Backtesting Expected Shortfall

A comparative empirical evaluation of different backtests

Viktor Fredriksson & Jesper Johansson

Supervisor: Birger Nilsson

Department of Economics, Lund University, NEKN01, Master's Thesis (15 ECTS credits)

June 2020

Abstract

This paper empirically evaluates whether different backtests for Expected Shortfall (ES) produce similar results. In 2016, the Basel Committee on Banking Supervision proposed a shift from Value-at-Risk (VaR) to ES as the industry standard when calculating capital requirements for banks. However, ES has been found difficult to backtest. Since backtesting results form the basis for determining the capital requirements of banks it is important to elucidate whether the backtests produce similar results. We answer this question by performing six different daily backtests on the S&P 500 index for the period 1965-2020 and measuring correlations between the different backtests. We found a substantial divergence across different backtests. We also found that the correlations remain stable or increase during the global financial crisis. In the light of these results we recommend practitioners to diversify between multiple backtests.

Keywords: Expected Shortfall, Backtests, Value-at-Risk, Empirical, Risk

Acknowledgements

We would like to extend our sincerest gratitude to our supervisor Birger Nilsson. First and foremost for his excellent course in risk management that inspired this essay, and for his valuable and thoughtful supervision during the writing of this essay. We would also like to thank the Lund University School of Economics for providing us with the knowledge and resources needed to write this essay.

Contents

1 Introduction 3

2 Background 5

2.1 The History of Risk Measures 5

2.2 Value-at-Risk 7

2.3 Expected Shortfall 8

2.4 Backtesting 9

2.4.1 Backtesting Value-at-Risk 9

2.4.2 Backtesting Expected Shortfall 11

2.4.3 The Traffic Light System 12

3 Theory 14

3.1 The Design of Different Expected Shortfall Backtests 14

3.1.1 Acerbi and Szekelys' (2014) Non-parametric Tests 14

3.1.2 Constanzino and Curran's (2018) Traffic Light Tests 17

3.1.3 Emmer, Kratz and Tasche's (2015) Approximative Quantile Test 19

3.1.4 Moldenhauer and Pitera's (2019) Secured Position Test 21

3.1.5 Righi and Ceretta's (2015) Truncated Distribution Tests 23

4 Methodology and Data 26

4.1 Data and Descriptive Statistics 26

4.2 Estimation of VaR and ES 28

4.2.1 Parametric Estimation of VaR and ES 28

4.2.2 Non-parametric Estimation of VaR and ES 31

4.3 Monte Carlo Simulation 33

4.4 The Traffic Light System 35

5. Empirical Results 37

5.1 Backtesting 1965-2020 37

5.2 Backtesting The Global Financial Crisis 44

6. Conclusion 48

7. Discussion and Further Research 50

References 51

Appendix 55

1 Introduction

Risk is a central aspect of financial activities. There are many different types of risks and several ways of quantifying it. Since the early 1990s, Value-at-Risk (VaR) has been the predominant measure of market risk (Jorion, 2006). Simply put, VaR quantifies possible portfolio losses into a single number. However, VaR has been subject to criticism. It has been shown that the risk measure does not always encourage diversification. Also, losses greater than VaR are disregarded, thereby VaR fails to uncover “tail risk” (Artzner, Delbaen, Eber, and Heath, 1999). These drawbacks led to the introduction of the risk measure Expected Shortfall (ES). ES measures the average of losses larger than VaR, thereby considering a broader spectrum of potential losses. In 2016, the Basel Committee on Banking Supervision (BCBS) prescribed a shift from VaR to ES in calculating capital requirements for banks. However, while accounting for tail risk and encouraging diversification, ES has introduced another issue: it is complicated to backtest.

In this paper, we do not intend to participate in the ongoing debate on VaR versus ES. Nor is it an attempt to systematically investigate the quality of individual backtests. Instead, our goal is to determine whether different backtests for ES produce similar results. This is important to clarify, because if different backtests produce divergent results, the quality of a risk model and thereby the capital requirement of the bank is contingent on which particular test statistics are used.

We will analyse this problem by first describing six pre-existing backtests. We will create a coherent framework analogous to the Traffic Light System for VaR proposed by the Basel Committee in 1996. For some of the tests we will employ a Monte Carlo Simulation approach to derive the critical values corresponding to the Traffic Light System. We will then perform the backtests empirically on the US stock index S&P 500 and compare the Traffic Light responses from the various backtests. Lastly, we will provide some recommendations.

The remainder of the paper is organized as follows. Section 2 provides background on VaR and ES, and reviews the backtesting literature. In Section 3, we present and describe six prevalent ES backtests which are the basis of our empirical analysis. Section 4 outlines our methodological framework for estimating and backtesting ES. In Section 5, we present the results from the empirical analysis. Section 6 and Section 7 are conclusion and discussion, respectively.

2 Background

In this section, we will give a brief background on the history of risk measures, establish definitions of some important concepts, and provide an overview of the backtesting literature.

2.1 The History of Risk Measures

Financial risk management has been an important issue for regulators and financial institutions for a long time. However, VaR did not become a prevalent concept until the stock market crash of 1987 (Jorion, 2006). VaR was adopted to systematically measure a trading firm's risk exposures across its different portfolios. It is a holistic measure in the sense that it considers all types of exposures, and applies to market, credit and operational risk (Alexander, 2009). Before VaR, risk among commercial trading houses was measured and controlled on a desk-by-desk basis, basically neglecting the firm-wide exposures (Culp, Miller & Neves, 1998).

VaR is a measurement of the worst case loss at a predetermined confidence level of the profit and loss distribution (P&L) given a certain holding period (Alexander, 2009). In 1996, following new banking regulations for market risk, RiskMetrics declared VaR as its regular measure of risk, which then became an industry standard within international financial risk management, supported by the Basel Committee (Acerbi & Szekely, 2014).

However, the debate on appropriate risk measures has been, and still is, lively among researchers. VaR remains popular in practice due to its conceptual simplicity, universality and straightforward backtesting. Yet, VaR has been criticized because of several shortcomings. Artzner et al. (1999) point at two fundamental issues. Firstly, VaR neglects the shape of the tail. That is, if VaR is exceeded, it tells us nothing about the extent of those potential losses. Secondly, VaR lacks the mathematical property of subadditivity, which means that the VaR of a portfolio as a whole can be higher than the sum of the VaRs of its individual portfolios independently. Hence, it would not necessarily encourage diversification and thereby contradict modern portfolio theory.

Artzner et al. (1999) propose that a risk measure, p , is coherent if it is

(i) *monotonous* : $X, Y \in V, Y \geq X \Rightarrow p(Y) \leq p(X),$

(i) *subadditive* : $X, Y,\ X+Y \in V \Rightarrow p(X+Y) \leq p(X) + p(Y),$

(iii) *positively homogeneous* : $X \in V,\ h > 0,\ hX \in V \Rightarrow p(hX) = hp(X),$

(iv) *translation invariant* : $X \in V,\ a \in \mathbb{R} \Rightarrow p(X+a) = p(X) + a.$

They conclude that VaR is not a coherent risk measure since it does not fulfill the second property in all situations.

In 2001, ES was first introduced as an alternative downside risk measure to VaR. ES is defined as the conditional expectation of the loss for losses beyond the VaR level. By calculating the average of all losses larger than VaR, ES detects tail risk and satisfies the property of subadditivity. This led practitioners to start using ES in addition to VaR. In October 2013, The Basel Committee on Banking Supervision (BCBS) updated its bank trading book rules in the wake of the global financial crisis, aiming to better capture that type of extreme losses. This entailed a change from VaR to ES, which regulators believed will better capture the extreme losses that can occur during times of financial distress (Basel Committee, 2013).

Although ES solves some of the main issues with VaR, there are still mathematical and practical inconveniences. Gneiting (2011) demonstrates that ES is not elicitable. Elicitability is a mathematical property, satisfied by VaR but not ES. Specifically, a risk measure is elicitable if there exists a loss function such that the risk measure is the solution to minimizing the expected loss (Patton, Ziegel & Chen, 2019). This implies that backtesting ES is, if even possible, more complicated than backtesting VaR (Gneiting, 2011).

Given that VaR is not coherent and ES is not elicitable, researchers and regulators keep searching for a more suitable risk measure. In particular, the expectile has gained substantial attention. It was first introduced by Newey and Powell (1987) and has become an emerging alternative to VaR and ES. However, in this paper we are primarily interested in ES due to its growing importance within the Basel regulatory framework.

2.2 Value-at-Risk

VaR has two fundamental parameters. The significance level α (alternatively confidence level $1-\alpha$) and the risk horizon h, a period of time measured in trading days (Alexander, 2009). The significance level is usually determined by an external part. Under the Basel regulatory framework, VaR is backtested at the 1% significance level.

The statistical definition of VaR is given by

$$VaR_{\alpha}(L) = min\{ l : P(L > l) \leq 1 - \alpha \} .$$

The equation gives us the smallest portfolio loss value quantile conditioned on that it is larger than or equal to our significance level of choice. In terms of probability, VaR is simply a quantile of the loss distribution. As Alexander (2009) illustrates, a daily VaR with a 5% significance level is a loss level that we expect to occur with a frequency of 5%, holding the portfolio static for 24 hours. Likewise, with 95% confidence we believe that VaR will not be exceeded for the same portfolio and risk horizon. Alternatively, we expect a loss of the 5% VaR or more one day in 20 days.

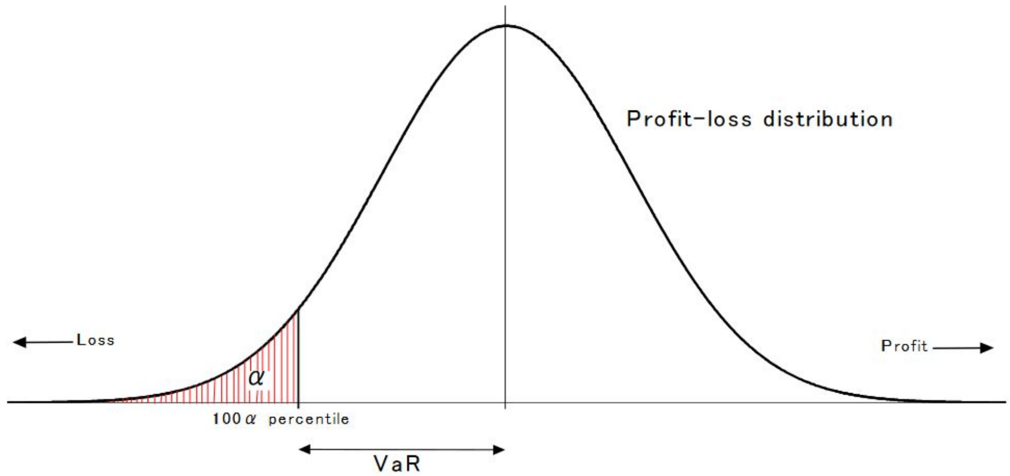

Figure 1. Profit-Loss distribution and VaR (Yamai & Yoshiba, 2002), where VaR is the value of the P&L distribution at percentile α.

2.3 Expected Shortfall

ES, also referred to as conditional VaR, was originally introduced by Rappoport (1993). While VaR asks “How bad can things get?”, ES instead asks the question “If things go bad, what is the expected loss?” (Hull, 2018). The predominant definition of ES is given by

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

which is equivalent to the expected VaR for all confidence levels larger or equal to α. This specification illustrates that VaR needs to be estimated in order to estimate ES. Figure 2 illustrates a P&L distribution for ES and VaR. A more intuitive definition is that ES is the expected value of losses conditional on losses greater VaR:

$$ES_\alpha(X) = E[X \mid X > VaR_\alpha(X)].$$

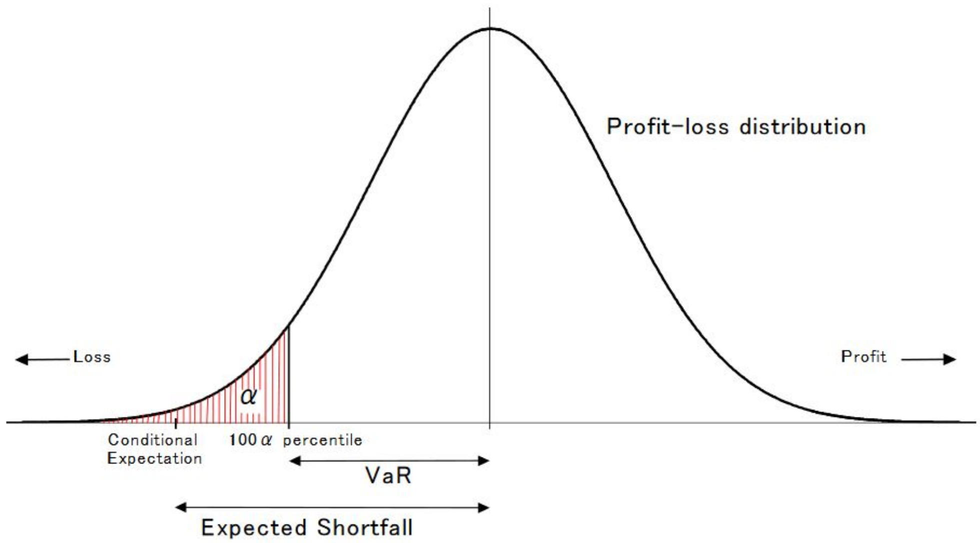

Figure 2. Profit-loss distribution, VaR and ES, (Yamai & Yoshiba, 2002), where ES is the integrated values of the P&L distribution to the left of α.

2.4 Backtesting

The term backtesting has several meanings within finance. It has been described as "a collection of disparate practices in the wait for a clear definition" (Acerbi & Szekely, 2017, p.2). As Christoffersen (2010) points out, it mainly refers to either an assessment of a trading strategy, or the evaluation of financial risk models. In this paper, backtesting refers to the evaluation of risk models.

The idea behind backtesting is straightforward. Jorion (2007) describes it as a statistical method aiming to check if the real losses, observed ex post, are in compliance with the forecasts of the risk measure. In this way, the accuracy of the risk model can be determined.

2.4.1 Backtesting Value-at-Risk

After VaR emerged as a popular risk measure in the early 1990s, the demand for backtesting methodologies was high. An early and influential backtesting procedure was proposed by Kupiec (1995). His POF (proportion of failures) test is an unconditional coverage test. These types of tests count the number of exceedances and compare them with confidence levels. Simply put, you test the quality of the VaR model by counting the number of days on which the realised portfolio loss is greater than the VaR forecast, given a determined confidence level and sample size.

If the number of exceedances differs substantially from $\alpha \times 100\%$ of the sample, the quality of the risk model must be investigated. The test statistic is formulated as

$$POF = 2log\left(\left(\tfrac{1-\widehat{\alpha}}{1-\alpha}\right)^{T-I(\alpha)}\left(\tfrac{\widehat{\alpha}}{\alpha}\right)^{I(\alpha)}\right),$$

$$\widehat{\alpha} = \tfrac{1}{T}I(\alpha),$$

$$I(\alpha) = \sum_{t=1}^{T} I_t(\alpha).$$

This reveals that if the proportion of VaR exceedances, $\widehat{\alpha} \times 100\%$ is exactly equal to $\alpha \times 100\%$, then the test takes the value zero, implying an adequate model. If the underlying model systematically understates or overstates risk, this is reflected in a discrepancy between the observed exceedance rate and the expected exceedance rate.

The POF test by Kupiec (1995) remains widely used and was built upon by the Basel Committee on Banking Supervision in 1996. However, unconditional coverage tests have some drawbacks. Most notably, they may fail to detect VaR measures that have dependent VaR exceedances. As Campbell (2005) points out, a streak of a small number of large unexpected losses over a short period may be a stronger indicator of insufficient risk management than a larger number of evenly occured losses over a relatively longer period.

The inability of unconditional coverage tests to distinguish violations of the independence property led to the development of tests that specifically check the independence property of the VaR exceedance series, $I_t(\alpha)$. A notable test in this category is Christoffersen's (1998) Markov test. It tests if the likelihood of a VaR exceedance is dependent on whether or not a VaR exceedance occurred on the previous day, where dependency indicates problems in the underlying model.

Another independence test was proposed by Christoffersen and Pelletier (2004). The idea behind their duration test is that if VaR exceedances are independent from each other, then the amount of time passing by between VaR exceedances should be independent of the amount of time passing by since the last violation.

Another influential part of the VaR literature is the conditional autoregressive VaR (CAViaR) class of models developed by Engle and Managanelli (2004). There are also conditional tests. These test for both unconditional coverage and independence. Some examples are the time between failures likelihood ratio test (Haas, 2001), the multivariate autocorrelations test (Hurlin & Tokpavi, 2006), the dynamic binary tests (Dumitrescu, Hurlin & Pham, 2012) and the generalized Markov tests (Pajhede, 2015). For a comprehensive review of different VaR backtests, see Zhang and Nadarajah (2018).

2.4.2 Backtesting Expected Shortfall

Although ES solves some of the main issues with VaR, there are still mathematical and practical inconveniences. As Du and Escanciano points out, "the major challenge in the implementation of the ES as the leading measure of market risk is the unavailability of simple tools for its evaluation (2017, p. 40)". When Gneiting (2011) demonstrated that ES, contrary to VaR, lacks the mathematical property of elicitability, many were concerned claiming that a non-elicitable functional cannot be backtestable. However, ES has been shown to be conditionally elicitable (Emmer, Kratz & Tasche, 2015) and jointly elicitable with VaR (Fissler & Ziegel, 2015). Besides, Acerbi and Szekely (2014) argue that elicitability is not inherently related to backtesting itself but rather a way to rank the forecasting performance of different risk models.

The literature on ES backtesting is smaller than the literature on VaR backtesting, but constantly growing since the shift from VaR to ES in trading book capital rules. Some early attempts are the residual approach (McNeil & Frey, 2000), the censored Gaussian approach (Berkowitz, 2001) and the functional delta approach (Kerkhof & Melenberg, 2004). However, as noted by Righi and Ceretta (2015) and Novales and Garcia-Jorcano (2019), there are some pitfalls for these approaches. The reliance on asymptotic test statistics is a risk if the sample size is small, and the fact that the p-values are calculated based on the full sample size rather than conditioning on the number of exceptions might also cause inaccuracies.

In recent years, several approaches for backtesting ES have been suggested, although no coherent framework has been established. Some alternative approaches, which we do not consider in this essay, are the saddle-point techniques by Wong (2008) and Graham and Pál (2014), Acerbi and Szekely's (2017) ridgeback test, Du and Escaniano's (2017) tail risk method and the multinomial VaR implicit backtest (Kratz, Lok & McNeil, 2018).

2.4.3 The Traffic Light System

In 1996, the Basel Committee on Banking Supervision proposed a framework for supervisory interpretation of backtesting results. In their so-called Traffic Light Approach, three color zones are defined via cumulative probabilities of the number of actual VaR exceedances. The Basel Committee ranks backtesting outcomes according to green, yellow and red zones. By choosing multiple thresholds they try to balance type I and type II errors. They conclude that:

> there is no threshold number of exceptions that yields both a low probability of erroneously rejecting an accurate model and a low probability of erroneously accepting all of the relevant inaccurate models. It is for this reason that the Committee has rejected an approach that contains only a single threshold (The Basel Committee on Banking Supervision, 1996, p. 7).

The green zone is defined as the number of exceedances under the null hypothesis whereby the cumulative probability of obtaining that many exceedances or fewer is less than 95%. This result is to a very high certainty consistent with an accurate model. The red zone is defined by a cumulative probability of 99.99% or more, and indicates a risk model that is almost certainly problematic. The yellow zone corresponds to a probability between 95% and 99.99%, and may be consistent with either an accurate or an inaccurate model.

3 Theory

3.1 The Design of Different Expected Shortfall Backtests

In this section we will define and describe six prevalent backtests for ES. They are all simple enough to implement, which is a relevant factor considering the growing demand for viable backtesting procedures for ES within the financial industry. The tests are the foundation of our empirical analysis.

3.1.1 Acerbi and Szekelys' (2014) Non-parametric Tests

Acerbi and Szekely (2014) propose three different non-parametric backtests for ES. The three tests are classified as conditional, unconditional and quantile tests. For this study, we will empirically evaluate the conditional and unconditional test. We omit the third test, since it is not a direct test of ES but rather a test of the full distribution.

Test 1 – Conditional Test

The first test proposed by Acerbi and Szekely (2014) is considered conditional since it requires the estimation of VaR beforehand. They define ES and the test statistic as

$$ES_{\alpha,t} = -E[X_t \mid X_t < -VaR_t],$$

$$Z_1 = \frac{1}{N_T} \sum_{t=1}^{T} \frac{X_t I_t}{ES_{\alpha,t}} + 1,$$

where, N_T is the number of VaR exceedances in the evaluation period, and X_t is the P&L distribution in the evaluation period. I_t is defined as an indicator function $I_t = 1_{\{X_t < -VaR_t\}}$ which is equal to one if $X_t < -VaR_t$ and zero otherwise. $ES_{\alpha,t}$ is the expected value of the losses in the estimation window conditional on VaR exceedances. The test statistic is an equally weighted average of the rescaled losses conditional on VaR exceedances. The null hypothesis is formulated as:

$$H_0 : P_t = F_t \, , \; \forall \; t \, ,$$

where P_t and F_t are the expected and observed distributional tail. The null is formulated against

$$H_1 : ES^F_{\alpha,t}(X_t) \geq ES_{\alpha,t}(X_t) \;\; and \;\; VaR^F_{\alpha,t}(X_t) = VaR_{\alpha,t}(X_t) \; ,$$

where F_t is the realised and P_t is the expected distribution of P&L. Under H_o, the test does not reject the estimation method for both ES and VaR while under H_1, the test rejects ES without rejecting VaR. Acerbi and Szekely (2014) highlight that since the test is an expected value of exceedances, it is completely insensitive to the number of VaR exceedances. Notice that $E_{H_o}[Z_1] = 0$ and $E_{H_1}[Z_1] < 0$. This indicates that the expected value of the test statistic is zero under H_0, and negative under H_1.

The test requires a Monte Carlo Simulation approach to attain the critical values for the backtest. The practical implementation of the simulations is further explained in Section 4.3.

Test 2 – Unconditional Test

The second test proposed by Acerbi and Szekely (2014) is unconditional because it tests ES directly without any need to first backtest VaR. They define ES and the test statistic as:

$$ES_{\alpha,t} = -E[\tfrac{X_t I_t}{\alpha}] \, ,$$

$$Z_2 = \frac{1}{T\alpha} \sum_{t=1}^{T} \frac{X_t I_t}{ES_{\alpha,t}} + 1 \, ,$$

where T is the number of observations in the estimation window, and α is the significance level. The difference between the first and second test can be observed in the weighting parameter. In the conditional test, the losses are scaled by the number of VaR exceedances, yielding an equally weighted average of exceedances. However, in the unconditional test the losses are scaled by the expected number of exceedances a priori given a significance level α and estimation window T $\left(\frac{1}{T\alpha}\right)$. Consequently, the unconditional test is more sensitive to the number of VaR exceedances, e.g. it will presumably reject a large number of small VaR exceedances while the conditional test is insensitive to the number of VaR exceedances. Therefore, the unconditional test is not only sensitive to the magnitude of exceedances but also the number of exceedances. The null and alternative hypothesis are defined as:

$$H_0 : P_t = F_t \,,\ \forall\ t \,,$$
$$H_1 :\ ES^F_{\alpha,t} \geq ES_{\alpha,t}\ \ \forall\ t \text{ and } > \text{ for some } t \text{ and } VaR^F_{\alpha,t} \geq VaR_{\alpha,t} \,,$$

where F_t and P_t are defined as in the conditional test. Note again that $E_{H_o}[Z_2] = 0$ and $E_{H_1}[Z_2] < 0$. The difference between the conditional and unconditional regarding H_1 is due to the difference in the weighting parameter. For the unconditional test, rejecting H_0 means rejecting both the estimated ES and VaR.

A convenient result of the second test is the stability of the p-values across different distributions. Therefore, Monte Carlo simulations to obtain the critical values are not required (Acerbi & Szekely, 2014). The critical values simulated by Acerbi and Szekely (2014) are illustrated in Table 1.

Degrees of freedom	Significance level	
v	5%	0.01%
3	-0.82	-4.4
5	-0.74	-2
10	-0.71	-1.9
100	-0.7	-1.8
N(0,1)	-0.7	-1.8

Table 1: Acerbi and Szekely's (2014) left tail critical values.

3.1.2 Constanzino and Curran's (2018) Traffic Light Tests

Constanzino and Curran (2015) propose a backtest for any spectral risk measure, that is a risk measure which is a weighted average of outcomes, including ES, using a finite-sample distribution. In accordance with established Basel practice, Constanzino and Curran (2018) expand on their findings with a Traffic Light test. The backtest is an extension of the unconditional coverage test for VaR, but also measures the severity of an exceedance. The critical values are derived from both an approximative asymptotic distribution and a finite sample distribution. Since the test is an extension of the VaR coverage test, the individual and summarized VaR indicator functions for a significance level α need to be defined:

$$I_{VaR}^{(i)}(\alpha) = 1_{\{X_i \le -VaR_i(\alpha)\}},$$

$$I_{VaR}^{N}(\alpha) = \sum_{i=1}^{N} 1_{\{X_i \le -VaR_i(\alpha)\}},$$

where X_i and VaR_i is the P&L distribution and VaR estimate respectively for the period i. Analogous to $I_{VaR}^{(i)}$, Constanzino and Curran (2018) define the ES exceedance indicator, $I_{ES}^{(i)}$, and $ES(\alpha)$ for significance level α as:

$$ES(\alpha) = \int_0^{\alpha} VaR(p)dp,$$

$$I_{ES}^{(i)}(\alpha) = \frac{1}{\alpha}\int_0^{\alpha} 1_{\{X_i \le -VaR_i(p)\}}dp \rightarrow (1 - \frac{F_x(X_i)}{\alpha})\, 1_{\{X_i \le -VaR_i(\alpha)\}},$$

$$I_{ES}^{(i)}(\alpha) = \theta^{(i)}(\alpha)\, I_{VaR}^{(i)},$$

where $\theta^{(i)}$ is the measure of severity of the exceedance and F_L is the cumulative distribution of the P&L distribution. To exemplify the role of the exceedance indicator, Constanzino and Curran (2018) provide the following examples:

$$X_i = VaR_i \rightarrow F_X(X_i) = 1 \rightarrow \theta^{(i)}(\alpha) = 0 \rightarrow I_{ES}^{(i)}(\alpha) = 0.$$

This makes intuitively sense as the loss is equivalent to the VaR, hence there is no exceedance and ES should be equal to zero. Further, they show that as $X_i \rightarrow -\infty$ $\theta^{(i)} \rightarrow 1 \rightarrow I_{ES}^{(i)} = 1$. Hence, the severity of an individual loss value for $I_{ES}^{(i)} \in (0,1)$ is continuous. Generalizing the equation to N trading days yields the following sum:

$$I_{ES}^{N}(\alpha) = \sum_{i=1}^{N} \theta^{(i)}(\alpha)\, I_{VaR}^{(i)}(\alpha).$$

The value of ES for N trading days is the sum of exceedances with their respective severity. The backtesting procedure is a modification of the POF test for VaR exceedances proposed by Kupiec (1995), described in Section 2.4.1. However, since the ES exceedance indicator is continuous, they adjust the test by choosing and inverting the quantile to attain the corresponding exceedance value (Constanzino & Curran, 2018). The boundaries are computed under the null hypothesis:

$$H_o : \{I_{ES}^{(i)}\}_{i=1}^{N} \; iid \; \forall \; i \neq j \quad \text{and} \; P[\, X_i \leq VaR_i(p)\,] = p \in [0, \alpha].$$

They use the result derived in Constanzino and Curran (2015) to show that you can make a normal approximation for any $\alpha \in (0,1)$:

$$\sqrt{N}(\, I_{ES}^{N}(\alpha) - \mu_{ES}) \rightarrow_d N(0, \sigma_{ES}^2),$$
$$\mu_{ES} = \tfrac{1}{2}\alpha N,$$
$$\sigma_{ES}^2 = \alpha\left(\tfrac{4-3\alpha}{12}\right).$$

From the above approximation, the Z-score test statistic can be derived as

$$\sqrt{N}\frac{(I_{ES}^{N}(\alpha) - \mu_{ES})}{\sigma_{ES}} \sim N(0,1).$$

Constanzino and Curran (2018) derive the asymptotic boundaries for the Traffic Light test using the above approximation, yielding:

$$\text{Green light: } I^N_{ES} < 5.4768\,,$$
$$\text{Yellow light: } 5.4768 \leq I^N_{ES} < 9.229\,,$$
$$\text{Red light: } I^N_{ES} \geq 9.229\,.$$

Constanzino and Curran (2018) also derive the finite sample boundaries for the Traffic Light test using a numerical root finding procedure, yielding:

$$\text{Green light: } I^N_{ES} < 5.7049\,,$$
$$\text{Yellow light: } 5.7049 \leq I^N_{ES} < 9.8833\,,$$
$$\text{Red light: } I^N_{ES} \geq 9.8833\,.$$

3.1.3 Emmer, Kratz and Tasche's (2015) Approximative Quantile Test

Emmer, Kratz and Tasche (2015) propose a simple approximative method for backtesting ES based on a representation of ES as integrated VaR:

$$ES_\alpha(X) = \frac{1}{1-\alpha}\int_\alpha^1 q_u(X)\,du \approx \frac{1}{4}[\,q_\alpha(X) + q_{0,75\alpha+0,025}(X) + q_{0,5\alpha+0,5}(X) + q_{0,25\alpha+0,75}(X)\,],$$

where $q_\alpha(X)$ is the VaR of the P&L distribution at quantile α. A similar approach for backtesting has been proposed by the Basel Committee in the Financial Trading Book Review (2013) with the VaR quantile levels 97.5 % and 99 %. Emmer, Kratz and Tasche (2015) suggest using four quantile levels. Still, there is no conclusive guidance in the literature regarding the optimal number of VaR levels. Since we compare different backtests for ES, we want a more precise approximation. Hence, we choose to use five VaR quantile levels. In line with Kratz, Lok and McNeil (2018), the VaR quantile levels are derived from the formula

$$\alpha_j = \alpha + \frac{j-1}{N}(1-\alpha),\ j = 1\ldots N\ ,$$

where α_j is the $j-th$ order significance level. The approximative ES at a 97.5 % confidence level can then be written as:

$$ES_{0.975,t}(X) \ = \ \tfrac{1}{5}[VaR_{0,975}(X) + VaR(X)_{0,98} + VaR(X)_{0.985} + VaR(X)_{0.99} + VaR(X)_{0.995}].$$

That is, VaR 97.5%, 98%, 98.5%, 99% and 99.5% are jointly backtested. This constitutes a linear approximation of ES.

This method is attractive due to its practical simplicity. There exist well established methods for backtesting VaR and it requires no Monte Carlo Simulation for generating the statistical parameters needed for backtesting (Emmer, Kratz & Tasche, 2015). Also, since VaR is elicitable, the approximative ES is too. Consequently, this backtest circumvents the discussion of whether ES is backtestable or not.

Similar to Kupiec's POF test the number of observed exceedances are compared with the cumulative probability of observing that number of exceedances for every VaR level. The cumulative probabilities of each VaR level is illustrated in Table 2.

Number of Exceedances	Value-at-Risk quantile levels 0,975	0,98	0,985	0,99	0,995
0	0,0018	0,0064	0,0229	0,0811	0,2856
1	0,0129	0,0384	0,1086	0,2837	0,6426
2	0,0479	0,1188	0,2700	0,5380	0,8666
3	0,1213	0,2539	0,4733	0,7517	0,9607
4	0,2374	0,4247	0,6662	0,8868	0,9905
5	0,3849	0,5982	0,8132	0,9554	0,9980
6	0,5415	0,7458	0,9069	0,9846	0,9997
7	0,6848	0,8537	0,9583	0,9953	0,9999
8	0,7998	0,9230	0,9831	0,9987	1,0000
9	0,8822	0,9628	0,9937	0,9997	…
10	0,9356	0,9834	0,9979	0,9999	…
11	0,9672	0,9931	0,9993	1,0000	…
12	0,9843	0,9973	0,9998	…	…
13	0,9930	0,9990	0,9999	…	…
14	0,9970	0,9997	1,0000	…	…
15	0,9988	0,9999	…	…	…
16	0,9995	1,0000	…	…	…
17	0,9998	…	…	…	…
18	0,9999	…	…	…	…
19	1,0000	…	…	…	…
20	…	…	…	…	…

Table 2. Cumulative probabilities of the VaR quantile levels. The test yields a green light if the cumulative probability of the observed number of exceedances are below 95% for all VaR levels. Red light corresponds to a cumulative probability above 99.99% for at least one of the VaR levels. Yellow light corresponds to a cumulative probability above 95% and below 99.99% for at least one of the VaR levels.

To perform the backtest, the number of exceedances are estimated for each quantile and compared with the cumulative probability of observing that number of exceedances. Contingent on the number of exceedances, the risk model is classified with a green, yellow or red light in line with the suggested Traffic Light System.

3.1.4 Moldenhauer and Pitera's (2019) Secured Position Test

Moldenhauer and Pitera (2019) argue that a backtesting procedure for ES should be transparent, holistic and intuitive such that it is applicable to all asset classes. Moreover, they conclude that proposed methods in previous literature require an advanced mathematical framework, certain model assumptions, reference estimation processes or large samples. Therefore, they advocate a simple non-parametric approach with a Traffic Light test resembling the POF test for VaR proposed by Kupiec (1995).

To backtest ES, they define the secured position, Y , as a combination of the realised financial position and the capital reserve:

$$Y = X + \widehat{p},$$

where X is the P&L distribution of the evaluation period, and $\widehat{p}$ is the estimated capital reserve derived from a chosen internal modelling approach, which can be any distributional risk measure like VaR or ES. It can be calculated using e.g. Historical Simulation, normal approximation or Monte Carlo Simulation methods. The backtesting procedure is performed by counting the number of exceedances, where an exceedance is observed when $Y < 0$. Intuitively, an exceedance occurs when the daily P&L are negative such that $\widehat{p} - X < 0$.

To perform the backtests, we assume that we have x_t P&L for $t = 1, \ldots, t-1$. Further, we estimate $\widehat{p}_t$ as our daily ES forecast. The daily secured position is given by

$$y_t = x_t + \widehat{p}_t .$$

To perform the backtest, the test statistic G_t is defined as

$$G_T = \sum_{t=1}^{T} \frac{1_{\{y_{[1]} + \ldots + y_{[t]}\}}}{T},$$

where y_t is our sample of secured positions. The test is carried out by taking the sum of the worst secured position realisations and observing the sign of the sum. Notice that we do not only test the number of worst exceedances, but also the severity.

Moldenhauer and Pitera (2019) provide the critical values corresponding to the Traffic Light System. For $T = 250$ and $ES = 2.5\%$ the expected number of exceedances are 6,25. A confidence band around the number of expected exceedances provides us with the three color zones for the Traffic Light System:

Green Light: $\sum_{t=1}^{12} min\{G_T\} > 0$, which is expected to happen in 90 % of all cases

$G_T \in [0, 0.05)$,

Yellow Light: $\sum_{t=1}^{12} min\{G_T\} < 0$ but $\sum_{t=1}^{25} min\{G_T\} > 0$, which is expected to happen at $\sim 10\ \%$ of all cases $G_T \in [0.05, 0.1)$,

Red light: $\sum_{t=1}^{25} min\{G_T\} < 0$, which is expected to happen at 0.01 % of all cases

$G_T \in [0.1, 1]$.

If the sum of the 12 worst secured positions are positive, the model is classified with a green light. If the sum of the 12 worst secured positions are negative but the sum of the 25 worst are positive the model is classified with a yellow light. Finally, if the sum of the 25 worst secured positions are negative, the model is classified with a red light.

3.1.5 Righi and Ceretta's (2015) Truncated Distribution Tests

Righi and Ceretta (2013) propose a parametric ES backtest which uses the expectation and dispersion of a return distribution truncated by the VaR upper limit. Previous backtests use the full conditional distribution standard deviation as a dispersion measure, see e.g. McNeil and Frey (2000) and Wong (2008). Righi and Ceretta's (2015) backtest is different by only considering the dispersion conditional on a VaR exceedance. To estimate the dispersion around the expected value of the truncated return distribution, they introduce a measure they denote the Shortfall Deviation (SD). Righi and Ceretta (2015) define VaR, ES and SD as

$$VaR_t = \mu_t + \sigma_t F^{-1}(\alpha),$$
$$ES_t = \mu_t + \sigma_t E[z_t \mid z_t < F^{-1}(\alpha)],$$
$$SD_t = (\sigma_t^2\, var[z_t \mid z_t < F^{-1}(\alpha)])^{\frac{1}{2}},$$

where $F^{-1}(\alpha)$ is the inverse of a probability distribution which dictates the white noise process z_t. They argue that the SD of the truncated distribution is a superior dispersion metric than the previously proposed full distribution conditional standard deviation. It merely accounts for the extreme losses, and risk managers and legislators are primarily interested in limiting the extremely bad cases.

Righi and Ceretta (2015) suggest applying a simple model-free numerical simulation approach to estimate the values of VaR, ES and SD. Using Monte Carlo simulations, it is possible to simulate a large number of P&L from a known distribution and estimate ES and SD as the expected value and standard deviation of the P&L distribution conditional on the occurence of a VaR exceedance. The simulated values can then be tested using a standard t-test:

$$BK = \frac{X - ES}{SD}.$$

The simulations result in a large number of values of BK which are used to find the critical values. We perform the numerical simulations using a normal and Student's t-distribution to attain estimates of ES, SD and BK and by extension the critical levels used in the Traffic Light System. The practical implementation of the simulations are further explained in Section 4.3. Righi and Ceretta (2015) identify that a problem with this method is that it ignores the stylized fact of heteroscedasticity and variance clustering in financial data.

Righi and Ceretta (2015) further derive a parametric approach. This results in the analytical expressions for ES and SD with applications to the normal and the Student's t-distribution. The analytical expressions allow us to simulate the critical levels from a predictive distribution. Due to the mathematical complexity involved with the Student's t-distribution, we limit our study of the analytical approach to the standard normal distribution case.

Let Φ be the probability density function and ξ be the cumulative distribution function of the normal distribution. The truncated expectation by the superior limit theorem is then given by

$$E[X \mid X < F^{-1}(p)] = -\frac{\Phi(F^{-1}(p))}{\xi(F^{-1}(p))},$$

where F is the cumulative distribution function of the truncated P&L distribution X. If we treat X_t as equivalent to z_t, and note that $\xi^{-1}(p) = F^{-1}(p)$, we can plug it into ES equation above and derive the analytical form of the ES:

$$ES_t = \mu_t - \sigma_t \frac{\Phi(F^{-1}(p))}{p}.$$

Similarly, Right and Ceretta (2015) derive the analytical form for the SD:

$$SD_t = [\sigma_t^2(1 - \xi^{-1}(p)\frac{\Phi(\xi^{-1}(p))}{p} - (\frac{\Phi(\xi^{-1}(p))}{p})^2)]^{\frac{1}{2}}.$$

If we plug in the chosen ES significance level, $\alpha = 0.025$, we attain the analytical values of $ES_t = -2.34$ and $SD_t = 0.3416$ for the normal distribution. The values are subsequently used in a Monte Carlo Simulation to estimate values of the backtesting statistic from which we attain the critical values for the Traffic Light System.

The backtest of Righi and Ceretta (2015) follows the BK test statistic described above. Assume X is a P&L distribution for $t = 1 \ldots T$ trading days. We estimate $\widehat{ES}$ as the average losses in the evaluation window conditional on a VaR exceedance. Using the values of ES and SD from the numerical simulation and the analytical approach, we plug the values into the BK test statistic:

$$BK = \frac{\widehat{ES} - ES}{SD}.$$

The resulting value is compared with the critical levels estimated by the numerical and analytical simulation. The test is a one-sided test with the null and alternative hypotheses:

$$H_0 : \widehat{ES} = ES,$$

$$H_1 : \widehat{ES} > ES.$$

4 Methodology and Data

In this section, we will present the empirical data used in our analysis, describe the different methods used to estimate VaR and ES, and provide our methodological framework for comparing the backtests.

4.1 Data and Descriptive Statistics

We used data of the daily closing price of the US stock index S&P 500. Our data was obtained from Thomson Reuters Eikon Datastream over the period from 1 December 1965 to 18 May 2020. The data was logged, first differenced and scaled by a factor of 100. S&P 500 is one of the broadest stock indices in the world, with a long history. Consequently, the index reflects significant shocks that hit the world economy during the time period, e.g. the 1973 and 1979 oil crises, the stock market crash of 1987, the 1997 Asian financial crisis, the dot-com bubble around the turn of the millenium, the September 11 terror attacks, the 2007-2008 global financial crisis, and the COVID-19 pandemic. However, any other index, asset or asset class with a continuous loss distribution can be applied. Figure 3 displays the daily log returns of S&P 500. The plot reveals several turbulent periods, most notably the aforementioned crises. As expected, the series also exhibit volatility clustering. The daily closing price of the S&P 500 is visualized in Figure 1 in Appendix. Descriptive statistics are presented in Table 3. Notably, the log returns show excess kurtosis. Figure 4 shows the daily log returns of S&P 500 during the global financial crisis.

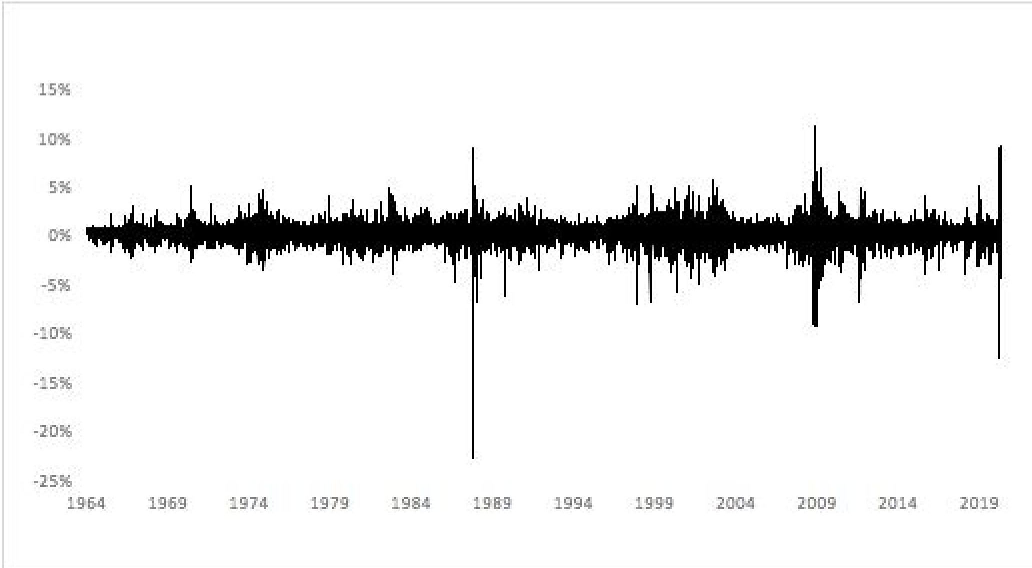

Figure 3. Daily log returns (%) of S&P 500 for our entire dataset (1/1/1964–18/5/2020).

	Raw data	Log returns (%)
No. of obs.	14708	14708
Mean	756.502	0.025
Median	409.330	0.015
Std.dev.	770.465	1.026
Skewness	1.151	-1.047
Kurtosis	0.576	28.025
Maximum	3386.150	10.957
Minimum	62.280	-29.00

Table 3. Descriptive Statistics for the returns and log returns of the stock index S&P 500 for our entire dataset (1/1/1964–18/5/2020).

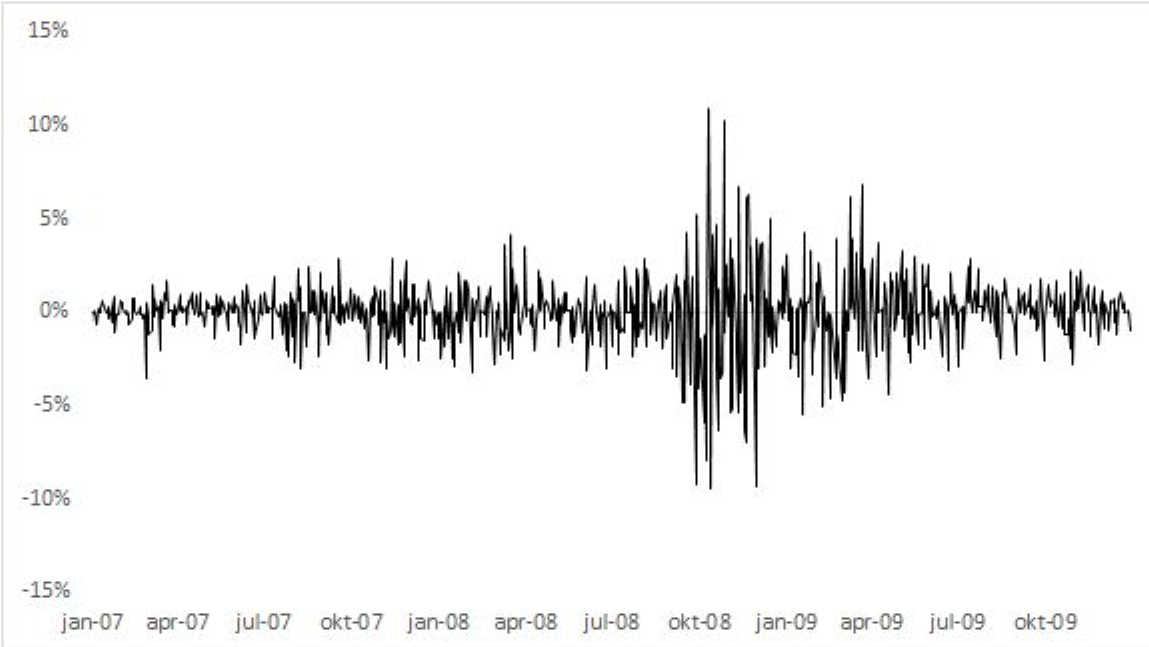

Figure 4. Daily log returns (%) of S&P 500 during the global financial crisis (1/1/2007–31/12/2009).

4.2 Estimation of VaR and ES

We estimated a 1-day ahead forecast of VaR and ES. There are various ways to do this. We used both parametric and non-parametric methods. Parametric methods rely on some distributional assumption, while non-parametric methods do not. For the parametric methods the standard normal and Student's t-distribution were used. For the non-parametric method applied Historical Simulation with 500 trading days as our estimation window. As emphasized by Righi and Ceretta (2015), the estimation window is a potential source of model risk. The Basel Committee proposes an estimation window of minimum 250 trading days (Basel Committee, 1996).

4.2.1 Parametric Estimation of VaR and ES

Normal distribution

The normal distribution relies on two parameters: the mean (μ) and the standard deviation (σ). Assuming that the return variable is normally distributed, we estimated the 1-day ahead forecasts of VaR and ES for the percentile p using the following formulas:

$$VaR^{p}_{t+1} = \mu_t + \sigma_t\, \xi^{-1}(p)\,,$$

$$ES^{p}_{t+1} = \mu_t + \sigma_t\, \frac{\Phi(\xi^{-1}(p))}{1-p}\,,$$

where μ is the estimated mean and σ_t is the standard deviation. ξ and Φ are the cumulative distribution and the probability density function of the standard normal, respectively.

Student's t-distribution

The normal distribution is generally considered an unrealistic assumption due to the stylized fact that financial data exhibits excess kurtosis and variance clustering. The Student's t-distribution converges to a normal distribution as the degrees of freedom increase. For smaller degrees of freedom, however, it allows for fat tails and asymmetry which makes it a more reasonable assumption within finance (Righi & Ceretta, 2015). We used the Student's t-distribution to estimate the 1-day ahead forecasts VaR and ES for percentile p using the following formulas:

$$VaR_{p,t} = \mu + \sqrt{\tfrac{\nu-2}{\nu}}\sigma_t t^{-1}_{p,\nu}\,,$$

$$ES_{p,t} = \mu + \sqrt{\tfrac{\nu-2}{\nu}}\sigma_t \frac{t_{tp,\nu}(t^{-1}_{p,\nu})}{1-p}\left(\frac{\nu+[\,t_{p,\nu}\,]^2}{\nu-1}\right),$$

where ν is the degrees of freedom, $t_{t_{p,\nu},\nu}$ and $t^{-1}_{p,\nu}$ are the probability density and inverse probability density function of the Student's t-distribution, respectively.

Volatility and EWMA

To estimate VaR and ES for our backtests daily estimates of the volatility are required. One possibility is to use a rolling window technique where the forecasted volatility of today is the sample volatility of the estimation window. This is simple to apply practically. However, the method suffers in that it gives equal weight to all observations in the estimation window. Consequently, the method does not incorporate the current market condition particularly well (Hull, 2018).

An alternative method is to use a geometrical decay parameter such that values in the past have less effect the further they are from today. An example of a volatility forecasting method that incorporates a decay parameter is the exponentially weighted moving average method (EWMA), first introduced by Roberts (1959). The EWMA method is given by

$$\sigma_t^2 = \lambda\sigma_{t-1}^2 + (1-\lambda)\varepsilon_{t-1}^2 ,$$

where the volatility forecast at time t is given by the volatility forecast at $t-1$ and the most recent shock, ε_{t-1}^2. $\lambda \in [0, 1]$, is a weighing parameter. Using recursive substitution, Hull (2018) shows that the equation can be written as

$$\sigma_t^2 = (1-\lambda)\sum_{i=1}^{m}\lambda^{i-1}\varepsilon_{t-i}^2 + \lambda^m\sigma_{t-m}^2 .$$

We observe that $\lambda^m\sigma_{t-m}^2 \to 0$ as $m \to \infty$. Hence, there is geometrical decay at rate λ for the past volatility forecasts. The chosen value of λ governs the responsiveness of the volatility forecast to the most recent change, where a small value puts a lot of emphasis on the most recent shock, and a larger value puts more emphasis on past volatility forecasts. We use $\lambda = 0.94$, a conventional value suggested by J.P. Morgan (1996) since it comes closest to the realised variance rate. An alternative method is to use GARCH(1,1) of which EWMA is a special case. GARCH models have the benefit of incorporating the mean reversion but requires estimation of three unknown parameters (ω, α, β).

Hence, in addition to using a standard sample volatility with an estimation window of the 500 previous days, we applied the EWMA forecasting approach with $\lambda = 0.94$ to attain a new sample of volatilities, more responsive to current market conditions.

4.2.2 Non-parametric Estimation of VaR and ES

Historical Simulation

Historical Simulation is the most widely used non-parametric VaR method (Taylor, 2008). The method requires no distributional assumptions. Instead, the estimation of VaR and ES are based on the sample of actual observed losses. Assume a P&L distribution X for T trading days . To estimate VaR and ES we divided the sample into an estimation and evaluation period. Using the P&L distribution in the estimation window the VaR and the ES was estimated for the first day in the evaluation window using the functions

$$VaR_{p,t}(X) = min\{X : Pr(X > x) \leq 1 - p\},$$

$$ES_{p,t}(X) = E[X \mid X > VaR_p(X)],$$

where VaR is the value of the P&L distribution at quantile p given our chosen significance level and ES is the expected value of the losses conditional that they are bigger than the estimated VaR. Using a rolling window technique, this process was repeated for each day in the evaluation period.

There are pros and cons with Historical Simulation. It is easy to implement and does not make any distributional assumptions. However, it is highly reliant on the sample period, and there is a risk of a "ghosting effect". That is, a very volatile/calm period remaining in the estimation window might result in too high/low values of VaR and ES. Moreover, the choice of estimation window length is subjective. A too long period might include return moments that are no longer relevant while a too short period is prone to a larger sampling error (Taylor, 2008).

Figure 5 illustrates the daily logged losses of S&P 500 and the 1-day ahead forecasts of and ES using the parametric methods with sample volatility and EWMA and Historical Simulation.

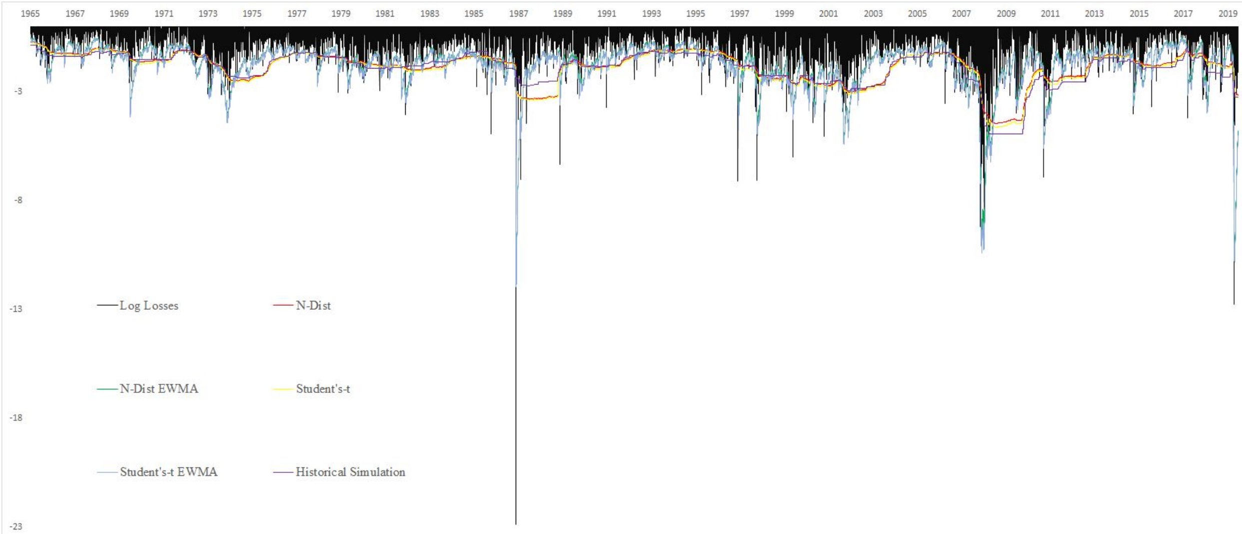

Figure 5. Daily logged losses of S&P 500 and the 1-day ahead forecasts of ES (1/12/1965–18/5/2020).

In Table 4, some descriptive statistics for the volatilities, log returns, VaR and ES for the different estimation methods are displayed. For VaR and ES, we observe that applying EWMA volatility on the normal distribution yields the smallest average losses. As expected, the EWMA estimates systematically produce the smallest minimum values, reflecting the superior ability to capture turbulent periods. Further, the larger kurtosis of the Student's t-distribution compared to the normal distribution is reflected in a higher mean for ES. Historical Simulation demonstrates the highest average ES, but the lowest maximum ES.

				Value-at-Risk					Expected Shortfall				
	σ	EWMA σ	Log returns (%)	N(0,1)	N(0,1) EWMA	Student's-t	Student's-t EWMA	Historical Simulation	N(0,1)	N(0,1) EWMA	Student's-t	Student's-t EWMA	Historical Simulation
Average	0,94	0,88	0,02	1,82	1,73	1,88	1,79	1,88	2,17	2,07	2,38	2,27	2,56
Min	0,37	0,21	-22,90	0,70	0,48	0,74	0,51	0,84	0,84	0,58	0,92	0,63	1,15
Max	2,22	6,08	10,96	4,46	11,86	4,65	11,96	4,94	5,30	14,16	5,92	17,00	6,75

Table 4. Descriptive statistics for our entire period (1/12/1965–18/5/2020). VaR and ES are measured in absolute (positive) terms.

Table 5 shows the same descriptive statistics during the global financial crisis. We observe that unlike the entire period the average EWMA is higher than the sample volatility. The parametric methods that employ EMWA display a higher VaR and ES than the sample volatility. Historical

Simulation demonstrates the highest average ES for the period while Student's t-distribution with EWMA produces the largest maximum value.

				Value-at-Risk					Expected Shortfall				
	σ	EWMA σ	Log returns (%)	N(0,1)	N(0,1) EWMA	Student's-t	Student's-t EWMA	Historical Simulation	N(0,1)	N(0,1) EWMA	Student's-t	Student's-t EWMA	Historical Simulation
Average	1,34	1,56	-0,03	2,65	3,08	2,77	3,23	3,03	3,15	3,67	3,50	4,07	4,09
Min	0,62	0,36	-9,47	1,16	0,67	1,19	0,68	1,17	1,39	0,80	1,47	0,84	1,49
Max	2,22	4,98	10,96	4,46	9,84	4,65	10,40	4,94	5,30	11,72	5,92	13,32	6,75

Table 5. Descriptive statistics for the financial crisis (1/1/2007-31/12/2009). VaR and ES are measured in absolute (positive) terms.

4.3 Monte Carlo Simulation

To derive the Traffic Light critical levels for the two tests proposed by Acerbi and Szekely (2014) and Righi and Ceretta (2015), we used a Monte Carlo Simulation approach. Monte Carlo Simulation methods are stochastic approaches that use a predetermined parametric or empirical distribution to generate a large series of stochastic paths that represent different outcomes (Konatantinos et al., 2007; Valerie Louisy-Louis, 1998; Hendricks, 1996 cited in Virdi 2011).

A general structure for our simulations is as follows:

1. We simulated 250 random numbers from either the standard normal or the Student's t-distribution with three, five or ten degrees of freedom. The distributional assumption was made a priori, hence the values of VaR and ES for the chosen level of significance (in our case 2.5%) are known beforehand (see Table 6).
2. The simulated values were treated as 250 i.i.d. trading day outcomes that we used to estimate the required parameters for the backtests.
3. The simulated outcomes, estimated and a priori known parameters was put into the test statistic. The result was one sample value of the test statistic.
4. This process was repeated 100 000 times to generate a large enough sample of the test statistic from which we derived the critical levels required for the Traffic Light System.

Degrees of freedom (v)	Critical values	
v	**VaR**	**ES**
3	-3.18	-5.04
5	-2.57	-3.52
10	-2.23	-2.82
N(0,1)	-1.96	-2.34

$\alpha = 0.025$

Table 6. Critical values of VaR and ES for the standard normal distribution and the Student's t-distribution.

For the two tests proposed by Acerbi and Szekely (2014) it is straightforward to simulate the test statistics. Assuming that the random numbers are trading day outcomes, we plugged the numbers into the test statistics and attained the corresponding test value. After the simulations the critical values were extracted as the test value sample percentile at the confidence levels specified in the Traffic Light System.

Righi and Ceretta (2015) provide us with the analytical values for the ES and SD. We used a numerical simulation approach to estimate the ES and SD for the standard normal and the Student's t-distribution with three, five and ten degrees of freedom. The simulation is a two-step approach where we first simulated random numbers from the distributions and estimated the ES and SD as the expected value and standard deviation of exceedances. The mean of the simulated ES and SD we used in the second step of the simulation are displayed in Table 7.

Analytical		Numerical							
N(0,1)		N(0,1)		Student's t ($v = 3$)		Student's t ($v = 5$)		Student's t ($v = 10$)	
ES	SD	ES	SD	ES	SD	ES	SD	ES	SD
-2.34	0.34	-2.33	0.31	-5.04	2.05	-3.51	0.92	-2.81	0.52

Table 7. The mean of the Expected Shortfall and the Shortfall Deviation from numerical simulation. The analytical values are derived from the analytical forms of the ES and the SD.

The second step is to estimate the test statistic using the analytical and simulated values. Like Acerbi and Szekely's (2014) tests we attained the critical levels by finding the percentile of the simulated test statistic values. The Traffic Light critical values for Acerbi and Szekely (2014) and Righi and Ceretta (2015) are displayed in Table 8.

4.4 The Traffic Light System

To compare the different ES backtests, we used a Traffic Light System, similar to the one introduced by the Basel Committee on Banking Supervision (1996), described in Section 2.4.3. In this way, we investigated whether the different backtests generate similar responses when ES is calculated based on empirical data on S&P 500. Following Acerbi and Szekely's (2014), Constanzino and Curran's (2018) and Moldenhauer and Pitera's (2019) ES application of the Basel Committee's VaR Traffic Light System, we used the color zone boundaries:

$$\text{Green zone: } p < 0.95\,,$$
$$\text{Yellow zone: } 0.95 \le p < 0.9999\,,$$
$$\text{Red zone: } p \ge 0.9999\,,$$

where p is the confidence level. We derived the critical values corresponding to the color zone boundaries for the different backtests. The critical values are presented in Table 8.

	N(0,1)		Student's t (v = 3)		Student's t (v = 5)		Student's t (v = 10)	
Confidence level	0.9999	0.95	0.9999	0.95	0.9999	0.95	0.9999	0.95
Acerbi and Szekely Test 1	-0.32	-0.11	-2.31	-0.42	-0.95	-0.26	-0.56	-0.17
Acerbi and Szekely Test 2	-1.8	-0.70	-4.4	-0.82	-2	-0.74	-1.90	-0.71
Constanzino and Curran Asymptotic test	9.30	5.70						
Constanzino and Curran Finite Sample test	9.88	5.70						
Righi and Ceretta Analytical test	-2.25	-0.78						
Righi and Ceretta Numerical test	-2.44	-0.88	-6.04	-1.04	-3.63	-0.99	-3.04	-0.92

Notes: Constanzino and Curran's two tests and Righi and Ceretta's Analytical test are parametric assuming normal distribution. Hence, there are no simulated critical values for the Student's t-distribution.

Table 8. Critical values for the color zone boundaries for different distributional assumptions in the Monte Carlo simulations.

For Emmer, Kratz and Tasche's (2015) Approximative Quantile Test, the color zones are given by the cumulative probabilities in Table 2, where a test is classified as green if the cumulative probabilities are green for all tests, yellow or red if any of the tests are yellow or red. For Moldenhauer and Pitera's (2019) Secured Position Test the color zones are defined as green if the sum of the 12 worst secured positions are positive, yellow if they are negative but the sum of the 25 worst secured positions are positive and red if the sum of the 25 worst secured positions are negative.

5. Empirical Results

In this section, we will present the results of the empirical evaluation of the backtests for the entire evaluation period and the subsample period during the global financial crisis.

5.1 Backtesting 1965-2020

Table 9 illustrates the proportions of green, yellow and red classifications for 13960 daily backtest for the entire evaluation period (1/12/1965–18/5/2020).

	Estimation methods														
	N(0,1)			**N(0,1) EWMA**			**Student's-t**			**Student's-t EWMA**			**Historical Simulation**		
Backtest	**Green**	**Yellow**	**Red**	**Green**	**Yellow**	**Red**	**Green**	**Yellow**	**Red**	**Green**	**Yellow**	**Red**	**Green**	**Yellow**	**Red**
Acerbi & Szekely Test 1 - N(0,1)	52,3%	35,2%	12,4%	52,3%	30,5%	17,2%	65,3%	26,5%	8,1%	66,9%	21,7%	11,4%	81,2%	16,3%	2,4%
- ν = 3	91,7%	8,3%	0,0%	89,0%	11,0%	0,0%	95,6%	4,4%	0,0%	94,4%	5,6%	0,0%	99,1%	0,9%	0,0%
- ν = 5	75,5%	23,3%	1,2%	76,8%	22,2%	1,0%	89,8%	9,1%	1,1%	85,2%	13,8%	1,0%	96,8%	2,8%	0,3%
- ν = 10	62,2%	34,0%	3,7%	61,3%	32,4%	6,3%	73,6%	23,7%	2,7%	76,0%	21,6%	2,4%	88,7%	10,9%	0,4%
Acerbi & Szekely Test 2 - N(0,1)	56,8%	20,3%	22,9%	55,4%	41,1%	3,5%	61,9%	20,3%	17,8%	68,7%	30,9%	0,4%	66,7%	21,4%	12,0%
- ν = 3	59,3%	36,1%	4,6%	62,6%	37,4%	0,0%	63,6%	34,2%	2,1%	74,4%	25,6%	0,0%	68,9%	31,0%	0,1%
- ν = 5	57,2%	21,5%	21,3%	57,6%	41,4%	1,0%	62,3%	24,0%	13,7%	70,1%	29,7%	0,2%	67,0%	22,2%	10,8%
- ν = 10	57,0%	21,1%	21,9%	56,2%	41,9%	1,8%	62,1%	23,2%	14,7%	68,9%	30,8%	0,3%	66,8%	22,0%	11,2%
Constanzino & Curran - Asymptotic	84,8%	7,7%	7,5%	92,0%	5,6%	2,4%	85,1%	8,1%	6,8%	92,4%	5,5%	2,1%	83,9%	7,1%	9,0%
Constanzino & Curran - Finite Sample	84,8%	8,8%	6,5%	92,0%	5,7%	2,3%	85,1%	9,4%	5,5%	92,4%	5,7%	1,9%	83,9%	7,7%	8,4%
Emmer et al.	49,9%	26,3%	23,8%	37,7%	55,8%	6,5%	67,9%	25,5%	6,6%	67,9%	31,7%	0,4%	59,9%	31,2%	8,8%
Moldenhauer & Pitera	53,5%	24,0%	22,5%	41,2%	52,5%	6,2%	65,3%	22,3%	12,4%	65,3%	32,4%	2,3%	69,7%	21,5%	8,8%
Righi & Ceretta Analytical - N(0,1)	73,6%	20,0%	6,4%	80,4%	14,5%	5,1%	70,7%	20,0%	9,3%	73,9%	20,0%	6,2%	66,5%	21,8%	11,6%
Righi & Ceretta Numerical - N(0,1)	73,6%	19,7%	6,7%	80,5%	14,1%	5,4%	70,7%	18,7%	10,6%	73,9%	19,6%	6,5%	66,6%	19,8%	13,6%
- ν = 3	100,0%	0,0%	0,0%	100,0%	0,0%	0,0%	100,0%	0,0%	0,0%	100,0%	0,0%	0,0%	100,0%	0,0%	0,0%
- ν = 5	97,1%	2,9%	0,0%	98,1%	1,9%	0,0%	94,4%	5,6%	0,0%	96,5%	3,5%	0,0%	92,5%	7,5%	0,0%
- ν = 10	90,6%	9,4%	0,0%	94,0%	4,9%	1,1%	86,2%	11,8%	2,0%	88,8%	9,7%	1,5%	78,6%	16,4%	5,1%

Table 9. Proportions of green, yellow and red lights for 13960 daily backtests for different estimation methods (1/12/1965–18/5/2020).

The results reveal some patterns among the backtests. Acerbi and Szekely’s Test 1 with three degrees of freedom, Constanzino and Curran’s two tests and Righi and Ceretta’s t-distributed Numerical Test show the largest proportion of green lights. The Quantile Approximation Test by

Emmer, Kratz and Tasche and The Secured Position Test by Moldenhauer and Pitera yield less than 70% green lights across all estimation methods.

Further, the tests display conflicting sensibilities to volatility type. For Acerbi and Szekely Test 2 and the tests by Constanzino and Curran, Emmer, Kratz and Tasche, and Moldenhauer and Pitera the proportion of red lights unambiguously decreases when EWMA volatilities are applied. This tendency is considerably apparent for Acerbi and Szekely Test 2. Acerbi and Szekely's first test has a tendency towards the opposite, where the proportion of red lights increases when EWMA volatilities are applied. This may be due to the fact that Test 1 is insensitive to the number of exceedances. This could possibly suggest that Test 1 is more sensitive to the magnitude of exceedances.

Table 10 shows the correlations between the different backtests performed daily over the entire evaluation period. Each correlation was calculated as the average correlation between each test across all estimation methods and distributions. Consequently, the correlation between one test and itself measures the test's stability across different estimation methods. Likewise, a correlation between two different tests is conditioned on all estimation methods. Notably, Acerbi and Szekely's Test 2, Emmer, Kratz and Tasche approximative test and Moldenhauer and Pitera's test exhibit the highest correlations (0.55-0.61). Acerbi and Szekely's Test 1, Constanzino and Curran and Righi and Ceretta exhibit weak correlations to other tests (a maximum of 0.35, 0.36 and 0.36 respectively).

Correlations

	A & S Test 1	A & S Test 2	C & C	Emmer et al.	M & P	R & C
A & S Test 1	*0.54*	...	...	...	...	...
A & S Test 2	*0.25*	*0.63*	...	...	...	...
C & C	*0.11*	*0.26*	*0.74*	...	...	...
Emmer et al.	*0.26*	*0.55*	*0.19*	*0.61*	...	...
M & P	*0.35*	*0.60*	*0.25*	*0.58*	*0.66*	...
R & C	*0.06*	*0.05*	*0.36*	*-0.01*	*0.03*	*0.38*

Table 10. Correlations between different backtests performed daily (1/12/1965–18/5/2020). Each correlation was calculated as the average correlation between each test across all estimation methods and distributional assumptions in the Monte Carlo simulations.

Tables 11-15 display the annual backtest results for different estimation methods. It is evident that the general performance differs across different estimation methods. For instance, a parametric method assuming normally distributed losses is more inclined to show green light, than a parametric method assuming Student's t-distributed losses with EWMA volatilities. Comparing the different backtests, irregardless of estimation method, we observe substantial differences in Traffic Light responses for a large number of years. Validating the output in Table 9, Acerbi and Szekely's test, Constanzino and Curran's Traffic Lights Test and Righi and Ceretta's Truncated Distribution Test generate a large proportion of green lights compared to Acerbi and Szekely's second test, Emmer, Kraz and Tasche's Quantile Approximation Test and Moldenhauer and Pitera's Secured Position Test. The proportion of red lights for Righi and Ceretta's test using simulated Student's t critical values diverges from the other tests. Notably, unlike the other tests it does not classify the period during the financial crisis as red.

	Acerbi & Szekely test 1				Acerbi & Szekely test 2				Constanzino & Curran		Emmer et al.	Moldenhauer & Pitera	Righi & Ceretta				
									Asymptotic	Finite sample			Analytical N(0,1)	Numerical N(0,1)			
	N(0,1)	v = 3	v = 5	v = 10	N(0,1)	v = 3	v = 5	v = 10						N(0,1)	v = 3	v = 5	v = 10
1966-12-30	Yellow	Green	Yellow	Yellow	Red	Red	Red	Red	Green	Green	Red	Red	Green	Green	Green	Green	Green
1967-12-29	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1968-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1969-12-31	Green	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
1970-12-31	Yellow	Green	Green	Yellow	Red	Yellow	Red	Red	Green	Green	Red	Red	Green	Green	Green	Green	Green
1971-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1972-12-29	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1973-12-31	Green	Green	Green	Green	Red	Yellow	Red	Red	Green	Green	Red	Red	Green	Green	Green	Green	Green
1974-12-31	Green	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Green	Yellow	Green	Green	Green	Green	Green
1975-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Green
1976-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1977-12-30	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1978-12-29	Green	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
1979-12-31	Yellow	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1980-12-31	Green	Green	Green	Green	Red	Yellow	Red	Red	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
1981-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1982-12-31	Yellow	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1983-12-30	Yellow	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1984-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1985-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1986-12-31	Yellow	Green	Yellow	Yellow	Red	Yellow	Red	Red	Green	Green	Red	Red	Green	Green	Green	Green	Green
1987-12-31	Red	Yellow	Yellow	Red	Red	Red	Red	Red	Yellow	Yellow	Red	Red	Green	Green	Green	Green	Green
1988-12-30	Yellow	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Red	Red	Green	Green	Yellow
1989-12-29	Red	Yellow	Yellow	Red	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Yellow
1990-12-31	Yellow	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
1991-12-31	Yellow	Green	Green	Yellow	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1992-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1993-12-31	Yellow	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1994-12-30	Yellow	Green	Green	Yellow	Yellow	Yellow	Yellow	Yellow	Green	Green	Red	Yellow	Green	Green	Green	Green	Green
1995-12-29	Yellow	Green	Green	Yellow	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1996-12-31	Red	Green	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Green	Green	Red	Yellow	Green	Green	Green	Green	Green
1997-12-31	Yellow	Green	Yellow	Yellow	Red	Yellow	Red	Red	Green	Green	Red	Red	Green	Green	Green	Green	Green
1998-12-31	Yellow	Green	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Green	Green	Green	Green	Green
1999-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Green
2000-12-29	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Yellow	Yellow	Yellow	Green	Green	Green
2001-12-31	Yellow	Green	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Green	Yellow	Yellow	Green	Green	Green
2002-12-31	Green	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Red	Red	Yellow	Yellow	Yellow	Yellow	Green	Green	Green
2003-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Yellow
2004-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Green
2005-12-30	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
2006-12-29	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Green	Green	Green	Green	Green	Green
2007-12-31	Yellow	Green	Yellow	Yellow	Red	Yellow	Red	Red	Green	Green	Red	Red	Green	Green	Green	Green	Green
2008-12-31	Red	Yellow	Yellow	Yellow	Red	Red	Red	Red	Red	Red	Red	Red	Yellow	Yellow	Green	Green	Green
2009-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Red	Red	Green	Yellow	Yellow
2010-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Red	Red	Green	Yellow	Yellow
2011-12-30	Yellow	Green	Yellow	Yellow	Red	Yellow	Red	Red	Red	Red	Red	Red	Yellow	Yellow	Green	Green	Green
2012-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Green
2013-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
2014-12-31	Yellow	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
2015-12-31	Yellow	Green	Green	Yellow	Red	Yellow	Red	Red	Green	Green	Yellow	Red	Green	Green	Green	Green	Green
2016-12-30	Yellow	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
2017-12-29	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
2018-12-31	Red	Green	Yellow	Yellow	Red	Red	Red	Red	Green	Green	Red	Red	Green	Green	Green	Green	Green
2019-12-31	Yellow	Green	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Green	Green	Green	Green	Green	Green
2020-05-18	Red	Yellow	Yellow	Red	Red	Red	Red	Red	Red	Red	Red	Red	Green	Green	Green	Green	Green

Table 11. Annual backtest results for the parametric estimation method assuming normal distribution.

	Acerbi & Szekely test 1				Acerbi & Szekely test 2				Constanzino & Curran		Emmer et al.	Moldenhauer & Pitera	Righi & Ceretta				
									Asymptotic	Finite sample			Analytical N(0,1)	Numerical N(0,1)			
	N(0,1)	v = 3	v = 5	v = 10	N(0,1)	v = 3	v = 5	v = 10						N(0,1)	v = 3	v = 5	v = 10
1966-12-30	Green	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Green	Green	Red	Yellow	Green	Green	Green	Green	Green
1967-12-29	Green	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
1968-12-31	Green	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
1969-12-31	Green	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow	Green	Green	Green	Green	Green	Green
1970-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1971-12-31	Yellow	Green	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
1972-12-29	Green	Green	Green	Green	Yellow	Green	Yellow	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
1973-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1974-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Green
1975-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1976-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1977-12-30	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Green	Green	Green	Green	Green	Green
1978-12-29	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1979-12-31	Yellow	Green	Green	Yellow	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
1980-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Green	Green	Green	Green	Green	Green
1981-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Green	Green	Green	Green	Green
1982-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1983-12-30	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1984-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1985-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1986-12-31	Yellow	Green	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
1987-12-31	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow	Red	Red	Red	Green	Green	Yellow
1988-12-30	Yellow	Green	Green	Yellow	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1989-12-29	Red	Yellow	Yellow	Red	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
1990-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Green	Green	Green	Green	Green
1991-12-31	Red	Yellow	Yellow	Red	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1992-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1993-12-31	Yellow	Green	Yellow	Yellow	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
1994-12-30	Yellow	Green	Green	Yellow	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
1995-12-29	Yellow	Green	Green	Yellow	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1996-12-31	Red	Green	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
1997-12-31	Yellow	Green	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
1998-12-31	Yellow	Green	Green	Yellow	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow	Yellow	Yellow	Yellow	Green	Green	Green
1999-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
2000-12-29	Green	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow	Yellow	Yellow	Yellow	Green	Green	Green
2001-12-31	Yellow	Green	Green	Yellow	Green	Green	Green	Green	Yellow	Yellow	Green	Yellow	Yellow	Yellow	Green	Green	Green
2002-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Red	Red	Green	Green	Yellow
2003-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
2004-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
2005-12-30	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
2006-12-29	Green	Green	Green	Green	Yellow	Green	Yellow	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
2007-12-31	Red	Green	Yellow	Yellow	Red	Yellow	Red	Red	Green	Green	Red	Red	Green	Green	Green	Green	Green
2008-12-31	Yellow	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Red	Red	Yellow	Yellow	Red	Red	Green	Yellow	Yellow
2009-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Red	Red	Green	Green	Yellow
2010-12-31	Green	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
2011-12-30	Yellow	Green	Yellow	Yellow	Red	Yellow	Yellow	Yellow	Green	Green	Yellow	Red	Yellow	Yellow	Green	Green	Green
2012-12-31	Yellow	Green	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
2013-12-31	Yellow	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
2014-12-31	Yellow	Green	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Green	Green	Red	Yellow	Green	Green	Green	Green	Green
2015-12-31	Yellow	Green	Green	Yellow	Yellow	Green	Yellow	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
2016-12-30	Red	Yellow	Red	Red	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
2017-12-29	Red	Yellow	Yellow	Yellow	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
2018-12-31	Red	Yellow	Yellow	Red	Yellow	Yellow	Yellow	Yellow	Green	Green	Red	Red	Green	Green	Green	Green	Green
2019-12-31	Red	Green	Yellow	Yellow	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
2020-05-18	Red	Green	Yellow	Yellow	Red	Yellow	Red	Red	Green	Green	Red	Red	Yellow	Yellow	Green	Green	Green

Table 12. Annual backtest results for the parametric estimation method assuming normal distribution with EWMA volatilities.

	Acerbi & Szekely test 1				Acerbi & Szekely test 2				Constanzino & Curran		Emmer et al.	Moldenhauer & Pitera	Righi & Ceretta				
									Asymptotic	Finite sample			Analytical N(0,1)	Numerical N(0,1)			
	N(0,1)	v = 3	v = 5	v = 10	N(0,1)	v = 3	v = 5	v = 10						N(0,1)	v = 3	v = 5	v = 10
1966-12-30	Yellow	Green	Green	Yellow	Red	Yellow	Red	Red	Green	Green	Red	Red	Green	Green	Green	Green	Green
1967-12-29	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1968-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1969-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1970-12-31	Green	Green	Green	Green	Red	Yellow	Yellow	Red	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
1971-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1972-12-29	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1973-12-31	Green	Green	Green	Green	Red	Yellow	Red	Red	Green	Green	Yellow	Red	Green	Green	Green	Green	Green
1974-12-31	Green	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Green	Yellow	Green	Green	Green	Green	Green
1975-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Green
1976-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1977-12-30	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1978-12-29	Green	Green	Green	Green	Yellow	Green	Yellow	Yellow	Green	Green	Green	Green	Green	Green	Green	Green	Green
1979-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1980-12-31	Green	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
1981-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1982-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1983-12-30	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1984-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1985-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1986-12-31	Yellow	Green	Green	Yellow	Red	Yellow	Yellow	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
1987-12-31	Red	Yellow	Yellow	Red	Red	Yellow	Red	Red	Red	Yellow	Yellow	Red	Yellow	Yellow	Green	Green	Green
1988-12-30	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Red	Red	Green	Yellow	Yellow
1989-12-29	Red	Yellow	Yellow	Red	Green	Green	Green	Green	Green	Green	Green	Green	Red	Red	Green	Green	Yellow
1990-12-31	Green	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Green	Green	Green	Yellow	Green	Green	Green	Green	Green
1991-12-31	Yellow	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1992-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1993-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1994-12-30	Green	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
1995-12-29	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1996-12-31	Yellow	Green	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
1997-12-31	Yellow	Green	Green	Yellow	Red	Yellow	Red	Red	Green	Green	Yellow	Red	Green	Green	Green	Green	Green
1998-12-31	Yellow	Green	Green	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Green	Green	Green	Green	Green
1999-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Green
2000-12-29	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green
2001-12-31	Yellow	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Yellow
2002-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Red	Yellow	Green	Green	Yellow	Red	Green	Green	Yellow
2003-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Red	Red	Green	Green	Yellow
2004-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Green
2005-12-30	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
2006-12-29	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Green	Green	Green	Green	Green	Green
2007-12-31	Yellow	Green	Green	Yellow	Red	Yellow	Red	Red	Green	Green	Red	Red	Green	Green	Green	Green	Green
2008-12-31	Red	Green	Yellow	Yellow	Red	Red	Red	Red	Red	Red	Red	Red	Yellow	Yellow	Green	Green	Green
2009-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Red	Red	Green	Yellow	Red
2010-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Red	Red	Green	Yellow	Yellow
2011-12-30	Yellow	Green	Green	Yellow	Red	Yellow	Yellow	Yellow	Red	Red	Yellow	Yellow	Yellow	Yellow	Green	Green	Green
2012-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Green
2013-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
2014-12-31	Green	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
2015-12-31	Yellow	Green	Green	Yellow	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
2016-12-30	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
2017-12-29	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
2018-12-31	Yellow	Green	Yellow	Yellow	Red	Red	Red	Red	Green	Green	Red	Red	Green	Green	Green	Green	Green
2019-12-31	Yellow	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
2020-05-18	Red	Yellow	Yellow	Red	Red	Yellow	Red	Red	Red	Red	Red	Red	Yellow	Yellow	Green	Green	Green

Table 13. Annual backtest results for the parametric estimation method assuming Student's t-distribution.

	Acerbi & Szekely test 1				Acerbi & Szekely test 2				Constanzino & Curran		Emmer et al.	Moldenhauer & Pitera	Righi & Ceretta				
									Asymptotic	Finite sample			Analytical N(0,1)	Numerical N(0,1)			
	N(0,1)	v = 3	v = 5	v = 10	N(0,1)	v = 3	v = 5	v = 10						N(0,1)	v = 3	v = 5	v = 10
1966-12-30	Green	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
1967-12-29	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1968-12-31	Green	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Green	Green	Green	Yellow	Green	Green	Green	Green	Green
1969-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1970-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1971-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1972-12-29	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1973-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1974-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Yellow
1975-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1976-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1977-12-30	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1978-12-29	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1979-12-31	Yellow	Green	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Green	Green	Green	Green	Green	Green
1980-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1981-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1982-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Green
1983-12-30	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1984-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1985-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1986-12-31	Yellow	Green	Yellow	Yellow	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
1987-12-31	Red	Green	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Green	Green	Green	Red	Red	Red	Green	Green	Yellow
1988-12-30	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Green
1989-12-29	Red	Green	Yellow	Yellow	Yellow	Green	Yellow	Yellow	Green	Green	Green	Yellow	Green	Green	Green	Green	Green
1990-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1991-12-31	Red	Yellow	Yellow	Yellow	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1992-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1993-12-31	Yellow	Green	Green	Yellow	Green	Green	Green	Green	Green	Green	Yellow	Green	Green	Green	Green	Green	Green
1994-12-30	Yellow	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Green	Green	Green	Yellow	Green	Green	Green	Green	Green
1995-12-29	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1996-12-31	Yellow	Green	Green	Yellow	Yellow	Green	Yellow	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
1997-12-31	Yellow	Green	Green	Yellow	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow	Yellow	Yellow	Yellow	Green	Green	Green
1998-12-31	Yellow	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Green	Green	Green	Yellow	Yellow	Yellow	Green	Green	Green
1999-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Green
2000-12-29	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Yellow
2001-12-31	Yellow	Green	Green	Yellow	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Yellow
2002-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Red	Red	Green	Green	Yellow
2003-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Green
2004-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
2005-12-30	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
2006-12-29	Green	Green	Green	Green	Yellow	Green	Green	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
2007-12-31	Yellow	Green	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
2008-12-31	Green	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Red	Red	Yellow	Yellow	Red	Red	Green	Yellow	Yellow
2009-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Red	Red	Green	Yellow	Yellow
2010-12-31	Green	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green
2011-12-30	Yellow	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow	Yellow	Yellow	Yellow	Green	Green	Green
2012-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
2013-12-31	Green	Green	Green	Green	Yellow	Green	Yellow	Yellow	Green	Green	Green	Green	Green	Green	Green	Green	Green
2014-12-31	Yellow	Green	Green	Yellow	Yellow	Yellow	Yellow	Yellow	Green	Green	Red	Yellow	Green	Green	Green	Green	Green
2015-12-31	Yellow	Green	Green	Yellow	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
2016-12-30	Red	Yellow	Red	Red	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
2017-12-29	Yellow	Green	Yellow	Yellow	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
2018-12-31	Red	Green	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
2019-12-31	Yellow	Green	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Green	Green	Green	Green	Green	Green
2020-05-18	Yellow	Green	Green	Yellow	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow	Red	Red	Red	Green	Green	Yellow

Table 14. Annual backtest results for the parametric estimation method assuming Student's-t distribution with EWMA volatilities.

	Acerbi & Szekely test 1				Acerbi & Szekely test 2				Constanzino & Curran		Emmer et al.	Moldenhauer & Pitera	Righi & Ceretta				
									Asymptotic	Finite sample			Analytical N(0,1)	Numerical N(0,1)			
	N(0,1)	v = 3	v = 5	v = 10	N(0,1)	v = 3	v = 5	v = 10						N(0,1)	v = 3	v = 5	v = 10
1966-12-30	Green	Green	Green	Green	Red	Yellow	Red	Red	Green	Green	Red	Red	Green	Green	Green	Green	Green
1967-12-29	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1968-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1969-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Green	Green	Green	Green	Green	Green
1970-12-31	Green	Green	Green	Green	Red	Yellow	Red	Red	Green	Green	Red	Yellow	Green	Green	Green	Green	Green
1971-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1972-12-29	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1973-12-31	Green	Green	Green	Green	Red	Yellow	Red	Red	Green	Green	Red	Red	Green	Green	Green	Green	Green
1974-12-31	Green	Green	Green	Green	Red	Yellow	Yellow	Red	Yellow	Yellow	Yellow	Yellow	Green	Green	Green	Green	Green
1975-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Green
1976-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1977-12-30	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1978-12-29	Green	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
1979-12-31	Yellow	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1980-12-31	Green	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
1981-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1982-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1983-12-30	Yellow	Green	Green	Yellow	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1984-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1985-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1986-12-31	Yellow	Green	Green	Yellow	Red	Yellow	Red	Red	Green	Green	Red	Red	Green	Green	Green	Green	Green
1987-12-31	Yellow	Green	Yellow	Yellow	Red	Yellow	Red	Red	Red	Red	Yellow	Red	Yellow	Yellow	Green	Green	Green
1988-12-30	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Red	Red	Green	Yellow	Red
1989-12-29	Red	Green	Yellow	Yellow	Green	Green	Green	Green	Green	Green	Green	Green	Red	Red	Green	Yellow	Yellow
1990-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Green	Green	Green	Green	Green	Green
1991-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1992-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1993-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1994-12-30	Green	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow	Green	Green	Green	Green	Green	Green
1995-12-29	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
1996-12-31	Yellow	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
1997-12-31	Yellow	Green	Green	Yellow	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
1998-12-31	Green	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Green	Green	Green
1999-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Yellow
2000-12-29	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green
2001-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Yellow
2002-12-31	Green	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Red	Red	Yellow	Green	Yellow	Yellow	Green	Green	Yellow
2003-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Yellow
2004-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Green
2005-12-30	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
2006-12-29	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Green	Green	Green	Green	Green	Green
2007-12-31	Yellow	Green	Green	Green	Red	Yellow	Red	Red	Green	Green	Red	Red	Green	Green	Green	Green	Green
2008-12-31	Yellow	Green	Green	Yellow	Red	Yellow	Red	Red	Red	Red	Red	Red	Red	Red	Green	Green	Yellow
2009-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Red	Red	Green	Yellow	Red
2010-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Red	Red	Green	Yellow	Red
2011-12-30	Green	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Red	Red	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow
2012-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Red	Red	Green	Green	Yellow
2013-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Green
2014-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
2015-12-31	Yellow	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
2016-12-30	Yellow	Green	Green	Yellow	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
2017-12-29	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
2018-12-31	Green	Green	Green	Green	Red	Yellow	Red	Red	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
2019-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Green
2020-05-18	Red	Green	Yellow	Yellow	Red	Yellow	Red	Red	Red	Red	Red	Red	Yellow	Yellow	Green	Green	Yellow

Table 15. Annual backtest results for the non-parametric estimation method applying Historical Simulation.

5.2 Backtesting The Global Financial Crisis

To further examine the backtests' performances during periods of financial stress, we considered the subsample period 1/1/2007–31/12/2009. Tables 16-20 illustrate the backtesting results for different estimation methods. One backtest was performed every 50 trading days. We observe similar patterns but also that the tests proposed by Acerbi and Szekely (2014), Emmer, Kratz and Tasche (2015) and Moldenhauer and Pitera (2019) react almost concurrently during the crisis, while Constanzino and Curran (2018) and Righi and Ceretta (2015) lag behind.

	Acerbi & Szekely test 1				Acerbi & Szekely test 2				Constanzino & Curran		Emmer et al.	Moldenhauer & Pitera	Righi & Ceretta				
									Asymptotic	Finite sample			Analytical N(0,1)	Numerical N(0,1)			
	N(0,1)	v = 3	v = 5	v = 10	N(0,1)	v = 3	v = 5	v = 10						N(0,1)	v = 3	v = 5	v = 10
2007-01-01	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Green	Green	Green	Green	Green	Green
2007-03-09	Yellow	Green	Green	Yellow	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
2007-05-18	Yellow	Green	Yellow	Yellow	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
2007-07-27	Yellow	Green	Green	Yellow	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
2007-10-05	Yellow	Green	Green	Yellow	Red	Yellow	Red	Red	Green	Green	Red	Red	Green	Green	Green	Green	Green
2007-12-14	Yellow	Green	Yellow	Yellow	Red	Yellow	Red	Red	Green	Green	Red	Red	Green	Green	Green	Green	Green
2008-02-22	Yellow	Green	Green	Yellow	Red	Red	Red	Red	Green	Green	Red	Red	Green	Green	Green	Green	Green
2008-05-02	Yellow	Green	Green	Yellow	Red	Red	Red	Red	Green	Green	Red	Red	Green	Green	Green	Green	Green
2008-07-11	Yellow	Green	Green	Yellow	Red	Red	Red	Red	Yellow	Yellow	Red	Red	Green	Green	Green	Green	Green
2008-09-19	Yellow	Green	Yellow	Yellow	Red	Yellow	Red	Red	Red	Red	Red	Red	Green	Green	Green	Green	Green
2008-11-28	Red	Green	Yellow	Yellow	Red	Red	Red	Red	Red	Red	Red	Red	Green	Green	Green	Green	Green
2009-02-06	Red	Yellow	Yellow	Yellow	Red	Red	Red	Red	Red	Red	Red	Red	Yellow	Yellow	Green	Green	Green
2009-04-17	Red	Yellow	Yellow	Yellow	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	Green	Green	Yellow
2009-06-26	Red	Yellow	Yellow	Yellow	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	Green	Green	Yellow
2009-09-04	Red	Yellow	Yellow	Yellow	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	Green	Yellow	Yellow
2009-11-13	Yellow	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Red	Red	Green	Yellow	Yellow
2009-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Red	Red	Green	Yellow	Yellow

Table 16. Backtest results for the parametric estimation method assuming normal distribution (1/1/2007–31/12/2009). One backtest is performed every 50 trading days.

	Acerbi & Szekely test 1				Acerbi & Szekely test 2				Constanzino & Curran		Emmer et al.	Moldenhaue r & Pitera	Righi & Ceretta				
									Asymptotic	Finite sample			Analytical N(0,1)	Numerical N(0,1)			
	N(0,1)	v = 3	v = 5	v = 10	N(0,1)	v = 3	v = 5	v = 10						N(0,1)	v = 3	v = 5	v = 10
2007-01-01	Green	Green	Green	Green	Yellow	Green	Yellow	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
2007-03-09	Red	Green	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
2007-05-18	Red	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
2007-07-27	Red	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Green	Green	Red	Yellow	Green	Green	Green	Green	Green
2007-10-05	Red	Green	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Green	Green	Red	Red	Green	Green	Green	Green	Green
2007-12-14	Red	Green	Yellow	Yellow	Red	Yellow	Red	Red	Green	Green	Red	Red	Green	Green	Green	Green	Green
2008-02-22	Green	Green	Green	Green	Red	Yellow	Yellow	Yellow	Green	Green	Red	Yellow	Green	Green	Green	Green	Green
2008-05-02	Green	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Red	Yellow	Yellow	Yellow	Green	Green	Green
2008-07-11	Green	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Red	Red	Yellow	Yellow	Yellow	Yellow	Green	Green	Green
2008-09-19	Green	Green	Green	Green	Red	Yellow	Yellow	Yellow	Red	Red	Red	Yellow	Yellow	Yellow	Green	Green	Green
2008-11-28	Green	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Red	Red	Yellow	Yellow	Red	Red	Green	Green	Yellow
2009-02-06	Yellow	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Red	Red	Yellow	Yellow	Red	Red	Green	Yellow	Yellow
2009-04-17	Green	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Red	Red	Yellow	Yellow	Red	Red	Green	Yellow	Red
2009-06-26	Green	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Red	Red	Yellow	Yellow	Red	Red	Green	Yellow	Red
2009-09-04	Green	Green	Green	Green	Yellow	Green	Green	Yellow	Red	Red	Green	Green	Red	Red	Green	Yellow	Red
2009-11-13	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Red	Red	Green	Yellow	Yellow
2009-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Red	Red	Green	Green	Yellow

Table 17. Backtest results for the parametric estimation method assuming normal distribution with EWMA volatilities (1/1/2007–31/12/2009). One backtest is performed every 50 trading days.

	Acerbi & Szekely test 1				Acerbi & Szekely test 2				Constanzino & Curran		Emmer et al.	Moldenhauer & Pitera	Righi & Ceretta				
									Asymptotic	Finite sample			Analytical N(0,1)	Numerical N(0,1)			
	N(0,1)	v = 3	v = 5	v = 10	N(0,1)	v = 3	v = 5	v = 10						N(0,1)	v = 3	v = 5	v = 10
2007-01-01	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Green	Green	Green	Green	Green	Green
2007-03-09	Yellow	Green	Yellow	Yellow	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
2007-05-18	Yellow	Yellow	Yellow	Yellow	Green	Green	Green	Green	Green	Green	Yellow	Green	Green	Green	Green	Green	Green
2007-07-27	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
2007-10-05	Yellow	Green	Yellow	Yellow	Red	Yellow	Red	Red	Green	Green	Red	Red	Green	Green	Green	Green	Green
2007-12-14	Yellow	Green	Yellow	Yellow	Red	Yellow	Red	Red	Green	Green	Red	Red	Green	Green	Green	Green	Green
2008-02-22	Yellow	Green	Green	Green	Red	Yellow	Red	Red	Green	Green	Red	Red	Green	Green	Green	Green	Green
2008-05-02	Yellow	Green	Green	Green	Red	Red	Red	Red	Green	Green	Red	Red	Green	Green	Green	Green	Green
2008-07-11	Yellow	Green	Green	Green	Red	Yellow	Red	Red	Red	Yellow	Red	Red	Green	Green	Green	Green	Green
2008-09-19	Yellow	Green	Green	Green	Red	Yellow	Red	Red	Red	Red	Red	Red	Green	Green	Green	Green	Green
2008-11-28	Red	Green	Green	Green	Red	Red	Red	Red	Red	Red	Red	Red	Yellow	Yellow	Green	Green	Green
2009-02-06	Red	Green	Green	Green	Red	Red	Red	Red	Red	Red	Red	Red	Yellow	Yellow	Green	Green	Yellow
2009-04-17	Red	Green	Green	Green	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	Green	Green	Yellow
2009-06-26	Red	Green	Green	Green	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	Green	Yellow	Yellow
2009-09-04	Red	Green	Green	Green	Red	Yellow	Red	Red	Red	Red	Red	Red	Red	Red	Green	Yellow	Yellow
2009-11-13	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Red	Red	Green	Yellow	Red
2009-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Red	Red	Green	Yellow	Red

Table 18. Backtest results for the parametric estimation method assuming Student's t-distribution (1/1/2007–31/12/2009). One backtest is performed every 50 trading days.

	Acerbi & Szekely test 1				Acerbi & Szekely test 2				Constanzino & Curran		Emmer et al.	Moldenhauer & Pitera	Righi & Ceretta				
									Asymptotic	Finite sample			Analytical N(0,1)	Numerical N(0,1)			
	N(0,1)	v = 3	v = 5	v = 10	N(0,1)	v = 3	v = 5	v = 10						N(0,1)	v = 3	v = 5	v = 10
2007-01-01	Green	Green	Green	Green	Yellow	Green	Green	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
2007-03-09	Red	Green	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
2007-05-18	Red	Yellow	Yellow	Yellow	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
2007-07-27	Red	Green	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
2007-10-05	Yellow	Green	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
2007-12-14	Yellow	Green	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
2008-02-22	Green	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow	Yellow	Yellow	Yellow	Green	Green	Green
2008-05-02	Green	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Green	Green	Green
2008-07-11	Green	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Red	Red	Yellow	Yellow	Yellow	Yellow	Green	Green	Green
2008-09-19	Green	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Red	Red	Yellow	Yellow	Yellow	Yellow	Green	Green	Green
2008-11-28	Green	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Red	Red	Yellow	Yellow	Red	Red	Green	Yellow	Yellow
2009-02-06	Green	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Red	Red	Yellow	Yellow	Red	Red	Green	Yellow	Red
2009-04-17	Green	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Red	Red	Yellow	Yellow	Red	Red	Green	Yellow	Red
2009-06-26	Green	Green	Green	Green	Green	Green	Green	Green	Red	Red	Green	Green	Red	Red	Green	Yellow	Red
2009-09-04	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Red	Red	Green	Yellow	Red
2009-11-13	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Red	Red	Green	Yellow	Yellow
2009-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Red	Red	Green	Yellow	Yellow

Table 19. Backtest results for the parametric estimation method assuming Student's t-distribution with EWMA volatilities (1/1/2007–31/12/2009). One backtest is performed every 50 trading days.

	Acerbi & Szekely test 1				Acerbi & Szekely test 2				Constanzino & Curran		Emmer et al.	Moldenhauer & Pitera	Righi & Ceretta				
									Asymptotic	Finite sample			Analytical N(0,1)	Numerical N(0,1)			
	N(0,1)	v = 3	v = 5	v = 10	N(0,1)	v = 3	v = 5	v = 10						N(0,1)	v = 3	v = 5	v = 10
2007-01-01	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Yellow	Green	Green	Green	Green	Green	Green
2007-03-09	Yellow	Green	Green	Yellow	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
2007-05-18	Yellow	Green	Green	Yellow	Green	Green	Green	Green	Green	Green	Yellow	Green	Green	Green	Green	Green	Green
2007-07-27	Green	Green	Green	Green	Yellow	Yellow	Yellow	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
2007-10-05	Green	Green	Green	Green	Red	Yellow	Yellow	Yellow	Green	Green	Yellow	Yellow	Green	Green	Green	Green	Green
2007-12-14	Yellow	Green	Green	Green	Red	Yellow	Red	Red	Green	Green	Red	Red	Green	Green	Green	Green	Green
2008-02-22	Green	Green	Green	Green	Red	Yellow	Red	Red	Green	Green	Red	Red	Green	Green	Green	Green	Green
2008-05-02	Green	Green	Green	Green	Red	Yellow	Red	Red	Green	Green	Red	Red	Green	Green	Green	Green	Green
2008-07-11	Green	Green	Green	Green	Red	Yellow	Red	Red	Red	Red	Red	Yellow	Green	Green	Green	Green	Green
2008-09-19	Yellow	Green	Green	Green	Red	Yellow	Red	Red	Red	Red	Red	Red	Yellow	Yellow	Green	Green	Green
2008-11-28	Yellow	Green	Green	Yellow	Red	Yellow	Red	Red	Red	Red	Red	Red	Yellow	Yellow	Green	Green	Green
2009-02-06	Yellow	Green	Green	Yellow	Red	Yellow	Red	Red	Red	Red	Red	Red	Red	Red	Green	Green	Yellow
2009-04-17	Yellow	Green	Green	Yellow	Red	Yellow	Red	Red	Red	Red	Red	Red	Red	Red	Green	Yellow	Yellow
2009-06-26	Yellow	Green	Green	Yellow	Red	Yellow	Red	Red	Red	Red	Red	Red	Red	Red	Green	Yellow	Red
2009-09-04	Yellow	Green	Green	Yellow	Red	Yellow	Red	Red	Red	Red	Yellow	Yellow	Red	Red	Green	Yellow	Red
2009-11-13	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Red	Red	Green	Yellow	Red
2009-12-31	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Red	Red	Green	Yellow	Red

Table 20. Backtest results for the non-parametric estimation method applying Historical Simulation (1/1/2007–31/12/2009). One backtest is performed every 50 trading days.

Correlations between the tests for the period are displayed in Table 21. In comparison with the entire sample, the correlations between Acerbi and Szekely's Test 2, Emmer, Kratz and Tasche, and Moldenhauer and Pitera slightly increase (from 0.55-0.61 to 0.62-0.64). Acerbi and Szekely's first test exhibits decreasing correlations to all tests, now almost uncorrelated to any other tests (-0.03-0.10). Moreover, the correlation between Constanzino and Curran and Righi and Ceretta increases from 0.36 to 0.49. Notably, Constanzino and Curran demonstrate a very high stability across different estimation methods during the global financial crisis (0.93).

Correlations

	A & S Test 1	A & S Test 2	C & C	Emmer et al.	M & P	R & C
A & S Test 1	*0.26*	...	...	...	...	...
A & S Test 2	*0.07*	*0.71*	...	...	...	...
C & C	*0.01*	*0.28*	*0.93*	...	...	...
Emmer et al.	*0.10*	*0.63*	*0.11*	*0.63*	...	...
M & P	*0.15*	*0.64*	*0.16*	*0.62*	*0.65*	...
R & C	*-0.03*	*-0.03*	*0.49*	*-0.17*	*-0.12*	*0.53*

Table 21. Correlations between different backtests performed daily (1/1/2007–31/1/2009). Each correlation was calculated as the average correlation between each test across all estimation methods and distributions.

6. Conclusion

In 2016, the Basel Committee on Banking Supervision (BCBS) prescribed a shift from VaR to ES in determining capital requirements for banks. However, while solving some issues associated with VaR, ES has been declared difficult to backtest. Still, several backtests have been proposed and considering the growing importance of the risk measure within the Basel regulatory framework, there are probably more to come.

The purpose of this essay was to determine whether different backtests for ES produce similar results. This is important to elucidate, because if different backtests produce conflicting results, the quality of a risk model and thereby the capital requirement of the bank is contingent on which particular test statistics are used.

We have backtested six different ES backtests using data of the daily closing price of the US stock index S&P 500, applying a Traffic Light Approach. We found a substantial divergence across different backtests. For the period 1/12/1965–18/5/2020, we found that Acerbi and Szekely's Test 2, Emmer, Kratz and Tasche Approximative Quantile Test and Moldenhauer and Pitera Secured Position Test are the highest correlated tests (0.55-0.61). Acerbi and Szekely's Test 1, Constanzino and Curran's Traffic Light Tests and Righi and Ceretta's Truncated Distribution Tests exhibit weak correlations to other tests (a maximum of 0.35, 0.36 and 0.36 respectively). Also, we found diverse sensitivity to volatility type among different tests.

To further examine the backtests' performances during periods of financial stress we performed backtests for the period 1/1/2007–31/12/2009. We found that the correlations remain stable or slightly increase during the financial crisis. Moreover, the approach proposed by Constanzino and Curran displays a very high stability across different estimation methods. Righi and Ceretta's tests demonstrate a disparate proportion of green lights compared to the other tests.

In the light of the conflicting performances of the various backtests, we conclude that regulators and practitioners need to be careful when choosing a backtest. We recommend diversification between several tests. In particular, tests that are weakly correlated are suitable pairings. If only one test must be chosen, we suggest using Constazino and Curran's approach due to its stability across different estimation methods.

7. Discussion and Further Research

In this essay, we showed that different backtests for ES produce different results. We examined six different backtests. However, these do not encompass all the proposed backtests in the literature. A more comprehensive comparison would be illuminating. Also, it is important to bear in mind that a backtesting procedure involves many subjective decisions. The choice of estimation method, volatility forecasting method and estimation window length affect the results.

We noticed an anomaly with regard to Acerbi and Szekely's (2014) first test. The backtesting results indicate that the test is sensitive to the magnitude of the losses. One possible way to build upon our analysis is a more comprehensive simulation analysis of the backtests under different scenarios. This could systematically expose different backtests' performance under different market conditions. Furthermore, a study of the size and power of all backtests would be an interesting extension.

We have primarily focused on the empirical evaluation of backtests for ES. However, there is an ongoing debate on the mathematical property elicitability and its relation to backtestability. The relevance of elicitability in this context is crucial to determine. For now, ES remains the leading candidate to replace VaR as the standard risk measure used for regulatory purposes. A coherent and elicitable alternative risk measure to VaR and ES is expectiles. This is an important area of future research. Whichever risk measure chosen by practitioners and supervisors, we urge researchers to further illuminate us on this very relevant and fascinating topic.

References

Acerbi, C. and Szekely, B., 2014. Back-testing expected shortfall. *Risk*, *27*(11), pp.76-81

Acerbi, C. and Szekely, B., 2017. General properties of backtestable statistics. *Available at SSRN 2905109*

Alexander, C., 2009. Market Risk Analysis, Value at Risk Models (Vol. 4). John Wiley & Sons

Artzner, P., Delbaen, F., Eber, J.M. and Heath, D., 1999. Coherent measures of risk. *Mathematical finance*, *9*(3), pp.203-228

Basel Committee, 1996. Supervisory framework for the use of backtesting in conjunction with the internal models approach to market risk capital requirements. *Basel Committee on Banking and Supervision, Switzerland*, Available online: https://www.bis.org/publ/bcbs22.pdf [Accessed 10 May 2020]

Basel Committee, 2013. Fundamental review of the trading book: A revised market risk framework. Consultative Document, October, Available online: https://www.bis.org/publ/bcbs265.pdf [Accessed 10 May 2020]

Basel Committee, 2016. Minimal capital requirements for market risk. Standards, January, Available online: https://www.bis.org/bcbs/publ/d352.pdf [Accessed 10 May 2020]

Berkowitz, J., 2001. Testing density forecasts, with applications to risk management. *Journal of Business & Economic Statistics*, *19*(4), pp.465-474

Campbell, S.D., 2005. A review of backtesting and backtesting procedures. *Journal of Risk 9(2)*, pp.1-17

Christoffersen, P.F., 1998. Evaluating interval forecasts. *International economic review*, pp.841-862

Christoffersen, P. (2010). Backtesting. *Encyclopedia of Quantitative Finance*

Christoffersen, P. and Pelletier, D., 2004. Backtesting value-at-risk: A duration-based approach. *Journal of Financial Econometrics, 2*(1), pp.84-108

Costanzino, N. and Curran, M., 2015. Backtesting general spectral risk measures with application to expected shortfall. *Available at SSRN 2514403*

Costanzino, N. and Curran, M., 2018. A simple traffic light approach to backtesting expected shortfall. *Risks*, *6*(1), p.2

Culp, C.L., Miller, M.H. and Neves, A.M., 1998. Value at risk: Uses and abuses. *Journal of Applied Corporate Finance*, *10*(4), pp.26-38

Datastream. (2019) *Thomson Reuters Eikon Datastream*, Available online: Subscription Service [Accessed 16 May 2020]

Du, Z., & Escanciano, J. C. (2017). Backtesting expected shortfall: accounting for tail risk. *Management Science*, *63*(4), 940-958

Dumitrescu, E.I., Hurlin, C. and Pham, V., 2012. Backtesting value-at-risk: from dynamic quantile to dynamic binary tests. *Finance*, *33*(1), pp.79-112

Emmer, S., Kratz, M. and Tasche, D., 2015. What is the best risk measure in practice? A comparison of standard measures. *Journal of Risk*, *18*(2)

Engle, R.F. and Manganelli, S., 2004. CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, *22*(4), pp.367-381

Fissler, T., Ziegel, J.F. and Gneiting, T., 2015. Expected Shortfall is jointly elicitable with Value at Risk-Implications for backtesting. *arXiv preprint arXiv:1507.00244*

Gneiting, T., 2011. Making and evaluating point forecasts. *Journal of the American Statistical Association*, *106*(494), pp.746-762

Graham, A. and Pal, J., 2014. Backtesting value-at-risk tail losses on a dynamic portfolio. *The Journal of Risk Model Validation*, *8*(2), p.59

Haas, M., 2001. New methods in backtesting. *Financial Engineering Research Center, Bonn*

Hong, L.J., Hu, Z. and Liu, G., 2014. Monte Carlo methods for value-at-risk and conditional value-at-risk: a review. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, *24*(4), pp.1-37

Hull, J.C, 2018. 'Volatility' in Hull, J.C., Risk Management and Financial Institutions 5th edn., Wiley

Hurlin, C. and Tokpavi, S., 2006. Backtesting Value-at-Risk Accuracy: A New Simple Test (No. halshs-00257461)

Jorion, P., 2006. Value at risk: the new benchmark for controlling market risk. Chicago: IRWIN

Jorion, P., 2007. Financial risk manager handbook (Vol. 406). John Wiley & Sons

Kerkhof, J. and Melenberg, B., 2004. Backtesting for risk-based regulatory capital. *Journal of Banking & Finance*, *28*(8), pp.1845-1865

Kratz, M., Lok, Y. H., & McNeil, A. J. (2018). Multinomial VaR backtests: A simple implicit approach to backtesting expected shortfall. *Journal of Banking & Finance*, *88*, 393-407

Kupiec, Paul. (1995). Techniques for Verifying the Accuracy of Risk Measurement Models. *The Journal of Derivatives, 3*, 73-84

McNeil, A.J. and Frey, R., 2000. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of empirical finance*, *7*(3-4), pp.271-300

Moldenhauer, F. and Pitera, M., 2019. Backtesting Expected Shortfall: a simple recipe?. *Journal of Risk*, *22*(1)

Morgan, J.P., 1996. Riskmetrics technical document, Available online: https://www.msci.com/documents/10199/5915b101-4206-4ba0-aee2-3449d5c7e95a [Accessed 8 May 2020]

Newey, W.K. and Powell, J.L., 1987. Asymmetric least squares estimation and testing. *Econometrica: Journal of the Econometric Society*, pp.819-847

Novales, A. and Garcia-Jorcano, L., 2019. Backtesting extreme value theory models of expected shortfall. *Quantitative Finance*, *19*(5), pp.799-825

Pajhede, T., 2015. Backtesting Value-at-Risk: A Generalized Markov Framework. *Univ. of Copenhagen Dept. of Economics Discussion Paper*, (15-18)

Patton, A.J., Ziegel, J.F. and Chen, R., 2019. Dynamic semiparametric models for expected shortfall (and value-at-risk). *Journal of econometrics*, *211*(2), pp.388-413

Rappoport, P., 1993. A new approach: Average Shortfall. *JP Morgan Fixed Income Research Technical Document*

Righi, M. and Ceretta, P.S., 2013. Individual and flexible expected shortfall backtesting. *Journal of Risk Model Validation*, *7*(3), pp.3-20

Righi, M.B. and Ceretta, P.S., 2015. A comparison of Expected Shortfall estimation models. *Journal of Economics and Business*, *78*, pp.14-47

Roberts, S.W, 1959. Control Chart Tests Based on Geometric Moving Averages. *Technometrics,* 42, pp 97-101

Taylor, J.W., 2008. Estimating value at risk and expected shortfall using expectiles. *Journal of Financial Econometrics*, *6*(2), pp.231-252

Virdi, N.K., 2011. A review of backtesting methods for evaluating value-at-risk. *International Review of Business Research Papers*, *7*(4), pp.14-24

Wong, W. K. (2008). Backtesting trading risk of commercial banks using expected shortfall. *Journal of Banking & Finance*, *32*(7), 1404-1415

Yamai, Y. and Yoshiba, T., 2002. On the validity of value-at-risk: comparative analyses with expected shortfall. *Monetary and economic studies*, *20*(1), pp.57-85

Zhang, Y. and Nadarajah, S., 2018. A review of backtesting for value at risk. *Communications in Statistics-Theory and methods*, *47*(15), pp.3616-3639

Appendix

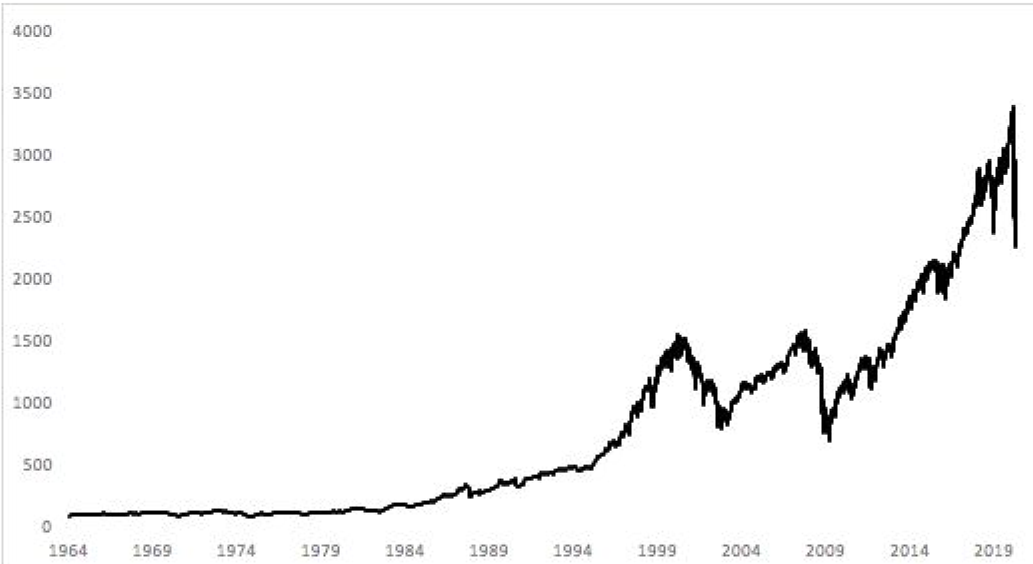

Figure 1: Daily closing price of S&P 500 (1/1/1964–18/5/2020).