations of previous losses, an EWMA approach where the volatility is estimated according to the prevailing unexpected losses and recent volatility in a GARCH process is implemented. Examining the results from the t-dist-EWMA approach suggests that applying the EWMA model as opposed to the rolling window approach, drastically improves the performance of the estimation approach. The t-dist-EWMA approach results in correct estimations in 79.3% of the examined year, indicating that out of the four parametric approaches investigated, modelling the loss observations on the assumption of a t-distribution and implementing an EWMA approach, produces the most accurate forecasts of ES. Furthermore, as with the implementation of the EWMA volatility to the N-dist approach, using EWMA volatility instead of a rolling window for the t-dist approach, results in zero overestimations of ES. Thus, as the number of overestimations is minimized and the underestimations are less severe compared to the underestimations produced by the N-dist, N-dist-EWMA and t-dist approaches, the data suggests that modelling the loss distribution according to the the t-dist-EWMA approach, is the preferred method for determining the capital requirements of a firm.

5.3 Backtest of Extreme Value Theory Approaches

Figure 3: The percentage of correct ES estimations for the Extreme Value Theory approaches for the years 1962-2019



As with the non-parametric approaches, the EVT approaches can be divided into "un-dynamic" and "dynamic" models, with he first implementing the regular POT approach and the latter attempting to account for volatility clustering by incorporating an EWMA model to the regular POT approach. Typically, POT is used to forecasts VaR and ES at high confidence levels with "extreme" events taking place with very low frequency. In this context it is not self-evident that making the model more dynamic and responsive to prevailing market conditions is a meaningful exercise.

The POT approach results in 44.8% and 24.1% under- and overestimations respectively for the years 1962 to 2019. In other words, the POT approach only results in correct estimations for 31.0% of the investigated years. Furthermore, for the very turbulent years of 1973, 1987 and 2008 the underestimations of ES are quite severe compared to most of the other parametric and non-parametric approaches.

Incorporating an EWMA approach to the regular POT approach, The conditional POT approach takes current market conditions into account to a higher degree. The data suggests that the accuracy of the ES forecasts is improved drastically by implementing this approach. The conditional POT ($\xi \neq 0$) and conditional POT ($\xi = 0$) result in 1.7% and 3.4% underestimations of ES respectively. Although the number of overestimations is slightly higher for the conditional POT than for the unconditional POT, the number of correct estimations is more than doubled. Additionally, the conditional POT performs quite well for the years 1973, 1987 and 2008, with the conditional POT being the only model together with VWHS to accurately forecast ES for the year 2008. The data thus suggests that there seems to be some merit to making the model more responsive to prevailing market conditions. Examining the results from the Traffic Light System, it is evident that the conditional POT produces the highest number of green lights out of all the models. Thus, from a regulator's standpoint, the data suggests that implementing the conditional POT approach is favourable.

6 Conclusion

Previous research on the the forecasting ability of various ES estimation approaches has largely been carried out from the perspective of financial regulators. That is, research has largely focused on finding a model that does not underestimate market risk. Although finding a model that does not underestimate market risk is undeniably important, both for the financial regulator attempting to ensure the stability of the financial system, and for the financial institution attempting to mitigate the consequences of turbulent market conditions, overestimation of market risk arguably also has certain drawbacks. Thus, when evaluating and drawing conclusions about the forecasting ability of the various ES estimation approaches, the relative importance of overestimation and underestimation has to be considered.

As for the non-parametric approaches, the data suggests that while the forecasting ability is improved when moving from BHS to AWHS, VWHS has a superior ability in producing accurate ES estimates. The data suggests that this is the case both when only the ability of the models not to underestimate market risk is considered as well as when the idea of finding a model that neither underestimates nor overestimates is introduced. As for the parametric approaches, the results augment the conclusions drawn by previous research about the drawbacks of modelling financial data according to probability distributions such as the Normal distribution or Student's t-distribution. Even when attempting to account for the empirical finding of volatility clustering by introducing an EWMA model to the aforementioned N-dist and t-dist approaches, the data suggests that VWHS still yields more accurate forecasts.

Addressing the accuracy of the EVT approaches, the results suggest that the forecasting ability of the regular POT approach is unsatisfactory form both an over- and underestimation perspective, displaying both severe underestimations for many of the investigated years as well as a tendency of the model to overestimate market risk in years following turbulent times. However, augmenting the model by introducing an EWMA to the regular POT, the conditional POT shows a greater ability to produce accurate forecasts. Although the model overestimates risk in many of the years, with less than 5% underestimations, the conditional POT is from the perspective of regulators, the most suitable model for estimating ES.

In conclusion, if one is simply concerned with implementing a model that does not underestimate market risk, the data suggests that the conditional POT approach is the preferred model. That is, if the sole concern is to prepare for and to mitigate the effects of financial havoc, conditional POT is arguably the most viable option. However, if a more holistic view is taken, where both over- and underestimation is taken into account, the data suggests that VWHS is more adequate in forecasting ES.

7 Further Research

As the results provided in this paper suggest that above average volatility in the current holding period is likely to be proceeded by above average volatility in the next holding period, and that the non-parametric VWHS approach has the greatest ability out of the investigated models to capture this phenomenon, further research focusing on different VWHS approaches would be of great interest. The VWHS approach investigated in this paper is restricted to the implementation of an EWMA model. This is of course by no means the only approach for accounting for volatility clustering. Thus, incorporating other ways of accounting for time-varying volatility such as other GARCH models would be of great interest.

When carrying out the respective estimation approaches, the sample data in this paper is restricted to the daily returns of the S&P500 index. Although, the daily returns of the S&P500 should offer evidence as to the accuracy of the respective models, limiting the sample data to the returns of an equity index may not offer results that are representative of other asset classes. Therefore, testing the forecasting ability of the various models on a series of hypothetical trading books, comprised of various asset classes and aimed at replicating the actual trading book of a financial institution, would provide further empirical evidence as to which model most accurately reflects the market risk of a financial institutions portfolio.

The results from the parametric approaches that assume a Normal distribution or Student's t-distribution do not, as the data suggests, provide accurate forecasts. The results do however suggest that accounting for the excess kurtosis typically exhibited by financial data does improve the accuracy of the model. Hence, modelling the underlying loss distribution on a probability distribution with greater leptokurtic properties may be of interest.

Lastly, when carrying out the non-parametric approaches, the sample size is restricted to the previous years losses, when carrying out the VWHS the decay rate (λ) is limited to the value 0.94 and when carrying out the EVT approaches, the 95^{th} quantile of the previous five years losses is used as a threshold value. For the purpose of future research, it would be of interest to focus on one approach, and attempt to optimize the forecasting ability of that particular model by investigating the effect of changing the underlying parameters.

References

Acerbi, C., & Szekely, B. (2014). Backtesting expected shortfall. Available at: https://dlu-umich.github.io/docs/Research_Insight_Backtesting_Expected_Shortfall_December_2014.pdf [Accessed: 2 May 2020].

Ahsanullah, M., Shakil, M., & Golam Kibria, B.M. (2014). *Normal and Student´s t Distributions and Their Applications. [Electronic resource]*. Atlantis Press (Atlantis Studies in Probability and Statistics: 4). Available at: https://search-ebscohost-com.ludwig.lub.lu.se/login.aspx?direct=true&db=cat07147a&AN=lub.5736263&site=eds-live&scope=site [Accessed: 28 May 2020].

Andersson, K. (2020). A comparative study of VaR and ES using extreme value theory, Unpublished, Bachelor Thesis, Department of Economics, Lund University, School of Economics and Management.

Anevski, D. (2017). *A concise introduction to mathematical statistics*. 1. ed. Lund: Studentlitteratur.

Artzner, P., Delbaen, F., Eber, J.-M., & Heath, D. (1999). Coherent Measures of Risk, *Mathematical Finance*, vol. 9, p. 203. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-9965.00068?casa_token=UA6hlWEM2fgAAAAA:avnk0vUuqrRaoACoHwQhMo0Rh_-GcNEuOPkLsi30xaX1sm_rQIN0314QsQ8DerdPrbDByjwTaw--B56b [Accessed: 15 May 2020].

Balkema, A., & de Haan, L. (1974). Residual Life Time at Great Age, *The Annals of Probability*, vol. 2, no. 5, p. 792. Available at: https://search-ebscohost-com.ludwig.lub.lu.se/login.aspx?direct=true&db=edsjsr&AN=edsjsr.2959306&site=eds-live&scope=site [Accessed: 15 May 2020].

Bank for International Settlements (BIS). (2013). Consultative Document - Fundamental review of the trading book: A revised market risk framework. Available at: https://www.bis.org/publ/bcbs265.pdf [Accessed: 23 April 2020].

Baran, J., & Witzany, J. (2011). A Comparison of EVT and Standard VaR Estimations Available at: https://ssrn.com/abstract=1768011 [Accessed: 15 May 2020].

Bredin, D., & Hyde, S. (2004). FOREX Risk: Measurement and Evaluation Using Value-at-Risk, *Journal of Business Finance & Accounting*, vol. 31, no. 9/10, p. 1389. Available at: https://search-ebscohost-com.ludwig.lub.lu.se/login.aspx?direct=true&db=bth&AN=15020970&site=eds-live&scope=site [Accessed: 15 May 2020].

Costanzino, N., & Curran, M. (2018). A Simple Traffic Light Approach to Backtesting Expected Shortfall, *Risks*, vol. 6, no. 1, p. 2. Available at: https://search-ebscohost-com.ludwig.lub.lu.se/login.aspx?direct=true&db=edsdoj&AN=edsdoj.2f52623527c418eb99d5805353bb903&site=eds-live&scope=site [Accessed: 15 May 2020].

Degiannakis, S., & Potamia, A. (2017). Multiple-days-ahead value-at-risk and expected shortfall forecasting for stock indices, commodities and exchange rates: Inter-day versus intra-day data, *International Review of Financial Analysis*, vol. 49, p. 176. Available at: https://search-ebscohost-com.ludwig.lub.lu.se/login.aspx?direct=true&db=edselp&AN=S1057521916301594&site=eds-live&scope=site [Accessed: 15 May 2020].

Dowd, K. (2005). *Measuring market risk. [Electronic resource]*. 2. ed. John Wiley & Sons. Available at: https://search-ebscohost-com.ludwig.lub.lu.se/login.aspx?direct=true&db=cat07147a&AN=lub.1875372&site=eds-live&scope=site [Accessed: 15 May 2020].

Fama, E. F. (1965). The Behavior of Stock-Market Prices, *The Journal of Business*, vol. 38, no. 1, p. 34. Available at: https://search-ebscohost-com.ludwig.lub.lu.se/login.aspx?direct=true&db=edsjsr&AN=edsjsr.2350752&site=eds-live&scope=site [Accessed: 15 May 2020].

Geisser, S., Johnson, W. O., & Wiley InterScience (Online service). (2006). *Modes of parametric statistical inference. [Electronic resource]*. Wiley-Interscience (Wiley series in probability and statistics). Available at: https://search-ebscohost-com.ludwig.lub.lu.se/login.aspx?direct=true&db=cat07147a&AN=lub.6159545&site=eds-live&scope=site [Accessed: 15 May 2020].

Gnedenko, B. (1943). Sur La Distribution Limite Du Terme Maximum D'Une Serie Aleatoire, *Annals of Mathematics*, vol. 44, no. 3, p. 423. Available at: https://search-ebscohost-com.ludwig.lub.lu.se/login.aspx?direct=true&db=edsjsr&AN=edsjsr.10.2307.1968974&site=eds-live&scope=site [Accessed: 15 May 2020].

Guermat, C., & Harris, R. D. F. (2002). Forecasting value at risk allowing for time variation in the variance and kurtosis of portfolio returns, *International Journal of Forecasting*, vol. 18, no. 3, p. 409. Available at: https://search-ebscohost-com.ludwig.lub.lu.se/login.aspx?direct=true&db=bth&AN=11646971&site=eds-live&scope=site [Accessed: 15 May 2020].

Harmantzis, F. C., Linyan M., & Yifan, C. (2006). Empirical study of value-at-risk and expected shortfall models with heavy tails, *Journal of Risk Finance (Emerald Group Publishing Limited)*, vol. 7, no. 2, p. 117. Available at: https://search-ebscohost-com.ludwig.lub.lu.se/login.aspx?direct=true&db=edb&AN=20477943&site=eds-live&scope=site [Accessed: 15 May 2020].

Hull, J. (2018). *Risk management and financial institutions.* 5. ed. John Wiley Sons Inc.

Hull, J., & White, A. (1999). Incorporating volatility updating into the historical simulation method for VaR, *Journal of Risk*, vol. 1, no.1. Available at: https://www.researchgate.net/publication/2645882_Incorporating_volatility_updating_into_the_historical_simulation_method_for_VaR [Accessed: 15 May 2020].

Jadhav, D., Ramanathan, T. V., & Naik-Nimbalkar, U. V. (2009). Modified Estimators of the Expected Shortfall, *Journal of Emerging Market Finance*, vol. 8, no. 2, p. 87. Available at: https://journals.sagepub.com/doi/abs/10.1177/097265270900800201?casa_token=F5IFy5unjcUAAAAA:YzQiZGPeaDod8LIGGOgaLMbdhFi7iHMbkYdMnGD1oJc3-tuNUcS0TWyC5c4K3v4KTqWgWGhfJKLpXL0#articleCitationDownloadContainer [Accessed: 15 May 2020].

Jorion, P. (2007). *Value at risk: the new benchmark for managing financial risk*. 3. ed. New York: McGraw-Hill.

Longerstaey, J., & Spencer, M. (1996). *Riskmetrics - technical document*. *[Electronic resource]*. 4. ed. Morgan Guaranty Trust Company of New York. Available at: http://yats.free.fr/papers/td4e.pdf [Accessed: 15 May 2020].

LU, Z., & Huang, H. (2007). Estimating value at risk: From JP Morgan's Standard-EWMA to Skewed-EWMA Forecasting, *20th Australasian Finance & Banking Conference 2007 Paper*. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1009022 [Accessed: 15 May 2020].

Mandelbrot, B. (1963). The Variation of Certain Speculative Prices, *The Journal of Business*, vol. 36, no. 4, p. 394. Available at: https://search-ebscohost-com.ludwig.lub.lu.se/login.aspx?direct=true&db=edsjsr&AN=edsjsr.2350970&site=eds-live&scope=site [Accessed: 15 May 2020].

Marinelli, C., D'Addona, S., & Rachev, S. (2007). A comparison of some univariate models for value-at-risk and expected shortfall, *International Journal of Theoretical and Applied Finance*, vol. 10, no. 6, p. 1043. Available at: https://www.worldscientific.com/doi/abs/10.1142/S0219024907004548 [Accessed: 15 May 2020].

McNeil, A. J. (1997). Estimating the Tails of Loss Severity Distributions Using Extreme Value Theory, *Astin Bulletin*, vol. 27, no. 1, p. 117. Available at: https://search-ebscohost-com.ludwig.lub.lu.se/login.aspx?direct=true&db=edo&AN=ejs40614601&site=eds-live&scope=site [Accessed: 15 May 2020].

McNeil, A. J., & Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach, *Journal of Empirical Finance*, vol. 7, no. 3, p. 271. Available at: https://search-ebscohost-com.ludwig.lub.lu.se/login.aspx?direct=true&db=edselp&AN=S0927539800000128&site=eds-live&scope=site [Accessed: 15 May 2020].

Norton, M., Khokhlov, V., & Uryasev, S. (2018). Calculating CVaR and bPOE for Common Probability Distributions With Application to Portfolio Optimization and Density Estimation, working paper. Available at: https://search-ebscohost-com.ludwig.lub.lu.se/login.aspx?direct=true&db=edsarx&AN=edsarx.1811.11301&site=eds-live&scope=site [Accessed: 15 May 2020].

Richardson, M. P., Boudoukh, J., & Whitelaw, R. F. (1997). The Best of Both Worlds: A Hybrid Approach to Calculating Value at Risk. Available at: http://dx.doi.org/10.2139/ssrn.51420 [Accessed: 15 May 2020].

Sobreira, N., & Louro, R. (2020). Evaluation of volatility models for forecasting Value-at-Risk and Expected Shortfall in the Portuguese stock market, *Finance Research Letters*, vol. 32, Available at: https://search-ebscohost-com.ludwig.lub.lu.se/login.aspx?direct=true&db=edselp&AN=S1544612318305403&site=eds-live&scope=site [Accessed: 15 May 2020].

A Appendix

Figure 4: Expected Shortfall estimates and loss observations from the **Basic Historical Simulation** 1962-2019 (*Note: Positive values indicate a loss and negative values indicate a gain*)

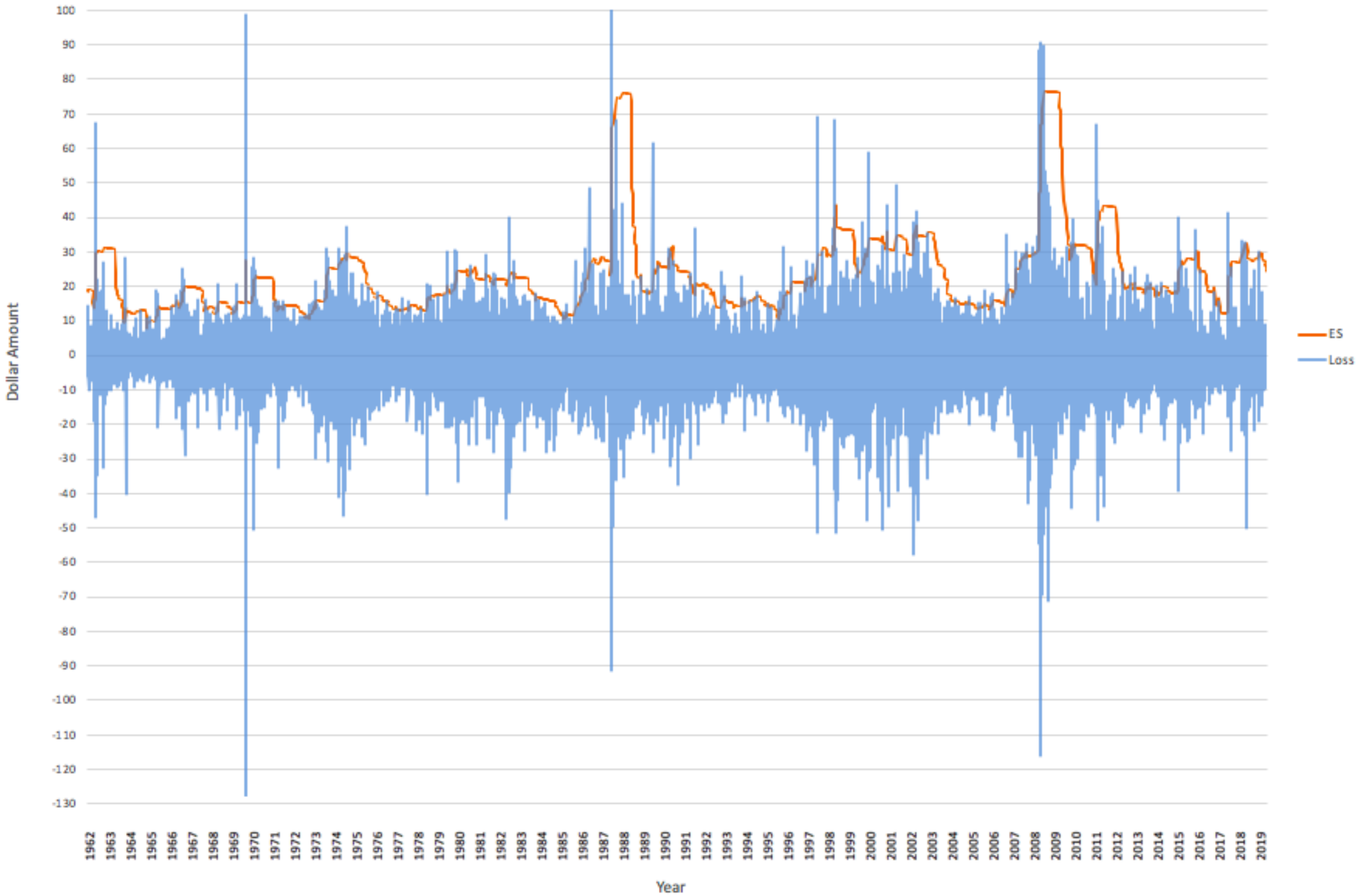


Figure 5: Expected Shortfall estimates and loss observations from the **Age-Weighted Historical Simulation** 1962-2019 (*Note: Positive values indicate a loss and negative values indicate a gain*)

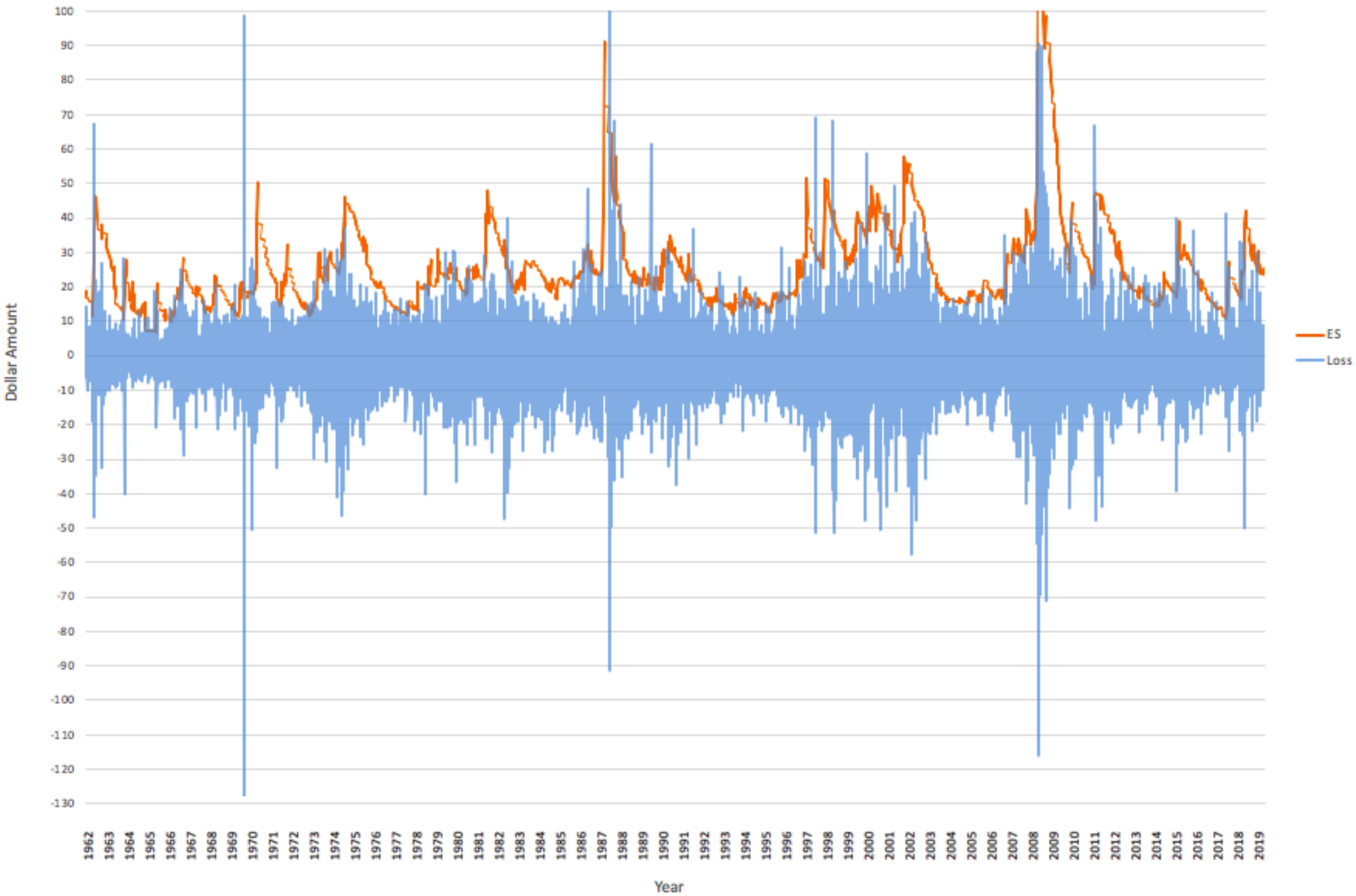


Figure 6: Expected Shortfall estimates and loss observations from the **Volatility-Weighted Historical Simulation** 1962-2019 (*Note: Positive values indicate a loss and negative values indicate a gain*)

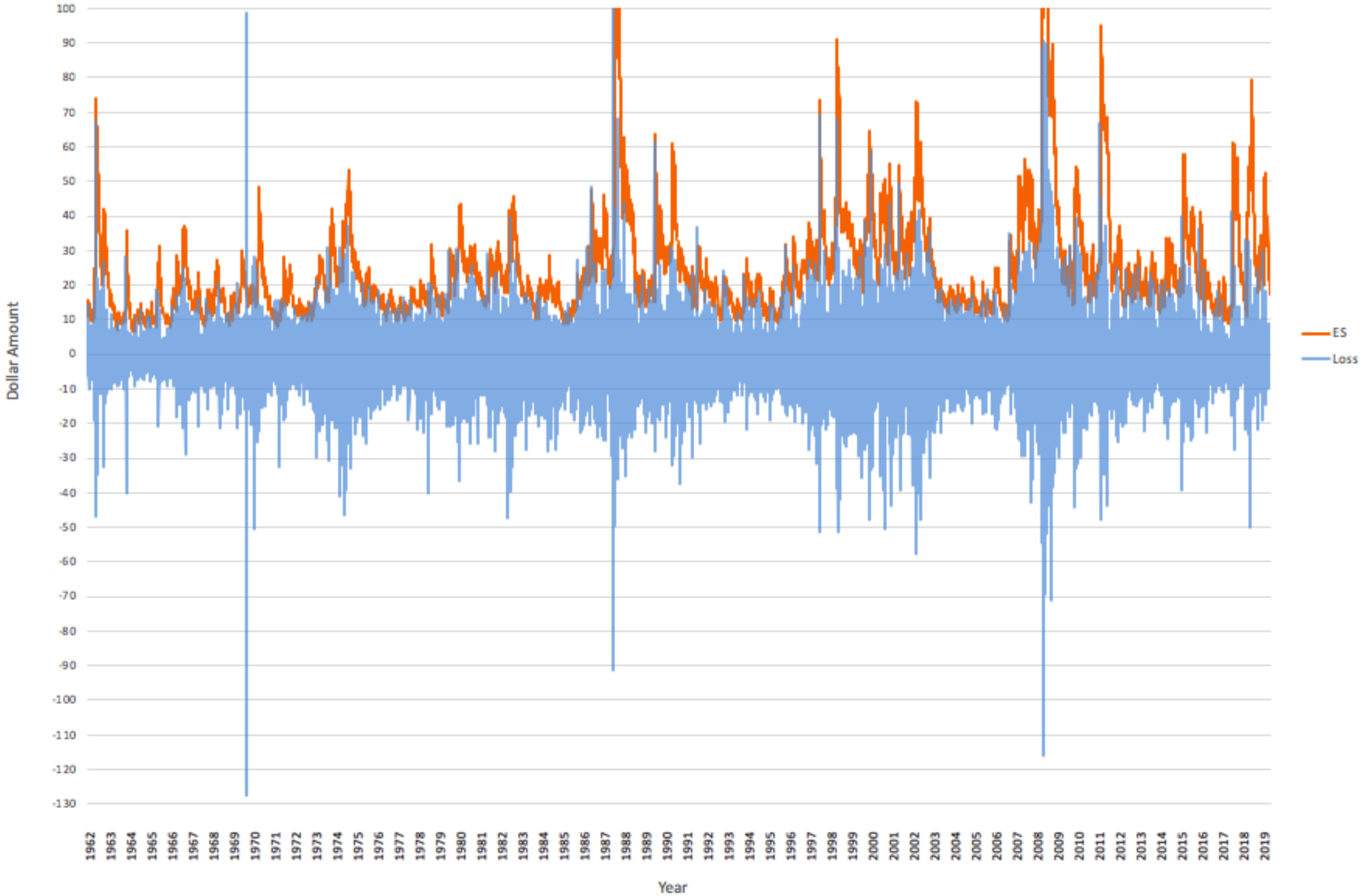


Figure 7: Expected Shortfall estimates and loss observations from the **N-dist** estimation 1962-2019 (*Note: Positive values indicate a loss and negative values indicate a gain*)

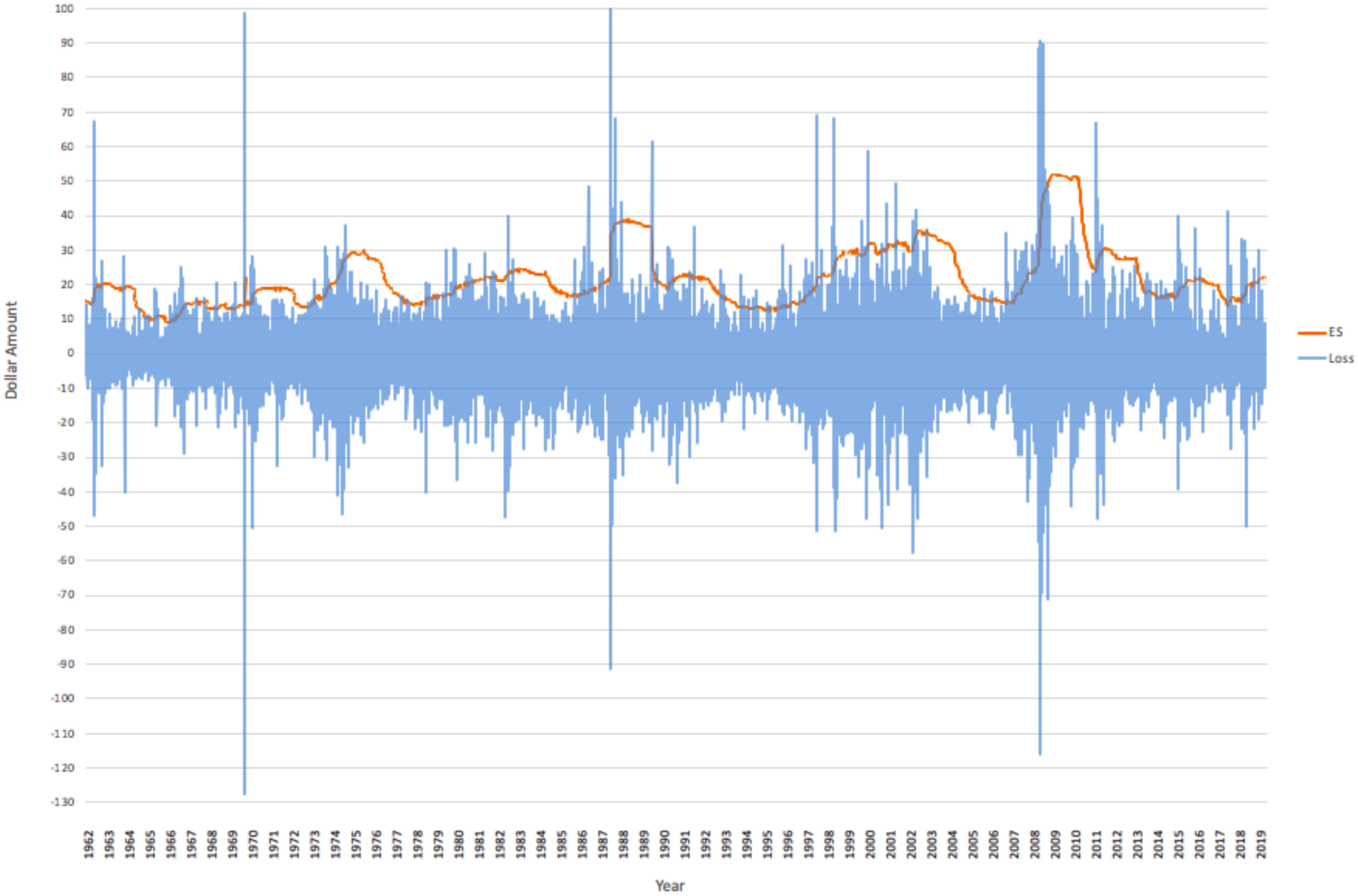


Figure 8: Expected Shortfall estimates and loss observations from the **N-dist-EWMA** estimation 1962-2019 (*Note: Positive values indicate a loss and negative values indicate a gain*)

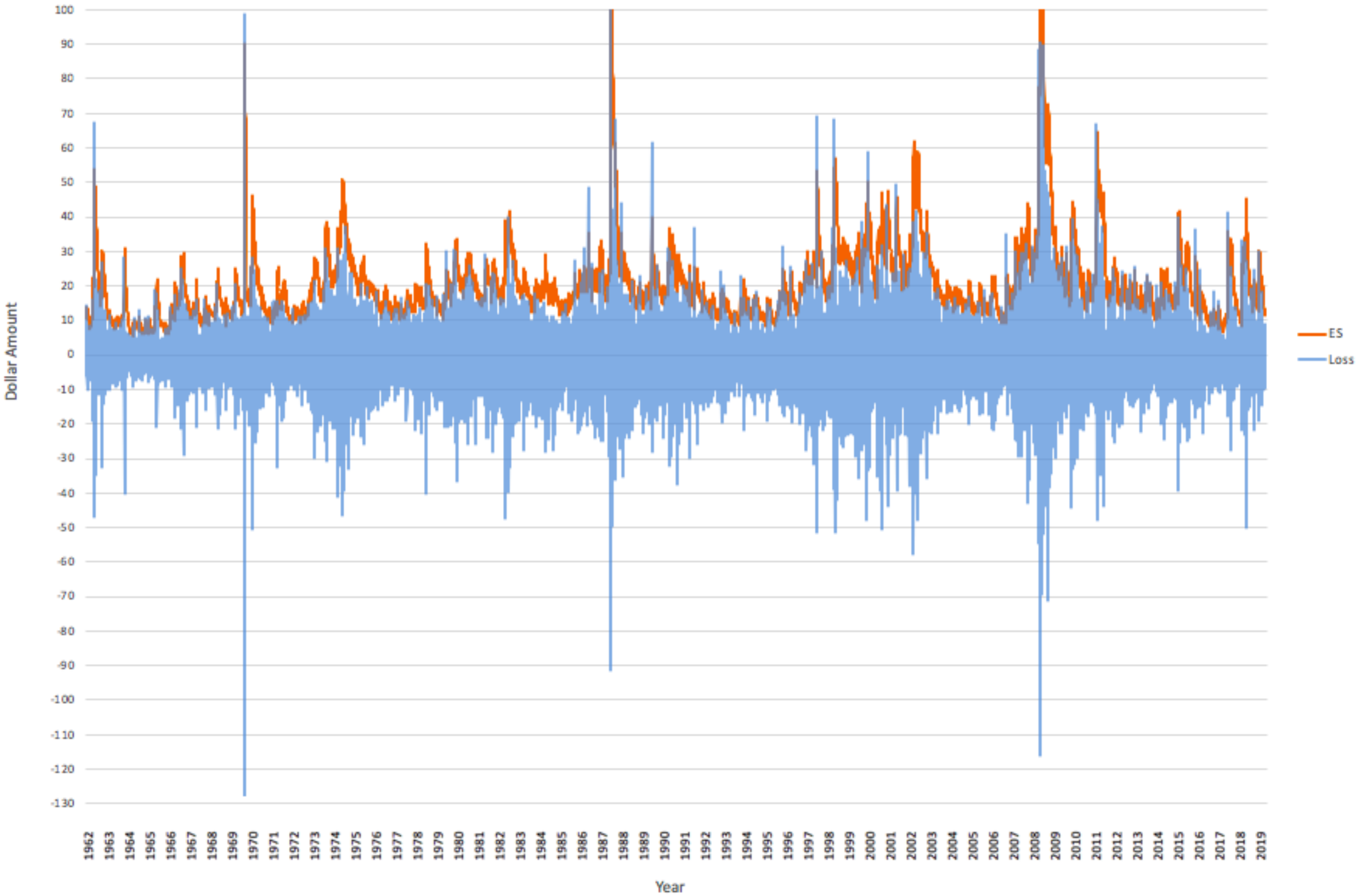


Figure 9: Expected Shortfall estimates and loss observations from the **t-dist** estimation 1962-2019 (*Note: Positive values indicate a loss and negative values indicate a gain*)

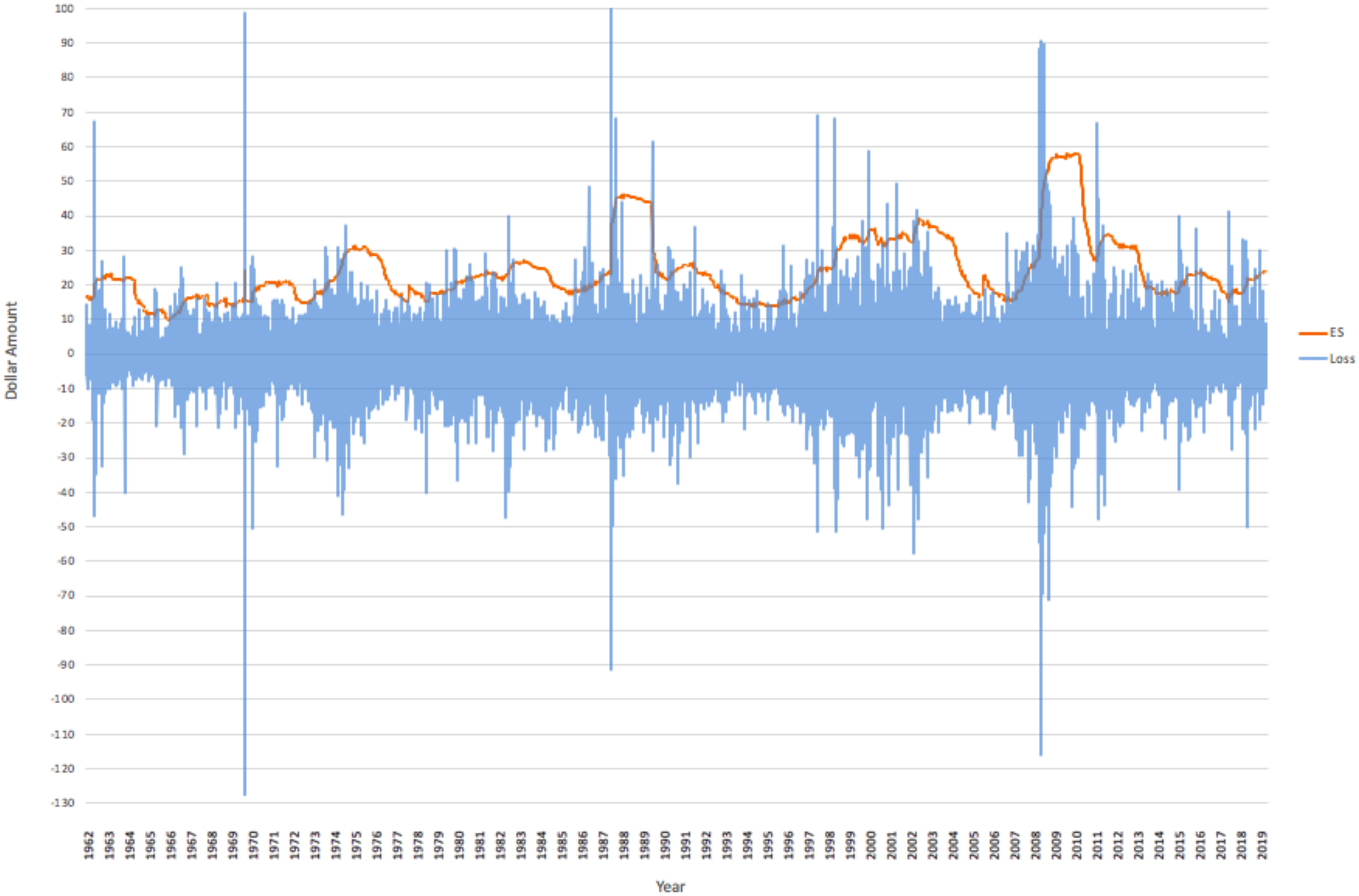


Figure 10: Expected Shortfall estimates and loss observations from the **t-dist-EWMA** estimation 1962-2019 (*Note: Positive values indicate a loss and negative values indicate a gain*)

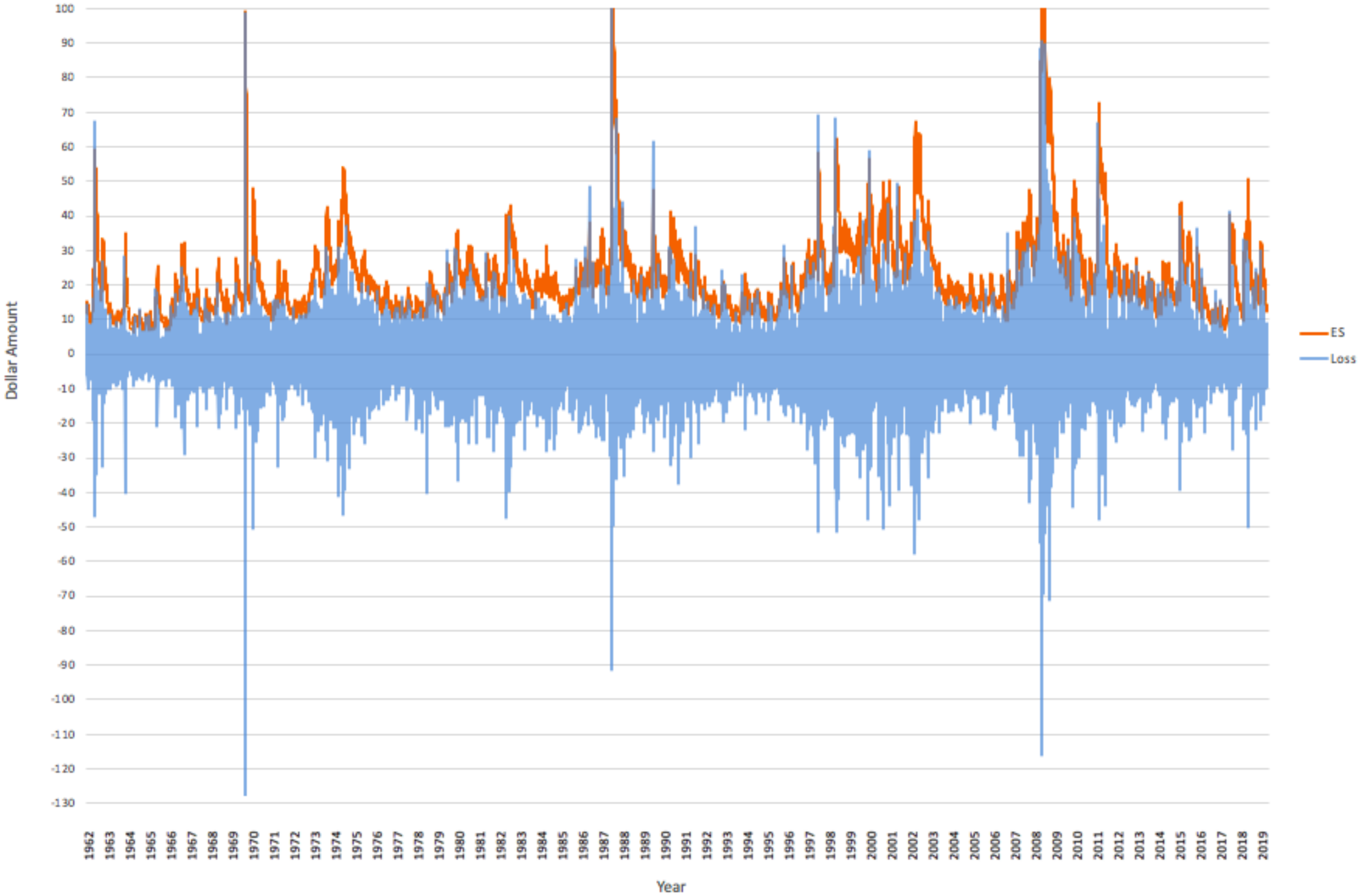


Figure 11: Expected Shortfall estimates and loss observations from the **POT** ($\xi = 0$) estimation 1962-2019 (*Note: Positive values indicate a loss and negative values indicate a gain*)

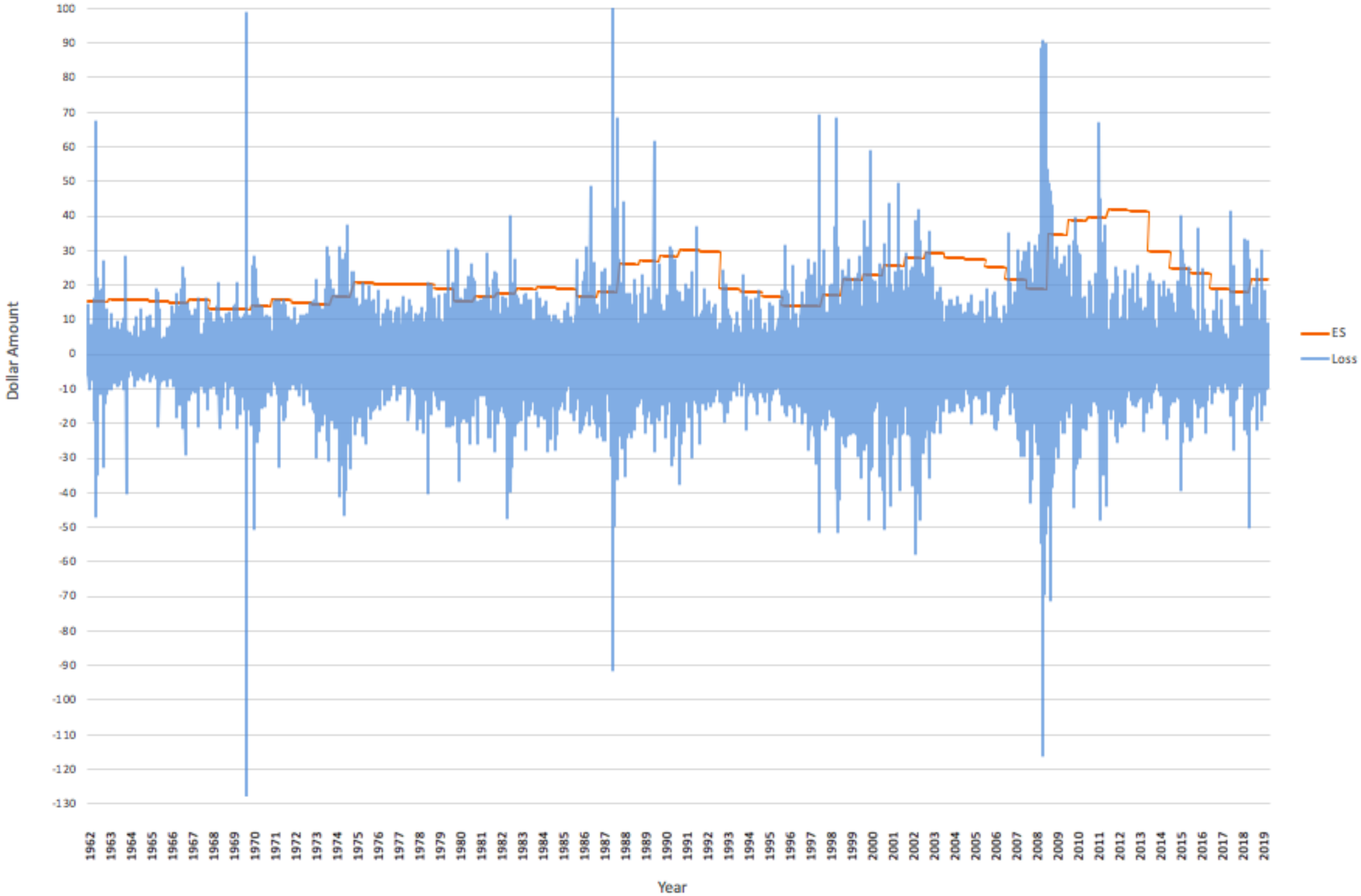


Figure 12: Expected Shortfall estimates and loss observations from the **POT** ($\xi \neq 0$) estimation 1962-2019 (*Note: Positive values indicate a loss and negative values indicate a gain*)

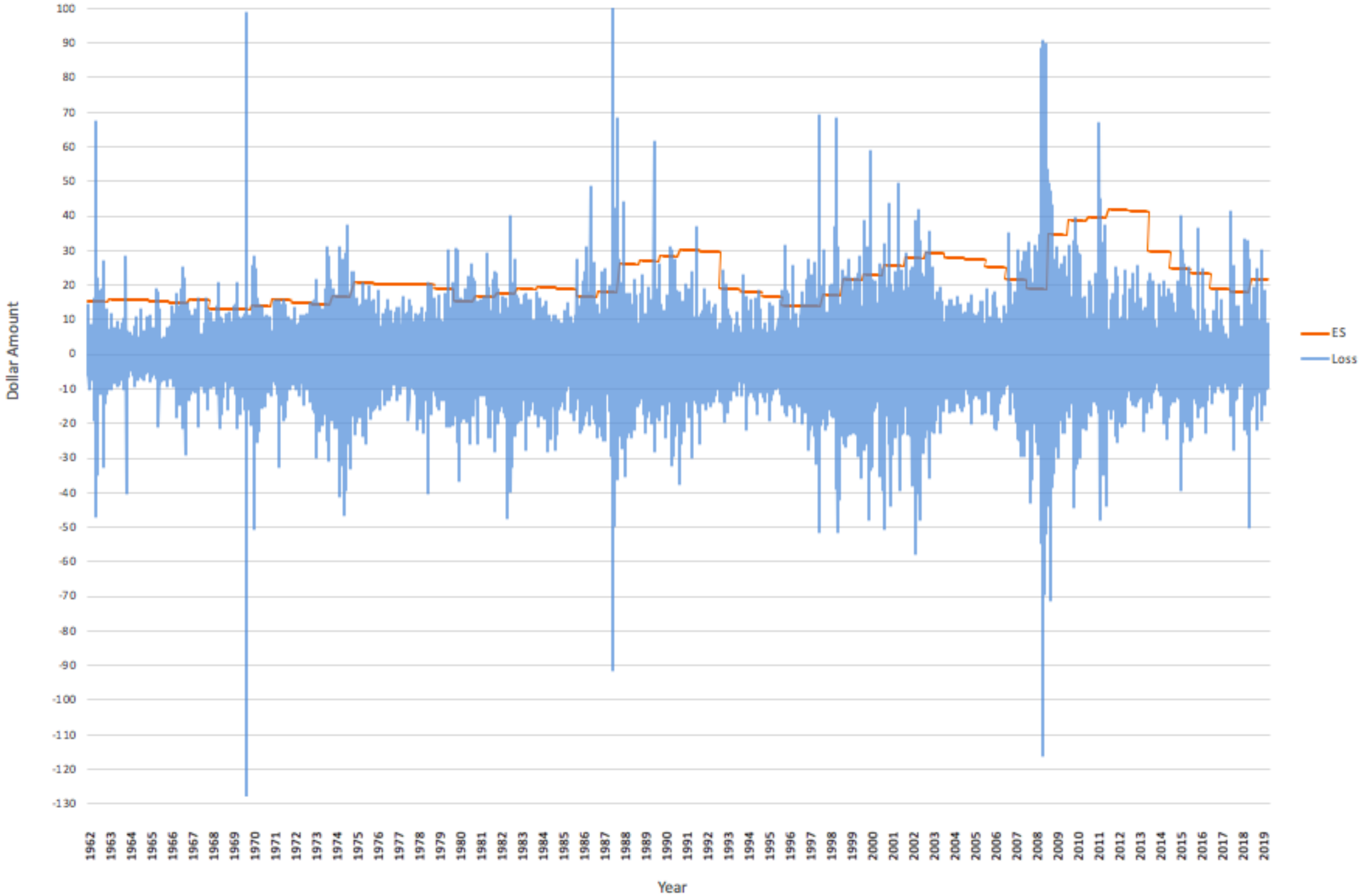


Figure 13: Expected Shortfall estimates and loss observations from the **Conditional POT** ($\xi = 0$) estimation 1962-2019 (*Note: Positive values indicate a loss and negative values indicate a gain*)

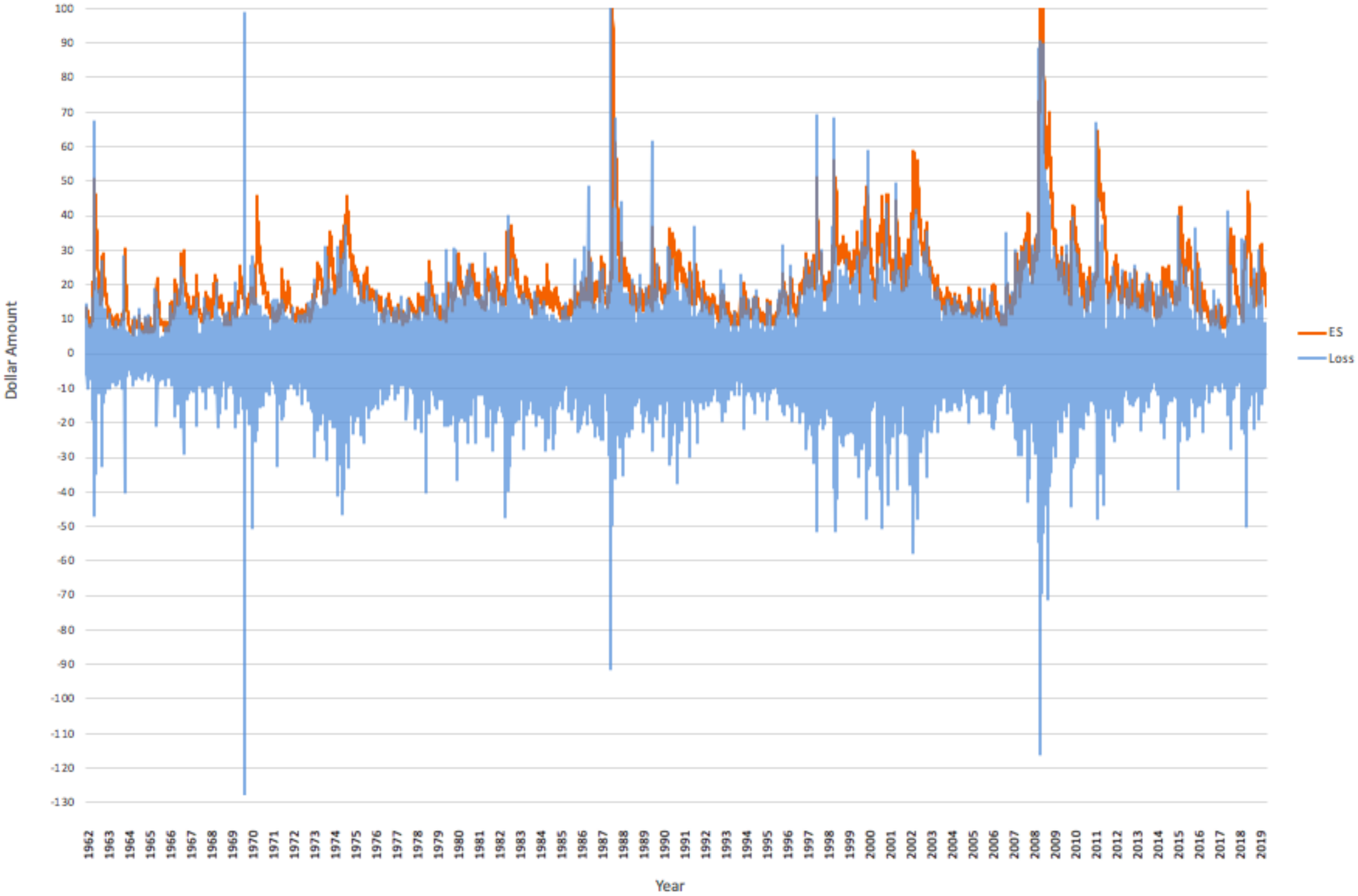
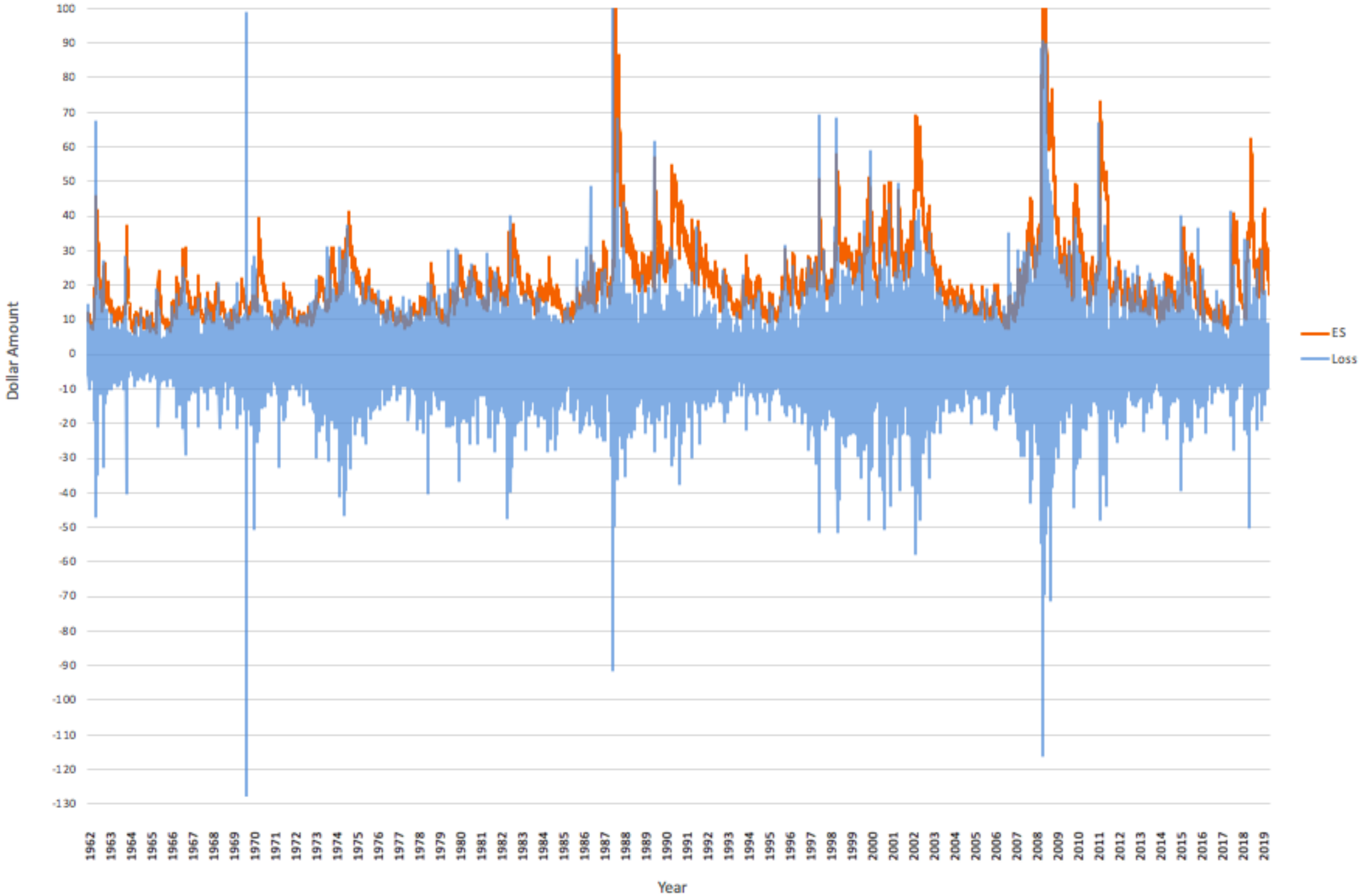


Figure 14: Expected Shortfall estimates and loss observations from the **Conditional POT** ($\xi \neq 0$) estimation 1962-2019 (*Note: Positive values indicate a loss and negative values indicate a gain*)



B Appendix

Table 2: The z-statistics for the eleven estimation approaches from 1962 to 2019 with green (-0.70 < Z < 0.59) indicating a correct estimation, yellow (0.59 < Z) overestimation and red (Z < -0.70) underestimation

	BHS	AWHS	VWHS	N-dist	N-dist-EWMA	t-dist	t-dist-EWMA	POT (xi≠0)	POT (xi=0)	Conditional POT (xi≠0)	Conditional POT (xi=0)
1962	-1.91	-0.32	-0.94	-2.64	-1.56	-2.30	-1.07	-4.95	-4.95	-0.19	-0.23
1963	-0.18	0.09	-0.37	0.76	-0.42	0.79	-0.02	0.48	0.48	0.69	0.62
1964	0.14	0.80	-0.15	0.70	-0.44	0.73	-0.26	0.87	0.87	0.28	0.12
1965	-0.62	-0.20	0.19	-0.76	-1.31	-0.41	-0.89	-0.11	-0.11	0.08	0.15
1966	-1.26	-1.10	-0.11	-4.76	-1.63	-3.59	-1.25	-3.19	-3.19	-0.08	0.03
1967	0.54	0.50	0.28	0.24	-0.66	0.56	-0.48	0.06	0.06	0.82	0.82
1968	-0.10	0.29	0.18	-0.40	-0.76	-0.27	-0.48	-0.99	-0.99	0.47	0.52
1969	-1.13	-0.22	-0.37	-1.20	-0.98	-1.01	-0.68	-2.15	-2.15	0.61	0.65
1970	-0.82	-0.22	0.37	-1.97	0.08	-1.84	0.12	-3.13	-3.13	0.82	0.84
1971	-0.27	0.00	-1.00	1.00	-0.58	1.00	-0.20	-0.17	-0.17	0.19	0.16
1972	0.55	0.59	-0.06	0.86	-0.40	1.00	0.10	0.61	0.61	0.64	0.67
1973	-3.36	-1.33	0.00	-3.48	-0.45	-2.82	-0.19	-3.57	-3.62	0.60	0.49
1974	-0.88	-0.23	0.36	-2.60	0.37	-1.89	0.53	-3.61	-3.61	0.80	0.81
1975	0.84	0.75	-0.33	1.00	0.23	1.00	0.40	0.09	0.09	0.50	0.51
1976	0.40	0.74	-0.11	1.00	0.36	1.00	0.39	0.62	0.62	0.63	0.65
1977	-0.30	0.15	-0.63	0.52	-0.61	0.53	-0.59	0.75	0.75	-0.57	-0.53
1978	-1.26	-0.35	0.43	-0.97	0.18	0.40	0.16	0.40	0.40	0.29	0.30
1979	0.00	0.12	-0.13	-0.08	-0.33	0.13	-0.12	0.18	0.18	-0.09	-0.15
1980	-0.91	0.07	0.05	-1.47	-0.49	-1.33	-0.27	-2.09	-2.09	-0.16	-0.14
1981	0.65	0.15	-0.06	0.03	-0.76	0.21	-0.53	-1.95	-1.95	0.46	0.45
1982	-0.25	-0.30	-0.06	-0.05	-0.22	-0.01	-0.18	-1.07	-1.07	0.29	0.29
1983	0.45	0.85	-0.59	0.82	0.54	0.83	0.58	-0.82	-0.82	0.85	0.84
1984	0.35	-0.60	0.09	1.00	0.18	1.00	0.24	0.32	0.32	0.52	0.48
1985	-0.15	0.57	-0.56	0.86	0.36	0.87	0.40	1.00	1.00	0.48	0.50
1986	-2.15	-0.14	-0.96	-1.85	-0.75	-1.67	-0.64	-2.62	-2.62	-0.52	-0.48
1987	-2.21	-0.72	-0.64	-3.82	-1.26	-3.39	-0.93	-2.28	-2.28	-0.46	-0.71
1988	0.75	0.88	0.20	0.53	0.04	0.60	0.19	-0.95	-0.95	0.61	0.42
1989	-0.23	0.09	-0.73	0.73	-0.91	0.77	-0.60	0.01	0.01	0.08	0.04
1990	-0.48	-0.20	0.13	-1.58	-0.55	-0.70	-0.38	-0.99	-0.99	0.46	0.49
1991	0.25	0.03	-0.26	0.43	-0.14	0.49	-0.01	0.40	0.40	0.63	0.59
1992	0.39	0.74	-0.19	0.86	-0.11	1.00	0.21	0.90	0.90	0.69	0.68
1993	0.14	0.18	-0.35	0.58	-0.23	0.62	0.00	0.64	0.64	0.32	0.10
1994	-0.66	-0.20	-0.29	-1.19	-0.94	-1.05	-0.81	-0.53	-0.53	-0.12	-0.08
1995	0.32	0.25	0.13	0.47	0.10	0.51	0.17	0.59	0.59	0.54	0.62
1996	-1.02	-0.49	-0.27	-1.49	-0.96	-1.15	-0.79	-2.16	-2.16	-0.22	-0.28
1997	-1.21	-0.79	-0.18	-2.08	-0.82	-1.56	-0.66	-2.47	-2.47	0.08	0.09
1998	-0.75	0.14	-0.40	-1.53	-1.18	-0.82	-0.36	-2.86	-2.85	-0.02	0.10
1999	0.55	0.52	0.26	0.70	-0.04	0.74	0.45	-1.46	-1.46	1.00	1.00
2000	-0.49	-0.47	-0.28	-0.44	-0.98	-0.28	-0.52	-1.98	-1.86	0.30	0.41
2001	-0.28	0.30	-0.08	-0.23	-0.15	-0.04	-0.08	-1.16	-1.16	0.63	0.60
2002	-1.21	-0.34	0.14	-0.77	-0.29	-0.62	-0.06	-2.66	-2.66	0.69	0.64
2003	0.84	1.00	-0.24	0.71	0.24	0.85	0.29	0.25	0.25	1.00	1.00
2004	0.42	0.26	-0.35	1.00	-0.27	1.00	-0.06	1.00	1.00	0.13	0.03
2005	0.17	0.03	-0.11	0.55	-0.10	0.58	-0.02	1.00	1.00	0.52	0.49
2006	-0.33	0.16	-0.23	0.00	-0.27	0.16	-0.15	0.88	0.88	-0.06	-0.07
2007	-2.00	-0.69	-0.88	-3.50	-1.69	-3.31	-1.58	-0.50	0.19	-1.66	-1.26
2008	-3.15	-0.93	-0.15	-7.10	-1.54	-6.19	-1.09	-6.06	-6.27	-0.10	0.11
2009	1.00	0.77	0.50	0.08	-0.33	0.41	0.42	-1.29	-1.31	0.85	0.84
2010	-0.02	0.25	-0.39	1.00	-1.21	1.00	-0.84	-0.02	-0.02	0.26	0.31
2011	-1.48	-0.28	-0.36	-2.15	-1.48	-1.56	-1.09	-0.94	-0.94	0.02	-0.10
2012	0.84	0.66	-0.06	0.73	-0.51	0.88	-0.22	1.00	1.00	0.65	0.67
2013	0.33	0.64	-0.22	0.72	-0.69	0.87	-0.39	1.00	1.00	0.44	0.33
2014	-0.57	-0.32	-0.49	-0.75	-1.46	-0.64	-1.29	0.44	0.44	-0.29	-0.21
2015	-0.84	0.21	0.22	-1.74	-0.73	-1.20	-0.63	-0.66	-0.66	0.17	0.28
2016	0.23	0.54	-0.10	-0.28	-0.11	-0.18	-0.03	-0.25	-0.25	0.49	0.55
2017	0.21	0.62	-0.29	0.86	-0.35	0.87	0.00	0.60	0.60	0.28	0.24
2018	-1.91	-1.37	-0.75	-5.05	-1.56	-4.02	-1.16	-3.40	-3.40	-0.02	0.02
2019	0.39	0.61	0.24	0.02	-0.60	0.09	-0.35	-0.50	-0.51	0.36	0.46

Table 3: The z-statistics and corresponding Basel type traffic light color for the eleven estimation approaches from 1962 to 2019 with green (-0.70 < Z), yellow (-1.80 < Z < -0.70) and red (Z < -1.80)

	BHS	AWHS	VWHS	N-dist	N-dist-EWMA	t-dist	t-dist-EWMA	POT (xi≠0)	POT (xi=0)	Conditional POT (xi≠0)	Conditional POT (xi=0)
1962	-1.91	-0.32	-0.94	-2.64	-1.56	-2.30	-1.07	-4.95	-4.95	-0.19	-0.23
1963	-0.18	0.09	-0.37	0.76	-0.42	0.79	-0.02	0.48	0.48	0.69	0.62
1964	0.14	0.80	-0.15	0.70	-0.44	0.73	-0.26	0.87	0.87	0.28	0.12
1965	-0.62	-0.20	0.19	-0.76	-1.31	-0.41	-0.89	-0.11	-0.11	0.08	0.15
1966	-1.26	-1.10	-0.11	-4.76	-1.63	-3.59	-1.25	-3.19	-3.19	-0.08	0.03
1967	0.54	0.50	0.28	0.24	-0.66	0.56	-0.48	0.06	0.06	0.82	0.82
1968	-0.10	0.29	0.18	-0.40	-0.76	-0.27	-0.48	-0.99	-0.99	0.47	0.52
1969	-1.13	-0.22	-0.37	-1.20	-0.98	-1.01	-0.68	-2.15	-2.15	0.61	0.65
1970	-0.82	-0.22	0.37	-1.97	0.08	-1.84	0.12	-3.13	-3.13	0.82	0.84
1971	-0.27	0.00	-1.00	1.00	-0.58	1.00	-0.20	-0.17	-0.17	0.19	0.16
1972	0.55	0.59	-0.06	0.86	-0.40	1.00	0.10	0.61	0.61	0.64	0.67
1973	-3.36	-1.33	0.00	-3.48	-0.45	-2.82	-0.19	-3.57	-3.62	0.60	0.49
1974	-0.88	-0.23	0.36	-2.60	0.37	-1.89	0.53	-3.61	-3.61	0.80	0.81
1975	0.84	0.75	-0.33	1.00	0.23	1.00	0.40	0.09	0.09	0.50	0.51
1976	0.40	0.74	-0.11	1.00	0.36	1.00	0.39	0.62	0.62	0.63	0.65
1977	-0.30	0.15	-0.63	0.52	-0.61	0.53	-0.59	0.75	0.75	-0.57	-0.53
1978	-1.26	-0.35	0.43	-0.97	0.18	0.40	0.16	0.40	0.40	0.29	0.30
1979	0.00	0.12	-0.13	-0.08	-0.33	0.13	-0.12	0.18	0.18	-0.09	-0.15
1980	-0.91	0.07	0.05	-1.47	-0.49	-1.33	-0.27	-2.09	-2.09	-0.16	-0.14
1981	0.65	0.15	-0.06	0.03	-0.76	0.21	-0.53	-1.95	-1.95	0.46	0.45
1982	-0.25	-0.30	-0.06	-0.05	-0.22	-0.01	-0.18	-1.07	-1.07	0.29	0.29
1983	0.45	0.85	-0.59	0.82	0.54	0.83	0.58	-0.82	-0.82	0.85	0.84
1984	0.35	-0.60	0.09	1.00	0.18	1.00	0.24	0.32	0.32	0.52	0.48
1985	-0.15	0.57	-0.56	0.86	0.36	0.87	0.40	1.00	1.00	0.48	0.50
1986	-2.15	-0.14	-0.96	-1.85	-0.75	-1.67	-0.64	-2.62	-2.62	-0.52	-0.48
1987	-2.21	-0.72	-0.64	-3.82	-1.26	-3.39	-0.93	-2.28	-2.28	-0.46	-0.71
1988	0.75	0.88	0.20	0.53	0.04	0.60	0.19	-0.95	-0.95	0.61	0.42
1989	-0.23	0.09	-0.73	0.73	-0.91	0.77	-0.60	0.01	0.01	0.08	0.04
1990	-0.48	-0.20	0.13	-1.58	-0.55	-0.70	-0.38	-0.99	-0.99	0.46	0.49
1991	0.25	0.03	-0.26	0.43	-0.14	0.49	-0.01	0.40	0.40	0.63	0.59
1992	0.39	0.74	-0.19	0.86	-0.11	1.00	0.21	0.90	0.90	0.69	0.68
1993	0.14	0.18	-0.35	0.58	-0.23	0.62	0.00	0.64	0.64	0.32	0.10
1994	-0.66	-0.20	-0.29	-1.19	-0.94	-1.05	-0.81	-0.53	-0.53	-0.12	-0.08
1995	0.32	0.25	0.13	0.47	0.10	0.51	0.17	0.59	0.59	0.54	0.62
1996	-1.02	-0.49	-0.27	-1.49	-0.96	-1.15	-0.79	-2.16	-2.16	-0.22	-0.28
1997	-1.21	-0.79	-0.18	-2.08	-0.82	-1.56	-0.66	-2.47	-2.47	0.08	0.09
1998	-0.75	0.14	-0.40	-1.53	-1.18	-0.82	-0.36	-2.86	-2.85	-0.02	0.10
1999	0.55	0.52	0.26	0.70	-0.04	0.74	0.45	-1.46	-1.46	1.00	1.00
2000	-0.49	-0.47	-0.28	-0.44	-0.98	-0.28	-0.52	-1.98	-1.86	0.30	0.41
2001	-0.28	0.30	-0.08	-0.23	-0.15	-0.04	-0.08	-1.16	-1.16	0.63	0.60
2002	-1.21	-0.34	0.14	-0.77	-0.29	-0.62	-0.06	-2.66	-2.66	0.69	0.64
2003	0.84	1.00	-0.24	0.71	0.24	0.85	0.29	0.25	0.25	1.00	1.00
2004	0.42	0.26	-0.35	1.00	-0.27	1.00	-0.06	1.00	1.00	0.13	0.03
2005	0.17	0.03	-0.11	0.55	-0.10	0.58	-0.02	1.00	1.00	0.52	0.49
2006	-0.33	0.16	-0.23	0.00	-0.27	0.16	-0.15	0.88	0.88	-0.06	-0.07
2007	-2.00	-0.69	-0.88	-3.50	-1.69	-3.31	-1.58	-0.50	0.19	-1.66	-1.26
2008	-3.15	-0.93	-0.15	-7.10	-1.54	-6.19	-1.09	-6.06	-6.27	-0.10	0.11
2009	1.00	0.77	0.50	0.08	-0.33	0.41	0.42	-1.29	-1.31	0.85	0.84
2010	-0.02	0.25	-0.39	1.00	-1.21	1.00	-0.84	-0.02	-0.02	0.26	0.31
2011	-1.48	-0.28	-0.36	-2.15	-1.48	-1.56	-1.09	-0.94	-0.94	0.02	-0.10
2012	0.84	0.66	-0.06	0.73	-0.51	0.88	-0.22	1.00	1.00	0.65	0.67
2013	0.33	0.64	-0.22	0.72	-0.69	0.87	-0.39	1.00	1.00	0.44	0.33
2014	-0.57	-0.32	-0.49	-0.75	-1.46	-0.64	-1.29	0.44	0.44	-0.29	-0.21
2015	-0.84	0.21	0.22	-1.74	-0.73	-1.20	-0.63	-0.66	-0.66	0.17	0.28
2016	0.23	0.54	-0.10	-0.28	-0.11	-0.18	-0.03	-0.25	-0.25	0.49	0.55
2017	0.21	0.62	-0.29	0.86	-0.35	0.87	0.00	0.60	0.60	0.28	0.24
2018	-1.91	-1.37	-0.75	-5.05	-1.56	-4.02	-1.16	-3.40	-3.40	-0.02	0.02
2019	0.39	0.61	0.24	0.02	-0.60	0.09	-0.35	-0.50	-0.51	0.36	0.46