

Discussion Paper

Model Risk of Expected Shortfall

November 2017

Emese Lazar

ICMA Centre, Henley Business School, University of Reading

Ning Zhang

ICMA Centre, Henley Business School, University of Reading

The aim of this discussion paper series is to disseminate new research of academic distinction. Papers are preliminary drafts, circulated to stimulate discussion and critical comment. Henley Business School is triple accredited and home to over 100 academic faculty, who undertake research in a wide range of fields from ethics and finance to international business and marketing.

admin@icmacentre.ac.uk

www.icmacentre.ac.uk

© Lazar and Zhang, November 2017

Model Risk of Expected Shortfall

Emese Lazar* and Ning Zhang[†]

November, 2017

Abstract

In this paper we study the model risk of Expected Shortfall (ES), extending the results of Boucher et al. (2014) on model risk of Value-at-Risk (VaR). We propose a correction formula for ES based on passing three backtests. Our results show that for the DJIA index, the smallest corrections are required for the ES estimates built using GARCH models. Furthermore, the 2.5% ES requires smaller corrections for model risk than the 1% VaR, which advocates the replacement of VaR with ES as recommended by the Basel Committee. Also, if the model risk of VaR is taken into account, then the correction made to ES estimates reduces by 50% on average.

Keywords: model risk, Expected Shortfall, backtesting.

JEL classification: C15, C22, C52, C53, G15.

*Correspondence to: Emese Lazar, ICMA Centre, Henley Business School, University of Reading, Whiteknights, Reading RG6 6BA, UK; e.lazar@icmacentre.ac.uk

[†]ICMA Centre, Henley Business School, University of Reading, Whiteknights, Reading RG6 6BA, UK; N.Zhang3@pgr.reading.ac.uk

1 Introduction

For risk managers and the regulatory authority, risk estimates like Value-at-Risk (VaR) and Expected Shortfall (ES)¹ are of great importance, since an accurate risk estimate can help calculate an appropriate buffer against unfavorable events. The asset prices are observable when the transactions are made, whilst risk estimates computed from various models are not materialized. The choice of the model is often subjective, leading to the possible inadequacy of the risk estimates. That is, model risk in risk models gives rise to inaccurate risk estimates.

The model risk of risk measures can be owed to misspecification of the underlying model (Cont, 2006), inaccuracy of parameter estimation (Berkowitz and O'Brien, 2002), or the use of inappropriate models (Dánielsson et al., 2016; Alexander and Sarabia, 2012). Kerkhof et al. (2010) decompose model risk into estimation risk, misspecification risk and identification risk².

In order to improve the accuracy of VaR estimates, the quantification of VaR model risk is followed by the adjustments of VaR estimates as discussed in several studies. One of the earliest works is Hartz et al. (2006), considering estimation error only, and the size of adjustments is based on a data-driven method. Alexander and Sarabia (2010)³, who propose a framework on quantifying VaR model risk and correcting VaR estimates, make the correction for estimation and specification errors mainly based on probability shifting. Complimentary to the research of Kerkhof et al. (2010), in which adjustments to VaR estimates are obtained based on regulatory backtesting measures, Boucher et al. (2014)⁴ suggest a correction for VaR model risk based on passing VaR backtests. Furthermore, they propose that their generalized methodology can be applied to correct ES for model risk. Using Taylor's expansion, Barrieu and Ravanelli (2015) derive the upper bound of the VaR adjustments, only taking specification error into account, whilst Farkas et al. (2016) derive confidence intervals for VaR and Median Shortfall and propose a test for model validation based on extreme losses.

¹An alternative is Median Shortfall (MS), as in So and Wong (2012).

²Estimation risk refers to the uncertainty of parameter estimates. Misspecification risk is the risk associated with inappropriate assumptions of the risk model, whilst identification risk refers to the risk that future sources of risk are not currently known and included in the model.

³Alexander and Sarabia (2010) assume that AGARCH(1,1) is the 'true' DGP and compute VaRs according to three different models: AGARCH, EWMA and RiskMetrics 'regulatory'.

⁴The authors assume (a) normal GARCH(1,1) as both the DGP and the estimated model to compute estimation error and (b) MS(2)-GARCH(1,1)-t as the DGP and normal GARCH(1,1) as the estimated model to compute estimation and specification errors jointly.

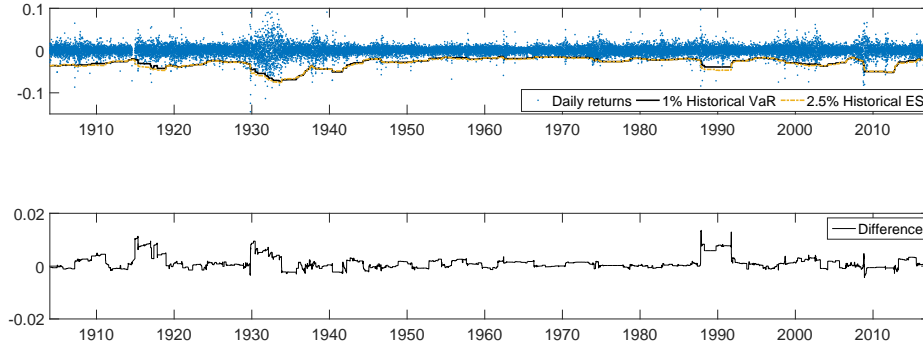


Figure 1: DJIA index daily returns, the daily historical VaR estimates ($\alpha = 1\%$) and the daily historical ES estimates ($\alpha = 2.5\%$) from 28/12/1903 to 23/05/2017, as well as the difference between the 2.5% historical ES and the 1% historical VaR are presented. We use a four-year rolling window to compute the risk estimates.

Though VaR has been a popular regulatory measure for many years, several shortcomings of the VaR measure are identified, most importantly that it is not a coherent measure as shown by Artzner et al. (1999). Recently, the Basel Committee recommends and advocates the use of ES (Basel, 2012). Specifically, they proposed the replacement of VaR at 1% significance level with ES at 2.5% significance level (Embrechts et al., 2014). With ES gaining more and more importance, the accuracy of the estimated ES becomes vital. To the best of our knowledge, no research on the adjustments of ES estimates for model risk has been conducted.

Figure 1 shows the DJIA index daily returns from 28/12/1903 to 23/05/2017, compared with the daily historical VaR and ES estimates with significance levels at 1% and 2.5% respectively. During the crisis periods, the difference between the historical ES and VaR becomes wider, which supports the replacement of the VaR measure with the ES measure; nevertheless, the clustering of exceptions when ES is violated is still noticeable. In other words, the historical ES does not react to adverse changes immediately when the market returns worsen and also it does not make effective adjustments when the market apparently goes back to normal.

An example is the time period around the 2007 financial crisis, presented in **Figure 2**, which shows the peaked-over-ES ($\alpha = 2.5\%$) and three tiers of corrections (labelled as #1, #2 and #3 on the right-hand side) made to the historical daily ES estimates ($\alpha = 2.5\%$), using a one-year rolling window. Adjustment #1 with a magnitude of 0.005 (about 18% in relative terms) added to the daily ES estimates can

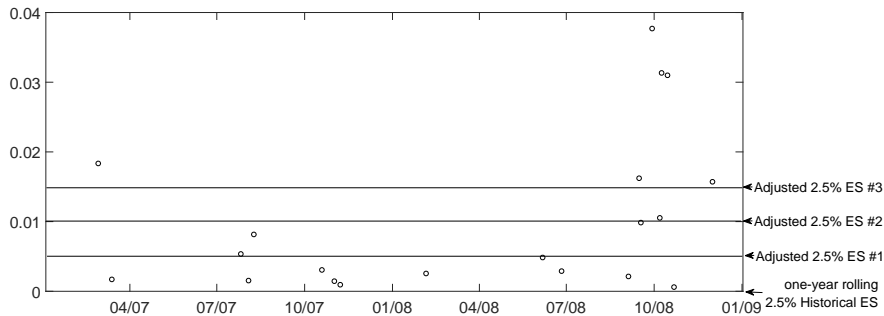


Figure 2: Peaked-over-ES and adjustments, based on the DJIA index from 01/01/2007 to 01/01/2009. One-year moving window is used to forecast daily historical ES ($\alpha = 2.5\%$).

avoid most of the exceptions that occur during this crisis. The higher the adjustment level (#2 and #3), the higher the protection from extreme losses, but even an adjustment of 0.015 (adjustment #3) still has too many exceptions. However, too much protection is not favorable to risk managers. This implies that ES estimates could be more effective when adjusted with an appropriate buffer against the model risk of the forecasted ES calculated using a specified model. The question arises how large the magnitude of the correction made to the ES estimates should be in order to ensure the accuracy of the ES estimates (not too large or too small).

We extend the research on model risk of VaR of Boucher et al. (2014) to investigate model risk of ES and also make corrections to ES estimates, thus improving their accuracy. A desirable ES estimate should satisfy three criteria: one referring to the expected number of exceptions, one regarding the absence of violation clustering, and one about the appropriate size of exceptions. By passing different ES backtests introduced by Du and Escanciano (2016) and Acerbi and Szekely (2014), we can compute the optimal size of correction made to the ES estimates, to accommodate for model risk.

The structure of the paper is as follows: section 2 analyzes the sources of ES model risk focusing on estimation and specification errors of ES estimates, and performs Monte Carlo simulations to quantify them; section 3 proposes a backtesting-based correction methodology for ES and also considers the impact of VaR model risk on model risk of ES; section 4 presents the empirical study and section 5 concludes.

2 Analysis of model risk of Expected Shortfall

2.1 Sources of model risk

Before investigating the sources of ES model risk, we first establish a general scheme (see **Figure 3**) in which the sources of model risk of risk estimates are shown. Consider a portfolio affected by risk factors, and the goal is to compute risk estimates such as VaR and ES. The first step is the identification of risk factors, and this process is affected by identification risk, which arises when some risk factors are not identified, with a very high risk of producing inaccurate risk estimates. The next step is the specification of risk factor models which, again, will have a large effect on the estimation of risk. This is followed by the estimation of the risk factor model (this, in our view, has a medium effect on the risk estimate). In step 3, the relationship between the portfolio P&L and the risk factors is considered and the formulation of this model will have a high effect on the estimation of the risk. The estimation of this will have a medium effect on the risk estimation. Step 4 links the risk estimation with the dependency of the P&L series on the risk factors.

For example, when computing the VaR of a portfolio of derivatives, step 1 would identify the sources of risk, step 2 would specify and estimate the models describing these risk factors (underlying assets most importantly), step 3 would model the P&L of the portfolio as a function of the risk factors, and in step 4 the risk model would transform P&L values into risk estimates.

The diagram shows that the main causes of model risk of risk estimates are (1) identification error, (2) model estimation error (for the risk factor model, the P&L model or the risk model), which arises from the estimation of the parameters of the model and (3) model specification error (for the risk factor model, the P&L model or the risk model), which arises when the true model is not known. Other sources of model risk that may give wrong risk estimates are, for example, granularity error, measurement error and liquidity risk (Boucher et al., 2014).

2.2 Bias and correction of Expected Shortfall

In fact, most academic research on the adequacies of risk models mainly focuses on two of the sources of model risk: estimation error and specification error. Referring to Boucher et al. (2014), the theoretical results about the two sources of VaR model risk are presented in **Appendix A**. In a similar vein, we investigate the impact of

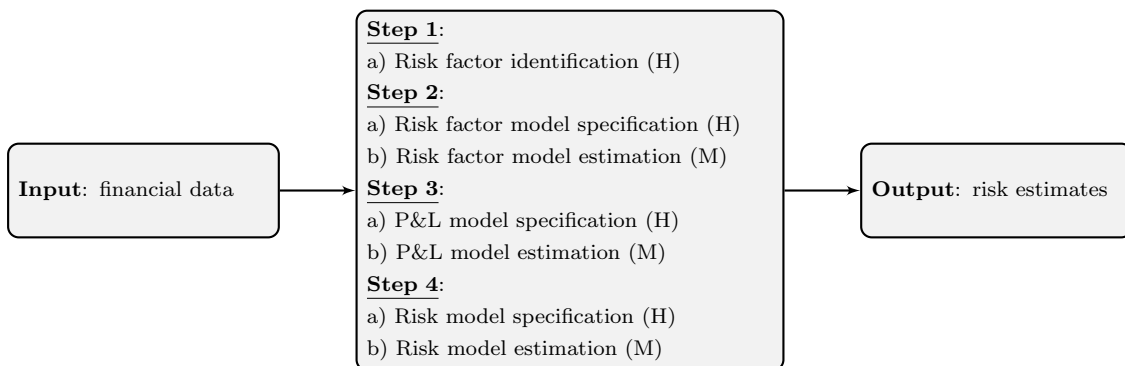


Figure 3: Risk estimation process

Notation: H and M represent high and medium impacts on risk estimates, respectively.

the earlier mentioned two errors on the ES estimates, thus deriving the theoretical formulae for estimation and specification errors, as well as correction of ES. VaR⁵, for a given distribution function F and a given significance level α , is defined as:

$$VaR_t(\alpha) = -inf\{q : F(q) \geq \alpha\}, \quad (1)$$

where q denotes the quantile of the cumulative distribution F . ES, as an absolute downside risk measure, measures the average losses exceeding VaR, taking extreme losses into account; it is given by:

$$ES_t(\alpha) = \frac{1}{\alpha} \int_0^\alpha VaR_t(u) du \quad (2)$$

Estimation bias of Expected Shortfall

Assuming that the data generating process (DGP), a model with a cumulative distribution F for the returns, is known and the true parameter values (θ_0) of this ‘true’ model are also known, the theoretical VaR, denoted by $ThVaR(\theta_0, \alpha)$ and the theoretical ES, denoted by $ThES(\theta_0, \alpha)$, both at a significance level α , can be computed as:

$$ThVaR(\theta_0, \alpha) = -q_\alpha^F = -F_\alpha^{-1} \quad (3)$$

$$ThES(\alpha) = \frac{1}{\alpha} \int_0^\alpha ThVaR(\theta_0, u) du \quad (4)$$

⁵The values of VaR and ES are considered positive in this paper.

Now, we assume that the DGP is known, but the parameter values are not known. The estimated VaR in this case is denoted by $VaR(\hat{\theta}_0, \alpha)$, where $\hat{\theta}_0$ is an estimate of θ_0 . The relationship between the theoretical VaR and the estimated VaR is:

$$ThVaR(\theta_0, \alpha) = VaR(\hat{\theta}_0, \alpha) + bias(\theta_0, \hat{\theta}_0, \alpha) \quad (5)$$

We also have that:

$$ThVaR(\theta_0, \alpha) - \mathbb{E}(VaR(\hat{\theta}_0, \alpha)) = \mathbb{E}(bias(\theta_0, \hat{\theta}_0, \alpha)) \quad (6)$$

where $\mathbb{E}[bias(\theta_0, \hat{\theta}_0, \alpha)]$ denotes the mean bias of the estimated VaR from the theoretical VaR as a result of model estimation error. Based on this, we can write the estimation bias of $ES(\hat{\theta}_0, \alpha)$, and we have that

$$ThES(\theta_0, \alpha) - \mathbb{E}[ES(\hat{\theta}_0, \alpha)] = \frac{1}{\alpha} \int_0^\alpha \mathbb{E}[bias(\theta_0, \hat{\theta}_0, v)] dv, \quad (7)$$

Ideally, correcting for the estimation bias, the ES estimate, denoted by $ES(\hat{\theta}_0, \alpha)$, can be improved as below:

$$ES^E(\hat{\theta}_0, \alpha) = ES(\hat{\theta}_0, \alpha) + \frac{1}{\alpha} \int_0^\alpha \mathbb{E}[bias(\theta_0, \hat{\theta}_0, v)] dv \quad (8)$$

Specification and estimation biases of Expected Shortfall

However, in most cases the 'true' DGP is not known, and the returns are assumed to follow a different model giving a cumulative distribution (\hat{F}) for the returns with estimated parameter values $\hat{\theta}_1$, where θ_0 and $\hat{\theta}_1$ can have different dimensions depending on the models used and their values are expected to be different. This gives the following value for the estimated VaR:

$$VaR(\hat{\theta}_1, \alpha) = -q_\alpha^{\hat{F}} = -\hat{F}_\alpha^{-1} \quad (9)$$

The relationship between the true VaR and the estimated VaR is given as:

$$ThVaR(\theta_0, \alpha) = VaR(\hat{\theta}_1, \alpha) + bias(\theta_0, \theta_1, \hat{\theta}_1, \alpha) \quad (10)$$

where θ_1 and $\hat{\theta}_1$ have the same dimension under the specified model, but θ_1 denotes the true parameter values different from the estimated parameter values of $\hat{\theta}_1$.

Similarly:

$$ThVaR(\theta_0, \alpha) - \mathbb{E}(VaR(\hat{\theta}_1, \alpha)) = \mathbb{E}(bias(\theta_0, \theta_1, \hat{\theta}_1, \alpha)) \quad (11)$$

where $\mathbb{E}[bias(\theta_0, \theta_1, \hat{\theta}_1, \alpha)]$ denotes the mean bias of the estimated VaR from the theoretical VaR as a result of model specification and estimation error. According to equation (2), the mean estimation and specification biases of ES can be formulated as below:

$$ThES(\theta_0, \alpha) - \mathbb{E}[ES(\hat{\theta}_1, \alpha)] = \frac{1}{\alpha} \int_0^\alpha \mathbb{E}[bias(\theta_0, \theta_1, \hat{\theta}_1, v)] dv \quad (12)$$

Correcting for these biases, the estimated ES, denoted by $ES(\hat{\theta}_1, \alpha)$, can be improved as:

$$ES^{SE}(\hat{\theta}_1, \alpha) = ES(\hat{\theta}_1, \alpha) + \frac{1}{\alpha} \int_0^\alpha \mathbb{E}[bias(\theta_0, \theta_1, \hat{\theta}_1, v)] dv \quad (13)$$

2.3 Monte Carlo simulations

In the following, we simplify the risk estimation process (**Figure 3**) so that only one risk factor exists. Thus, the identification risk and the P&L model specification and estimation risks are not modelled, and we are left with the specification and estimation risks for the risk factor model and, consequently, for the risk model, namely steps 2 and 4. Following the theoretical formulae for estimation and specification errors of the ES estimates, Monte Carlo simulations are implemented to investigate the impacts of these two errors on the estimated ES.

We simulate the daily return series assuming a model, thus knowing the theoretical ES. Then, the parameters are estimated using the same model as specified to generate the daily returns, giving the value of the estimation bias of ES, as in equation (7). We also estimate ES based on other models to examine the values of joint estimation and specification biases of ES, as in equation (12).

In our setup, a GARCH(1,1) model with normal disturbances (GARCH(1,1)-N) is assumed as the ‘true’ data generating process, given by:

$$r_t = \mu + \varepsilon_t \quad (14)$$

$$\varepsilon_t = \sigma_t \cdot z_t, \quad z_t \sim N(0, 1) \quad (15)$$

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \quad (16)$$

Table 1: Simulated bias associated with the ES estimates

Significance levels	Mean estimated ES(%)	Theoretical ES(%)	Mean bias(%)	Std. err of bias(%)
<i>Panel A. GARCH(1,1)-N DGP with estimated GARCH(1,1)-N ES: estimation bias</i>				
$\alpha=5\%$	23.82	23.83	0.01	1.73
$\alpha=2.5\%$	28.50	28.51	0.01	1.94
$\alpha=1\%$	34.07	34.08	0.01	2.20
<i>Panel B. GARCH(1,1)-N DGP with historical ES: specification and estimation biases</i>				
$\alpha=5\%$	28.92	23.83	-5.09	15.79
$\alpha=2.5\%$	36.38	28.51	-7.87	18.97
$\alpha=1\%$	45.77	34.08	-11.69	23.16
<i>Panel C. GARCH(1,1)-N DGP with Gaussian Normal ES: specification and estimation biases</i>				
$\alpha=5\%$	26.27	23.83	-2.44	14.86
$\alpha=2.5\%$	31.27	28.51	-2.76	16.84
$\alpha=1\%$	37.23	34.08	-3.15	19.20
<i>Panel D. GARCH(1,1)-N DGP with EWMA ES: specification and estimation biases</i>				
$\alpha=5\%$	21.68	23.83	2.15	2.54
$\alpha=2.5\%$	26.31	28.51	2.20	2.87
$\alpha=1\%$	31.82	34.08	2.26	3.28

Note: The results are based on the DJIA index from 01/01/1900 to 23/05/2017. These statistics are computed from simulating 1,000 paths of 1,000 daily returns according to the DGP of GARCH(1,1)-N. Then we forecast ES based on the GARCH(1,1)-N, historical, Gaussian Normal and EWMA ($\lambda = 0.94$) specifications, for $\alpha = 5\%$, 2.5% and 1% .

Using real data, we first estimate the parameters⁶ of this model. Next, we simulate 1,000 paths of 1,000 daily returns, compute one-step ahead ES forecasts under several different models and compare these forecasts with the theoretical ES.

The purpose of Monte Carlo simulations is to investigate the bias of ES estimates, that is, the perfect corrections for the model risk of ES forecasts. **Table 1** presents the annualized ES estimates at 5%, 2.5% and 1% in the second column, the annualized theoretical ES in the third column, the mean estimation bias of ES forecasts in **Panel A**, and the mean estimation and specification biases of the estimated ES depending on different model specifications in **Panel B, C and D**.

We compare the theoretical ES of the simulations with the estimated ES based on the GARCH(1,1)-N model in **Panel A**. The mean estimation bias is close to 0 for the 5%, 2.5% and 1% ES estimates. Also, the estimation bias can be reduced by increasing the size of the estimation period as suggested by Du and Escanciano (2016). The standard error of the bias decreases, as the value of α increases, as expected. In **Panel B**, the mean specification and estimation biases is computed from

⁶The parameters of GARCH(1,1)-N estimated from the DJIA index (1st Jan 1900 to 23rd May 2017) are : $\mu = 4.4521e^{-04}$; $\omega = 1.3269e^{-06}$; $\alpha = 0.0891$; and $\beta = 0.9017$.

the theoretical ES and the historical ES. The negative values of the bias show that the estimated ES is more conservative than the theoretical ES, whilst the positive values of the bias refer to an estimated ES lower than the theoretical ES. **Panel C** examines the specification and estimation biases of the Gaussian Normal ES estimates. In this case, the Gaussian Normal ES estimates are more conservative than the theoretical ES. The specification and estimation biases of ES estimates computed from EWMA is positive as shown in **Panel D**, which requires a positive adjustment to be added to the EWMA ES estimates.

Furthermore, the specification and estimation biases in **Panel B, C and D** are much higher than the estimation bias in **Panel A** in absolute values, which indicates that the specification error has a bigger importance than the estimation error. Overall, based on the results in the table, we conclude that an adjustment is needed to correct for the model risk of ES estimates.

3 Backtesting-based correction of ES

3.1 Backtesting-based correction methodology for ES

If a data generating process is known, then it is straightforward to compute the model risk of ES, as in **Table 1**, based on Monte Carlo simulations. In a realistic setup, the ‘true’ model is unknown, so we are not able to measure model risk directly. By correcting the estimated ES and forcing it to pass backtests, model risk is not broken into its components, but the correction would be for all the types of model risk considered jointly. In this way, the backtesting-based correction methodology for ES, proposed in this paper, provides corrections for all the sources of ES model risk.

Comparing the ex-post realizations of returns with the ex-ante forecasted ES, the accuracy of ES estimates is examined via backtesting. Then we can quantify the model risk of ES estimates through backtesting ES, by computing the correction required to be added to the ES estimate so that the corrected ES passes the backtest. The value of ES corrected via backtesting is written as:

$$ES^B(\hat{\theta}_1, \alpha, C_i) = ES(\hat{\theta}_1, \alpha) + C_i^* \quad (17)$$

The minimum correction is given by:

$$C_i^* = \min\{C_i | ES_t(\hat{\theta}_1, \alpha) + C_i \text{ passes the } i\text{th backtest}, t = 1, \dots, T\}, C_i^* \geq 0 \quad (18)$$

where $\{ES_t(\hat{\theta}, \alpha), t = 1, \dots, T\}$ denotes the forecasted ES during the period from 1 to T . C_i^* is the minimum correction required to pass the i th⁷ ES backtest. $C_i = C_i(\theta_0, \theta_1, \hat{\theta}_1, \alpha)$, a correction of ES estimates, is obtained when the i th backtest of the ES estimates is passed successfully. By learning from past mistakes, we can find the appropriate correction needed to be added to the ES estimates, through which the model risk of ES estimates can be quantified.

Backtesting checks whether ES forecasts satisfy certain properties. Here we consider that a good ES estimate should have a desirable frequency of exceptions, absence of volatility clustering in the tail and an appropriate magnitude of the violations (see **Table 5** in **Appendix B** for a summary of ES backtests).

3.2 Backtesting framework for ES

Exception frequency test

The unconditional coverage test (*UC* test) for VaR measure (Kupiec, 1995) is extended to an unconditional coverage test for ES, proposed by Du and Escanciano (2016). By analogy, they investigate the cumulation of violations and develop an unconditional coverage test statistic for ES. The estimated cumulative violations $\hat{H}_t(\alpha)$ are defined as:

$$\hat{H}_t(\alpha) = \frac{1}{\alpha}(\alpha - \hat{u}_t)\mathbf{1}(\hat{u}_t \leq \alpha) \quad (19)$$

where \hat{u}_t is the estimated probability level corresponding to the daily returns (r_t) in the estimated distribution (\hat{F}_t) with the estimated parameters ($\hat{\theta}_1$), and Ω_{t-1} denotes all the information available until $t - 1$.

$$\hat{u}_t = \hat{F}(r_t, \Omega_{t-1}, \hat{\theta}_1) \quad (20)$$

The null hypothesis of the unconditional coverage test for ES is given by:

$$\mathbb{E} \left[H_t(\alpha, \theta_0) - \frac{\alpha}{2} \right] = 0 \quad (21)$$

⁷In our paper, $i = 1, 2$ and 3 ; C_1 , C_2 and C_3 refer to the correction required to pass the unconditional coverage test for ES and the conditional coverage test for ES introduced by Du and Escanciano (2016), and the Z_2 test proposed by Acerbi and Szekely (2014), respectively.

Hence, the simple t-test statistic and its distribution is:

$$U_{ES} = \frac{\sqrt{n} \left(1/n \sum_{t=1}^n \hat{H}_t(\alpha) - \alpha/2 \right)}{\sqrt{\alpha(1/3 - \alpha/4)}} \sim N(0, 1) \quad (22)$$

Exception frequency and independence test

The conditional coverage test (*CC* test) for VaR is a very popular formal backtesting measure (Christoffersen, 1998). Inspired by this, Du and Escanciano (2016) propose a conditional coverage test for ES and give its test statistic. The null hypothesis of the conditional coverage test for ES is given by:

$$\mathbb{E} \left[H_t(\alpha, \theta_0) - \frac{\alpha}{2} | \Omega_{t-1} \right] = 0 \quad (23)$$

Du and Escanciano propose a general test statistic to test the m th-order dependence of the violations, following a Chi-squared distribution with m degrees of freedom. In the present context, the first order dependence of the violations is considered, so the test statistic follows $\chi^2(1)$. During the evaluation period from $t = 1$ to $t = n$, the basic test statistic $C_{ES}(1)$ is written as:

$$C_{ES}(1) = \frac{n^3}{(n-1)^2} \cdot \frac{\left(\sum_{t=2}^n (\hat{H}_t(\alpha) - \alpha/2)(\hat{H}_{t-1}(\alpha) - \alpha/2) \right)^2}{\left(\sum_{t=1}^n (\hat{H}_t(\alpha) - \alpha/2)(\hat{H}_t(\alpha) - \alpha/2) \right)^2} \sim \chi^2(1) \quad (24)$$

Escanciano and Olmo (2010) point out that the VaR (and correspondingly, ES) backtesting procedure may not be convincing enough due to estimation risk and propose a robust backtest. Based on this, Du and Escanciano (2016) derive robust test statistics for ES. They also agree with Escanciano and Olmo that estimation risk can be ignored and the basic test statistic is robust enough against the alternative hypothesis if the estimation period is much larger than the evaluation period. In this context, the estimation period (1,000) we use is much larger than the evaluation period (250), so the robust test statistic is not considered.

Exception frequency and magnitude test

Acerbi and Szekely (2014) directly backtest expected shortfall by using the test statistic (Z_2 test) as given below:

$$Z_2 = \sum_{t=1}^T \frac{r_t I_t}{T \alpha ES_{\alpha,t}} + 1 \quad (25)$$

I_t , the indicator function, is equal to 1 when the forecasted VaR is violated, otherwise, I_t is equal to 0. ES is jointly backtested in terms of the frequency and the magnitude of VaR exceptions. The ES backtesting method is non-parametric, thus easily implemented. It only needs the magnitude of the VaR violations ($r_t I_t$) and the predicted ES ($ES_{\alpha,t}$). The authors also demonstrate that there is no need to do Monte Carlo simulations to store the predictive distributions due to the stability of the p-values of the Z_2 test statistic across different distribution types. Clift et al. (2016) also support this test statistic (Z_2) introduced by Acerbi and Szekely by comparing the existing backtesting approaches for ES. The Z_2 score at a certain significance level can be determined numerically based on the simulated distribution of Z_2 . If the test statistic is smaller than the Z_2 score⁸, the model is rejected.

3.3 The impact of VaR model risk on the model risk of ES

The correction made to ES estimates can be regarded as a barometer of ES model risk, just as the backtesting-based correction methodology for ES shows. VaR has been an indispensable part of ES calculations and the three ES backtests used in this paper. For instance, the Z_2 test (Acerbi and Szekely, 2014) is commonly considered as a joint backtest of VaR and ES. For this reason, it is of much interest to explore whether the model risk of VaR is transferred to the model risk of ES. On the one hand, ES calculations may be affected by the model risk of VaR, since the inaccuracy of VaR estimates is carried over to the ES estimates as seen in equation (2). On the other hand, the wrong VaR estimates may have impacts on the backtesting results, thus leading to inappropriate corrections of ES estimates. More precisely, the measurement of the ES correction required to pass a backtest is likely to be affected by VaR model risk. To address this, as an additional exercise, we compute the optimal correction of VaR for model risk (estimated at the same significance

⁸The critical value for the Z_2 test is -0.7, which is stable for different distribution types (Acerbi and Szekely, 2014).

level as the corresponding ES) as in Boucher et al. (2014)⁹, and use the corrected VaR for ES calculation, thus giving the values of ES corrected for VaR model risk. Consequently, based on the backtesting-based correction framework, the optimal correction made to the ES corrected for VaR model risk is gauged as a measurement of ES model risk alone.

4 Empirical Analysis

Our purpose is to evaluate the backtesting-based correction methodology for ES by using the DJIA index from 01/01/1900 to 05/03/2017 (29,486 daily returns in total). We employ several well known parametric and nonparametric models for comparison, such as the Gaussian Normal distribution, the Student's t distribution, GARCH(1,1) with normal/student's t innovations, EWMA, Cornish Fisher expansion as well as the historical method¹⁰. First, one-step ahead ES forecasts are obtained in a rolling forecasting scheme with a four-year window. Based on equation (18), we compare the minimum corrections required to pass the ES backtests, where backtesting is done over a year.

Figure 4 shows the relative corrections made to the daily ES, estimated at different significance levels, of four models: EWMA, GARCH(1,1)-N, Gaussian Normal, and Student's t, when considering the frequency of the exceptions (passing the *UC* test). ES forecasts are computed with a four-year moving window and backtested using the entire sample. The level of relative corrections is decreasing when the alpha is increasing, implying that the ES at a smaller significance level may need a larger correction to allow for model risk. Not surprisingly, the dynamic approaches, GARCH(1,1)-N and EWMA, require smaller corrections than the two static models in general, though the Student's t distribution performs better at capturing the fat tails than the EWMA model, for example, at 1% and 1.5% significance levels.

Figure 5 presents the optimal corrections made to the daily ES estimates based on various forecasting models with regard to passing the unconditional coverage test for ES (*UC* test), the conditional test for ES (*CC* test) and the magnitude test (Z_2

⁹To find the optimal correction of VaR for VaR model risk, three VaR backtests are considered. The VaR backtests are Kupiec's unconditional coverage test (Kupiec, 1995), Christoffersen's conditional coverage test (Christoffersen, 1998) and Berkowitz's magnitude test (Berkowitz, 2001).

¹⁰The *UC* and *CC* tests for all the distribution-based ES are examined in the setting proposed by Du and Escanciano (2016), whilst the Cornish Fisher expansion and the historical method are entertained in the same setting but in a more general way. ES for the asymmetric and fat-tailed distributions (Broda and Paoletta, 2009) can be also examined by the three backtests.

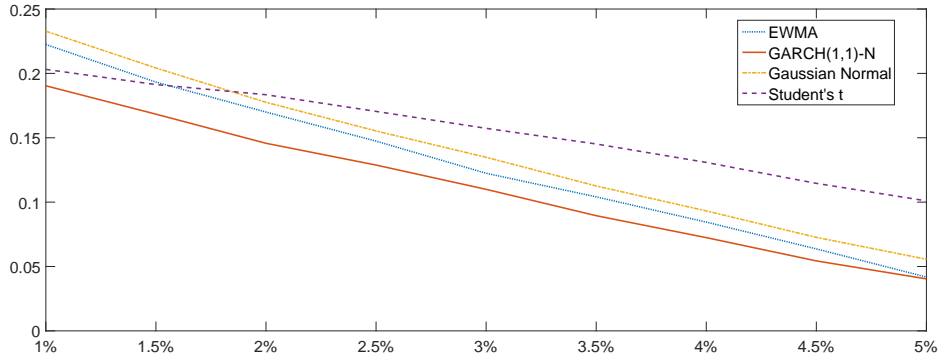
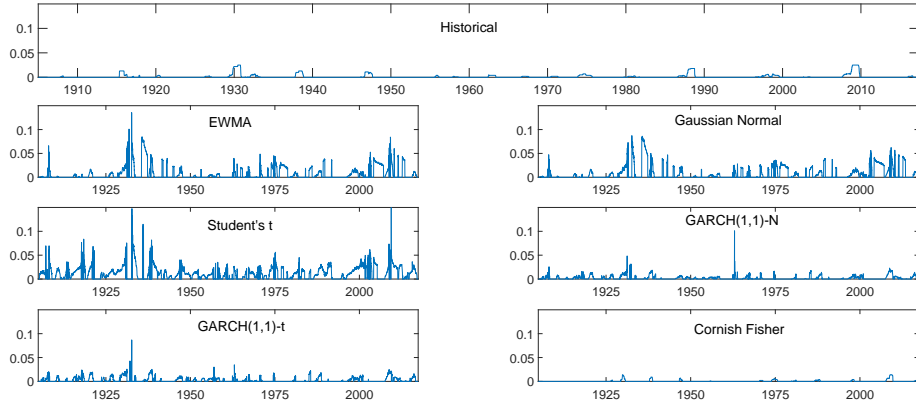


Figure 4: Relative corrections based on the UC test made to the daily ES associated with EWMA, GARCH(1,1)-N, Gaussian Normal, and Student's t along with a range of alpha levels, which is computed as the ratio of the absolute correction over the average daily ES.

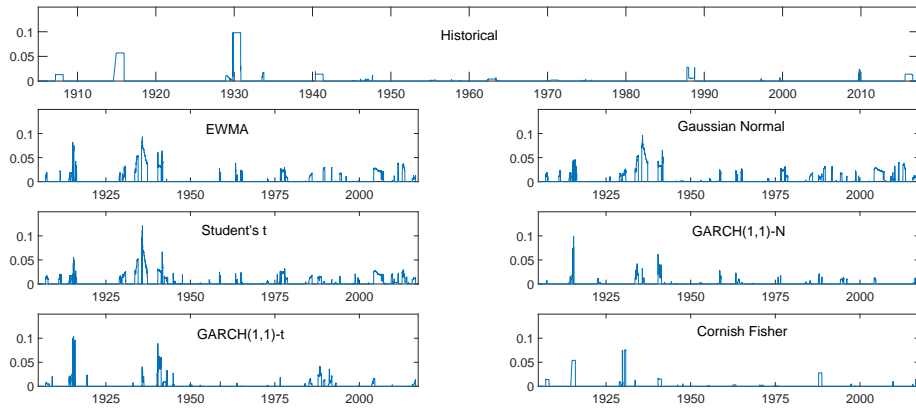
test), respectively, where ES is estimated at a 2.5% significance level using a four-year moving window¹¹ and the evaluation period for backtesting procedures is one year. The optimal corrections made to the ES estimates, found by minimizing equation (18), can be seen as a measurement of model risk. The smaller the correction, the more accurate the ES estimates, therefore the less the model risk of the ES forecasting model. This figure shows that a series of dynamic adjustments are needed for the daily ES ($\alpha = 2.5\%$) based on all different models, especially during the crisis periods. This is in line with our expectation of model inadequacy in the crisis periods. Among the models considered, the EWMA, Gaussian Normal and Student's t models require larger corrections than the others. That is, these three models have higher model risk than the other four models. Also, the Cornish Fisher expansion requires the smallest adjustments in order to pass the UC and CC tests, whilst GARCH(1,1)- t requires the smallest corrections when considering the Z_2 test.

Figure 6 presents the time taken to arrive to the peak of the optimal corrections, for the UC , CC and Z_2 tests, accordingly. It can be seen that more than a decade is needed to find the highest correction required to cover for model risk (also see **Appendix C, Table 6** for the dates when the highest corrections are required). When considering the UC test and the CC test, the highest values of the optimal corrections made to the daily ES based on various models are achieved before the 21st century (except that the highest value of the optimal corrections made to the

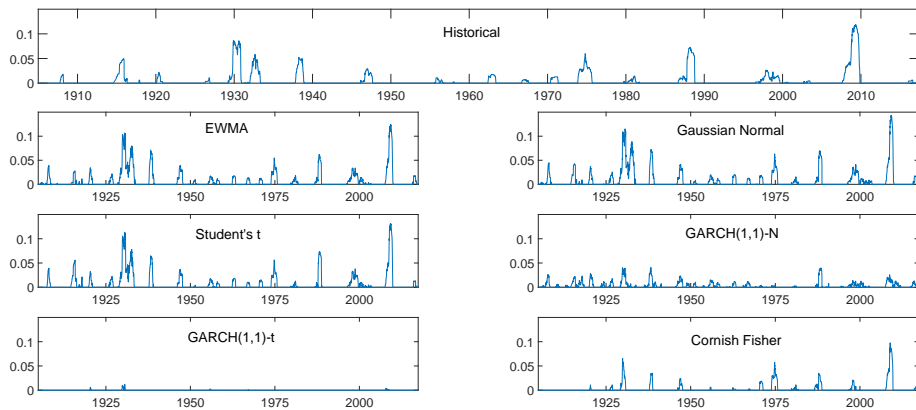
¹¹The results computed by using a five-year moving window and a three-year moving window are very similar to those required here. (available from the authors on request.)



(a) UC test for ES



(b) CC test for ES



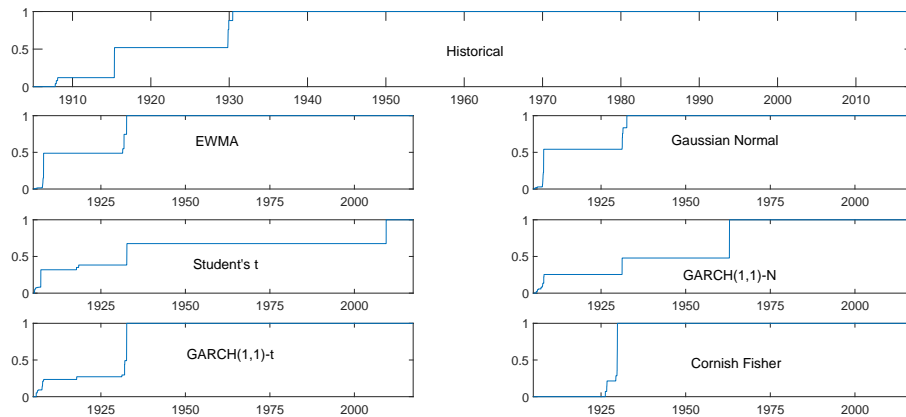
(c) Z_2 test

Figure 5: Optimal corrections made to the daily ES estimates ($\alpha = 2.5\%$) associated with various models for the DJIA index from 01/01/1900 to 23/05/2017, required to pass the UC , CC , and Z_2 tests, respectively. The parameters are re-estimated using a four-year moving window (1,000 daily returns) and the evaluation window for backtesting is one year.

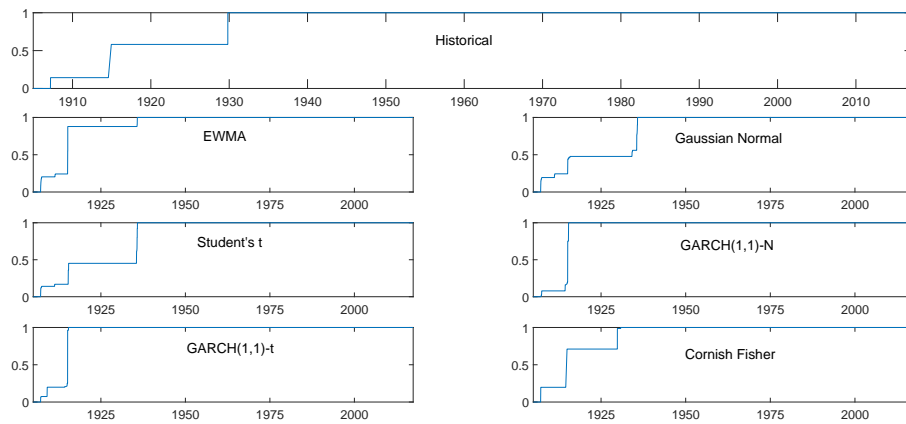
Student's t ES happened around 2008, required to pass the *UC* test), indicating that based on past mistakes we could have avoided the ES failures found by these two tests, for instance, in the 2008 credit crisis. Nevertheless, when considering the Z_2 test, all the models, except for the GARCH models, find the peak values of the optimal corrections around 2008. Therefore, the GARCH models are more favorable than the others in avoiding model risk. This is also supported by the results shown in **Appendix C, Figure 8**, which shows the left tails of the cumulative distribution of the negative of the optimal corrections.

Panel A and **Panel B** of **Table 2** compare the severity of the model risk of different ES estimates with respect to the three backtests discussed earlier. **Panel A** gives the maximum and mean values of the absolute optimal corrections to the daily ES ($\alpha = 2.5\%$) associated with several models, depending on the three ES backtests. The largest corrections are needed for the Gaussian Normal and Student's t models, which do not account for the volatility clustering, whilst the GARCH models perform well in capturing extreme losses. With the requirement of passing the three backtests jointly, the Cornish Fisher expansion performs best and requires a correction of 0.0975 made to the daily ES against model risk. However, the absolute model risk shown in **Panel A** may give an ambiguous understanding of the severity of ES model risk based on different forecasting models, since the values of ES estimates vary for various forecasting models. **Panel B** shows the maximum and mean values of the relative corrections made to the daily ES ($\alpha = 2.5\%$), where the relative corrections are expressed as the optimal corrections over the average daily ES. When looking at the three backtests jointly, the EWMA, Gaussian Normal and Student's t models face the highest ES model risk with the mean values of the relative corrections at 0.3070, 0.3577, and 0.3964, respectively, thereby needing the largest buffer; whilst GARCH(1,1)-t performs best, having the mean value of the relative optimal correction of 0.0868.

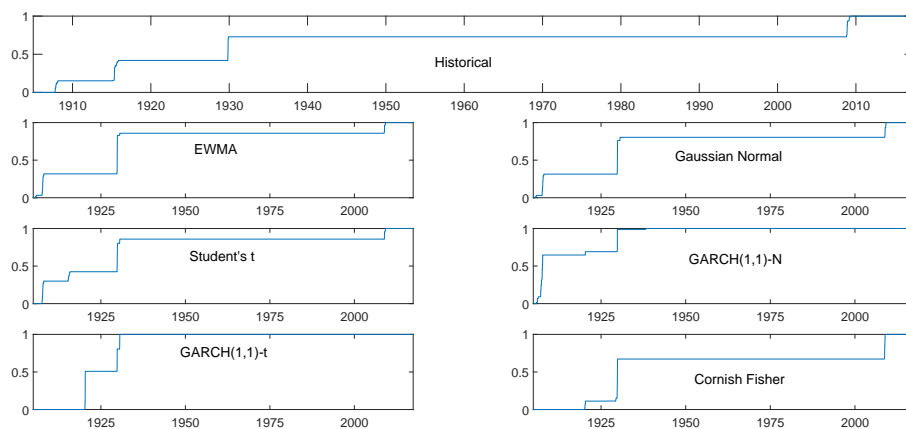
By using the backtesting-based correction methodology for VaR estimates, we compute the relative corrections made to one-step ahead VaR forecasts at a significance level of 1% by passing three VaR backtests as in Boucher et al. (2014), reported in **Panel C** of **Table 2**. The results show that the Cornish Fisher expansion and GARCH(1,1)-t outperform any other model, requiring the smallest corrections. Comparing **Panel B** and **Panel C** of **Table 2**, it can be seen that the peak values of the relative correction required to pass the *UC* and *CC* tests for VaR estimates are generally (with a few exceptions) smaller than the corresponding values for ES



(a) UC test for ES



(b) CC test for ES



(c) Z_2 test

Figure 6: Relative optimal adjustments required by passing the UC , CC , Z_2 tests, which is expressed as the ratio of the corrections over the maximum of the optimal corrections over the entire period.

estimates, whilst the ES estimates require much smaller corrections than the VaR estimates when considering the Z_2 test. That is, ES forecasts are more able to measure the size of the extreme losses. When the three backtests are considered jointly, it can be concluded that ES is less affected by model risk than VaR.

It is interesting to compare our results with those of Danielsson and Zhou (2015). In their Table 1, they show that VaR estimation has a higher bias than ES estimation, but a smaller standard error. However, this is based on a simulation study that focuses on estimation risk. The results presented in the empirical part of their paper somewhat contradict their theoretical expectation of VaR being superior to ES, and it can be argued that this is caused by the presence of specification error. So when only estimation error is considered, VaR is superior to ES, but when both estimation error and specification error are considered jointly, our results show that ES outperforms VaR, being less affected by model risk.

Supplementary to the backtesting-based correction methodology for ES, we examine the impact of VaR model risk on the model risk of ES in **Panel D, Table 2**. It can be seen that for the six models¹², the optimal corrections required to pass the three ES backtests jointly, made to the daily ES after correcting for VaR model risk, are smaller than the corrections made to the daily ES when VaR is not corrected for model risk. We can conclude that ES is less affected by model risk, when VaR model risk is removed in calculating and backtesting ES. Roughly speaking, the corrections for model risk to the ES estimates reduce by about 50% if the VaR estimates are corrected for model risk first. Moreover, the standard deviations of the corrections made to the daily ES before and after VaR model risk is removed are presented in **Appendix C, Table 7**, showing that the smaller corrections consistently come with smaller standard deviations. Ultimately, this supports the previous result that GARCH models are less affected by model risk, thus are preferred to make risk forecasts.

Additionally, we apply the methodology presented to different asset classes (namely equity, bond and commodity from 07/07/1986 to 07/07/2017), the FX (USD/GBP) and Microsoft stock from 01/01/1987 to 04/10/2017. We compare the model risk of ES, and give the dollar values of model risk of two specific assets in the case study. **Table 3** reports the absolute and relative corrections required to pass the three backtests, made to the daily GARCH(1,1)-t ES estimates ($\alpha = 2.5\%$) for the

¹²The model risk of ES forecasting models examined in this paper, except for the Cornish Fisher expansion, is considered to be affected by VaR model risk.

Table 2: Maximum and mean of the absolute and relative optimal corrections made to the daily ES ($\alpha = 2.5\%$), and maximum and mean of the relative optimal corrections to the daily VaR ($\alpha = 1\%$), based on various models and different backtests.

Methods	Mean ES (VaR)	Max C_1	Max C_2	Max C_3	Mean C_1	Mean C_2	Mean C_3
<i>Panel A: Maximum and mean of the absolute optimal corrections to the daily ES ($\alpha = 2.5\%$)</i>							
Historical	0.0306	0.0250	0.0980	0.1186	0.0013	0.0020	0.0053
EWMA ($\lambda=0.94$)	0.0244	0.1355	0.0930	0.1241	0.0069	0.0037	0.0074
Gaussian Normal	0.0245	0.0873	0.0964	0.1433	0.0072	0.0042	0.0084
Student's t	0.0303	0.2184	0.1212	0.1315	0.0113	0.0038	0.0073
GARCH(1,1)-N	0.0225	0.1011	0.0990	0.0408	0.0020	0.0008	0.0033
GARCH(1,1)-t	0.0314	0.0869	0.1041	0.0118	0.0029	0.0015	0.0001
Cornish Fisher	0.0499	0.0140	0.0760	0.0975	0.0005	0.0014	0.0029
<i>Panel B: Maximum and mean of the relative optimal corrections to the daily ES ($\alpha = 2.5\%$)</i>							
Historical	0.0306	0.9846	3.1899	4.3676	0.0568	0.0608	0.1818
EWMA ($\lambda=0.94$)	0.0244	3.1882	3.9932	5.3753	0.2599	0.1162	0.3070
Gaussian Normal	0.0245	2.6898	2.1429	6.7203	0.2739	0.1344	0.3577
Student's t	0.0303	4.7976	2.4108	4.8079	0.3964	0.0979	0.2550
GARCH(1,1)-N	0.0225	5.6038	3.9722	1.3372	0.0840	0.0341	0.1341
GARCH(1,1)-t	0.0314	1.5495	3.1743	0.2340	0.0868	0.0405	0.0016
Cornish Fisher	0.0499	0.5222	2.4011	3.3902	0.0175	0.0222	0.0977
<i>Panel C: Maximum and mean of the relative optimal corrections to the daily VaR ($\alpha = 1\%$)</i>							
Historical	0.0295	0.7820	2.8091	11.1037	0.0286	0.0766	0.5908
EWMA ($\lambda=0.94$)	0.0243	1.0183	2.9779	11.3926	0.0634	0.1075	0.8284
Gaussian Normal	0.0243	1.3942	4.2346	12.2224	0.0730	0.1433	0.8646
Student's t	0.0280	0.8911	3.6617	10.5877	0.0422	0.1004	0.6961
GARCH(1,1)-N	0.0224	0.5051	2.9813	11.4471	0.0229	0.0652	0.8325
GARCH(1,1)-t	0.0302	0.0714	1.7387	7.3335	0.0001	0.0149	0.4016
Cornish Fisher	0.0498	0.3660	1.8009	10.2159	0.0077	0.0241	0.3087
<i>Panel D: Maximum and mean of the absolute corrections to the daily ES, corrected for VaR model risk</i>							
Historical	0.0323	0.0140	0.0870	0.0605	0.0008	0.0018	0.0024
EWMA ($\lambda=0.94$)	0.0257	0.0853	0.0930	0.0672	0.0055	0.0036	0.0038
Gaussian Normal	0.0260	0.0861	0.0964	0.0726	0.0055	0.0041	0.0040
Student's t	0.0319	0.1368	0.1212	0.0642	0.0051	0.0035	0.0031
GARCH(1,1)-N	0.0230	0.0342	0.0706	0.0364	0.0015	0.0008	0.0023
GARCH(1,1)-t	0.0314	0.0644	0.1041	0.0094	0.0014	0.0015	0.0001

Note: The outcomes are computed based on the DJIA index from the 1st January 1900 to the 23rd May 2017, downloaded from DataStream. Based on various forecasting models, the risk estimates (ES and VaR) are forecasted with a four-year moving window (1,000 daily returns) and the mean ES and VaR are calculated over the entire sample. **Panel A** gives the absolute values of the maximum and mean of the historical series of minimum corrections (C_1 , C_2 and C_3) made to ES estimates, accordingly, for the unconditional coverage test (UC test), the conditional coverage test (CC test) and the magnitude test (Z_2 test), whilst **Panel B** gives the relative values of the maximum and mean of the historical series of minimum corrections for ES estimates, obtained by passing the aforementioned backtests, where the relative correction is the ratio of the optimal correction over the average daily ES; backtesting is done over 250 days. **Panel C** gives the relative values of the maximum and mean of the historical series of minimum corrections for VaR estimates, obtained by passing three VaR backtests. **Panel D** gives the absolute values of the maximum and mean of the historical series of minimum corrections for ES estimates, after correcting for VaR model risk.

Table 3: Maximum and mean of the absolute and relative corrections made to the daily GARCH(1,1)-t ES ($\alpha = 2.5\%$) for different asset classes based on different backtests.

Asset classes	Statistics of asset returns			Backtesting-based corrections						
	Std.	Skewness	Kurtosis	Mean ES	Max C_1	Max C_2	Max C_3	Mean C_1	Mean C_2	Mean C_3
<i>Panel A: Maximum and mean of the absolute corrections to the daily GARCH (1,1)-t ES ($\alpha = 2.5\%$)</i>										
equity	0.012	-0.363	8.966	0.0308	0.0246	0.0160	0.0071	0.0034	0.0006	0.0001
bond	0.003	0.029	4.477	0.0082	0.0055	0.0353	0.0000	0.0004	0.0005	0.0000
commodity	0.004	-0.556	7.045	0.0133	0.0105	0.0144	0.0068	0.0009	0.0004	0.0001
<i>Panel B: Maximum and mean of the relative corrections to the daily GARCH (1,1)-t ES ($\alpha = 2.5\%$)</i>										
equity	0.012	-0.363	8.966	0.0308	0.8282	0.5954	0.2835	0.0986	0.0219	0.0028
bond	0.003	0.029	4.477	0.0082	0.6275	3.0534	0.0000	0.0550	0.0548	0.0000
commodity	0.004	-0.556	7.045	0.0133	0.6939	1.5953	0.2759	0.0562	0.0344	0.0036

Note: The results are based on daily returns from 07/07/1986 to 07/07/2017, downloaded from DataStream. For the equity, we use a composite index with 95% “MSCI Europe Index” and 5% “MSCI World Index”; for the bond, we use the “Bank of America Merrill Lynch US Treasury & Agency Index”; for the commodity, we use the “CRB Spot Index”. The average daily ES of various asset classes is computed based on the GARCH(1,1)-t model in a four-year rolling forecasting scheme. **Panel A** gives the absolute values of the maximum and mean of the historical series of minimum corrections (C_1 , C_2 and C_3), accordingly, for the unconditional coverage test (UC test), the conditional coverage test (CC test) and the magnitude test (Z_2 test), whilst **Panel B** gives the relative values of the maximum and mean of the historical series of minimum corrections obtained by passing the three backtests; backtesting is done over 250 days. The relative correction is the ratio of the optimal correction over the average daily ES.

different asset classes¹³. In this case, the higher the corrections, the more unreliable the ES forecasts, thus the worse the fit of the forecasting model to the data. We find that equity has the highest model risk with the highest mean value of the relative optimal correction of 0.0986 required to pass the three tests jointly, when compared with bond and commodity ES, provided that a GARCH(1,1)-t model is used. This is consistent with the statistical properties of the dataset considered, where equity returns are the most volatile with a fat tail. To pass the magnitude test (Z_2 test), the corrections made to the daily ES associated with GARCH(1,1)-t are small, as expected.

To have a further look at the model risk of ES estimates of specific assets, we conduct a case study on the USD/GBP foreign currency as well as the Microsoft stock listed in the Nasdaq Stock Market. We consider that ES is estimated at a significance level of 2.5%, and we have a position of 1 million dollars for each asset. **Table 4** shows the exposure to model risk of the GARCH(1,1)-t ES when investing

¹³See the data source in the note to **Table 3**.

Table 4: Dollar exposure to model risk of GARCH(1,1)-t ES ($\alpha = 2.5\%$) of the USD/GBP exchange rate and Microsoft equity, based on various ES backtests.

Asset	Mean ES	Max C_1	Max C_2	Max C_3	Mean C_1	Mean C_2	Mean C_3
FX USD/GBP	17,654	13,997	13,288	6,924	1,001	700	100
Microsoft stock	64,069	61,837	49,485	17,044	1,701	3,305	100

Note: The data of USD/GBP spot rates and Microsoft stock prices from 01/01/1987 to 04/10/2017 are downloaded from DataStream and Bloomberg, respectively. All the outcomes are in monetary terms (\$) computed by using a four-year moving window and a one-year backtesting period, based on the GARCH(1,1)-t model. The table presents the dollar values of the maximum and mean of the optimal corrections (C_1 , C_2 and C_3) required to pass the UC , CC and Z_2 tests accordingly, when considering a position of 1 million dollars in the asset specified in the first column.

in the USD/GBP exchange rate or by purchasing the Microsoft stock, respectively. Using a significance level of 2.5%, the average ES of the two investments are \$17,654 and \$64,069, respectively. The mean model risk, according to the Z_2 backtest for ES, is only \$100 for both investments. The same backtest gives a maximum correction of \$6,924 for the FX investment and \$17,044 for the stock investment. These examples show why it is necessary for banks to introduce enough protection against model risk when calculating the risk-based capital requirement, as in Basel (2011).

Our empirical analysis shows that, when forecasting ES, GARCH models are preferred, whilst the static models (e.g. the Gaussian Normal and Student's t models) and EWMA should be avoided. However, when considering the model risk of VaR models, Boucher et al. (2014) recommend that EWMA VaR is preferred, whilst the Cornish Fisher expansion VaR should be avoided. Overall, we conclude that GARCH models are preferred for VaR and ES forecasting. Also, ES is the preferred measure of risk since it is less affected by model risk than VaR when both ES and VaR are estimated at the regulatory coverage levels, especially after VaR model risk is removed.

5 Conclusions

The superiority¹⁴ of ES allows it to replace VaR as the standard risk measure as documented in Basel (2013). Usually, only the risk estimates are considered, ignoring the risk of risk models, but this practice can lead to devastating losses. Whilst there are several papers that examine VaR model risk, the model risk of ES forecasting models is a relatively untouched topic. The general methodology in quantifying the model risk of imperfect VaR forecasts based on their backtesting performance, proposed by Boucher et al. (2014), is extended in this paper to study the model risk of ES estimates. Using Monte Carlo simulations, we compute the estimation bias as well as the joint specification and estimation biases of various ES estimates, indicating that the model risk of ES forecasting models is worth a great deal of consideration.

ES forecasts should satisfy certain requirements with respect to the following characteristics: 1) the frequency of exceptions; 2) the absence of autocorrelations in exceptions; 3) the magnitude of exceptions. We use the unconditional coverage test and the conditional coverage test for ES (Du and Escanciano, 2016), as well as the magnitude test proposed by Acerbi and Szekely (2014) to test the adequacy of ES estimates. The optimal corrections, which are considered as a proxy for model risk, are obtained by minimizing the adjustments needed to pass the backtests above. When considering the three tests jointly, the GARCH(1,1)-t model has the smallest model risk.

Comparing the model risk of ES and VaR risk models at the regulatory coverage levels, we find that the ES estimates are less affected by model risk, needing a smaller correction to pass the three ES backtests jointly. If VaR model risk is first removed from ES calculations and ES backtests, then ES model risk reduces further by approximately 50%. Our results are strengthened when the standard deviations of the corrections for model risk are considered: GARCH models not only require the smallest corrections for model risk, but the level of the corrections are the most stable, when compared to the other models considered in our study. Also, we obtain that the ES of equity investments are more vulnerable to the model risk of risk models, indicating that equity investors should be more aware of model inadequacy, when compared to bond, commodity and FX investors.

¹⁴Artzner et al. (1999) and Acerbi and Tasche (2002) suggest ES as a remedy for the VaR measure accused of non-subadditivity; Colletaz et al. (2013) and Danielsson and Zhou (2015) show that ES is more sensitive to the magnitude of losses in the tail than VaR.

Appendix A. Theoretical analysis of estimation and specification errors of VaR¹⁵

Estimation bias and correction of VaR

Based on equation (5) and (6), correcting for the estimation error, the VaR estimate can be written as:

$$VaR^E(\hat{\theta}_0, \alpha) = VaR(\hat{\theta}_0, \alpha) + \mathbb{E}(bias(\theta_0, \hat{\theta}_0, \alpha)) \quad (26)$$

This tells us that the mean bias of the forecasted VaR from the theoretical VaR is caused by estimation error.

Specification and estimation biases and correction of VaR

Based on equation (10) and (11), correcting for these biases (specification and estimation), the VaR estimate can be written as:

$$VaR^{SE}(\hat{\theta}_1, \alpha) = VaR(\hat{\theta}_1, \alpha) + \mathbb{E}(bias(\theta_0, \theta_1, \hat{\theta}_1, \alpha)) \quad (27)$$

The mean of the estimation bias and specification bias for VaR can be considered as a measurement of economic value of the model risk of VaR.

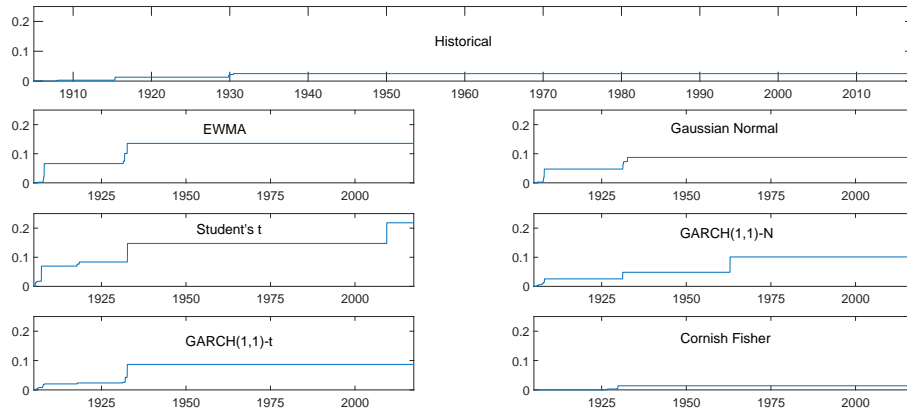
¹⁵The analysis is based on Boucher et al. (2014).

Appendix B. Backtesting measures of VaR and ES

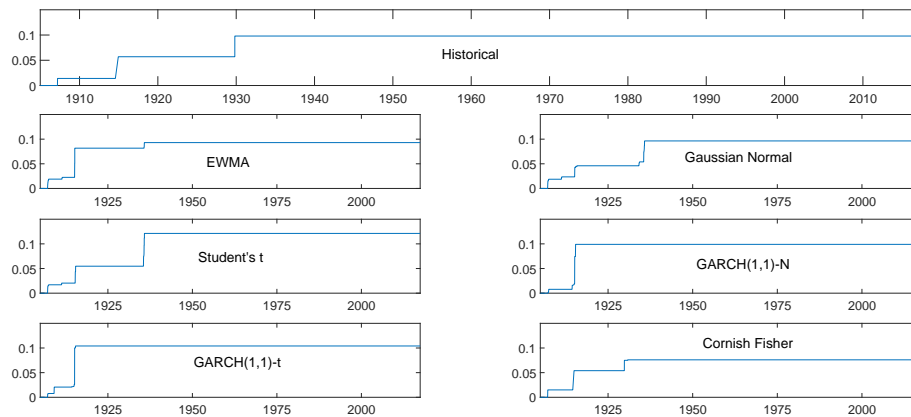
Table 5: Selected backtesting methodologies for VaR and ES

VaR backtests	ES backtests
<i>Exception Frequency Tests:</i>	<i>Exception Frequency Tests:</i>
1)UC test - Kupiec (1995)	1)UC test - Du and Escanciano (2016)
2)data-driven- Escanciano and Pei (2012)	
	2)risk map- Colletaz et al. (2013)
<i>Exception Independence Tests:</i>	<i>Exception Independence Tests:</i>
1)independence test-Christoffersen (1998)	
2)density test- Berkowitz (2001)	
<i>Exception Frequency and Independence Tests:</i>	<i>Exception Frequency and Independence Tests:</i>
1)CC test- Christoffersen (1998)	1)CC test- Du and Escanciano (2016); Costanzino and Curran (2015a,b)
2)dynamic quantile-Engle and Manganelli (2004)	
3)multilevel test- Campbell (2006)	
4)multilevel test-Leccadito et al. (2014)	
5)multinomial test-Kratz et al. (2016)	2)multinomial test-Kratz et al. (2016); Emmer et al. (2015); Clift et al. (2016)
6)two-stage test- Angelidis and Degiannakis (2006)	
<i>Exception Duration Tests:</i>	<i>Exception Duration Tests:</i>
1)duration-based test- Berkowitz et al. (2011)	
<i>Exception Magnitude Tests:</i>	<i>Exception Magnitude Tests:</i>
1)tail losses- Wong (2010)	1) tail losses- Wong (2008); Christoffersen (2009); McNeil and Frey (2000)
2)magnitude test-Berkowitz (2001)	
<i>Exception Frequency and Magnitude Tests:</i>	<i>Exception Frequency and Magnitude Tests:</i>
1)risk map- Colletaz et al. (2013)	
	1) Z_2 test-Acerbi and Szekely (2014)

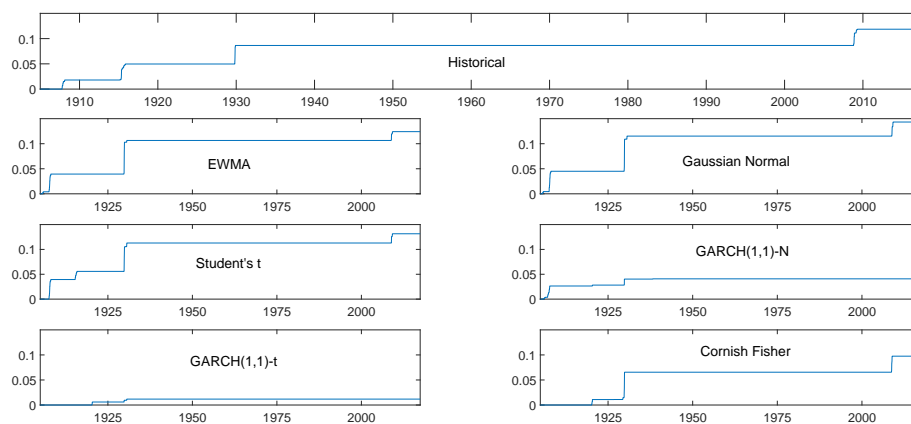
Appendix C. Empirical results



(a) UC test for ES



(b) CC test for ES



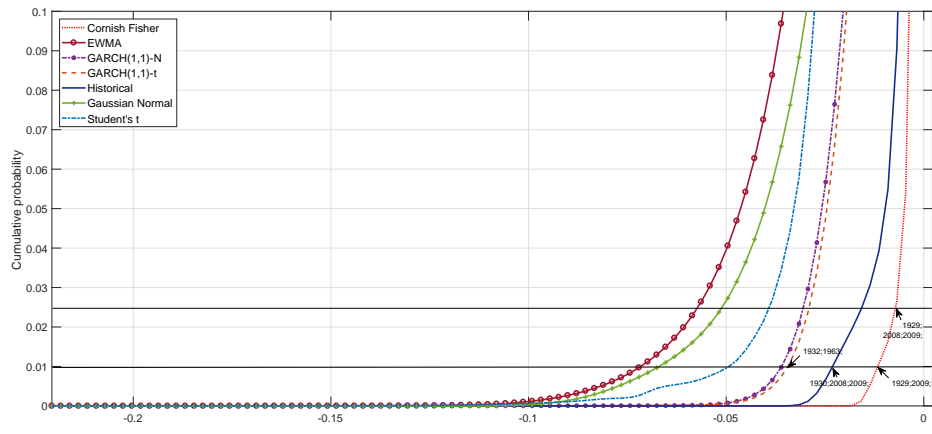
(c) Z_2 test

Figure 7: Historical maximum of the optimal corrections required by passing the UC , CC and Z_2 tests, respectively.

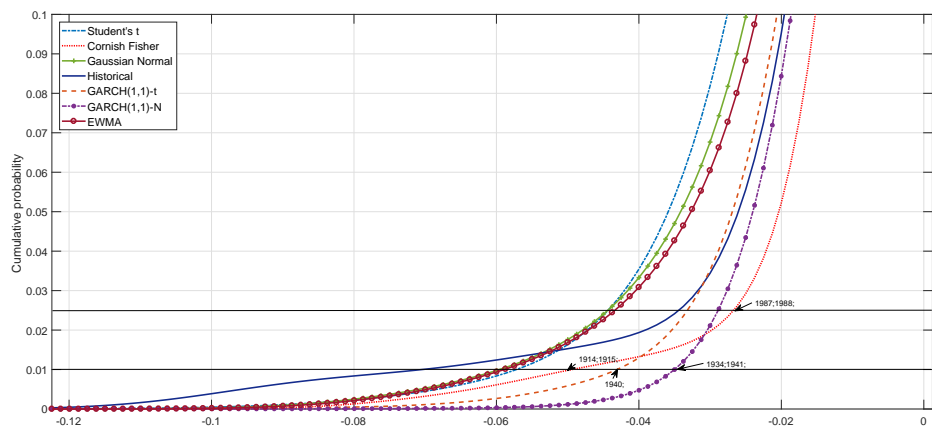
Table 6: The highest values of the absolute minimum corrections made to the daily ES ($\alpha = 2.5\%$) based on several models and different backtests.

Models	<i>UC</i> test		<i>CC</i> test		<i>Z2</i> test		
	Dates	C_1	Dates	C_2	Dates	C_3	
Historical	1	16/06/1930	0.0250	29/10/1929	0.0980	20/04/2009	0.1186
	2	11/09/2009	0.0240	14/12/1914	0.0570	30/03/2009	0.1176
	3	20/11/2008	0.0230	30/10/1930	0.0300	05/03/2009	0.1172
	4	12/12/1929	0.0220	13/12/1915	0.0280	19/05/2009	0.1167
EWMA	1	15/08/1932	0.1355	15/10/1935	0.0930	20/04/2009	0.1241
	2	08/08/1932	0.1196	18/10/1935	0.0898	05/03/2009	0.1238
	3	09/11/1931	0.1010	17/10/1935	0.0897	30/03/2009	0.1229
	4	22/06/1931	0.0744	16/10/1935	0.0893	05/05/2009	0.1225
Gaussian Normal	1	17/08/1932	0.0873	15/10/1935	0.0964	20/04/2009	0.1433
	2	13/09/1935	0.0861	18/10/1935	0.0927	05/03/2009	0.1431
	3	12/09/1935	0.0859	17/10/1935	0.0925	30/03/2009	0.1421
	4	16/09/1935	0.0850	16/10/1935	0.0921	05/05/2009	0.1418
Student's t	1	29/05/2009	0.2184	25/10/1935	0.1212	05/03/2009	0.1315
	2	15/09/1932	0.1475	04/10/1935	0.1118	20/04/2009	0.1308
	3	11/10/1932	0.1324	28/10/1935	0.1041	30/03/2009	0.1300
	4	08/09/1932	0.1206	29/10/1935	0.1005	02/03/2009	0.1299
GARCH(1,1)-N	1	14/12/1962	0.1011	02/06/1915	0.0990	29/03/1938	0.0408
	2	19/12/1962	0.0990	10/06/1915	0.0775	29/10/1929	0.0403
	3	27/03/1931	0.0484	01/03/1915	0.0744	14/04/1988	0.0397
	4	26/03/1931	0.0471	02/03/1915	0.0721	08/08/1930	0.0396
GARCH(1,1)-t	1	24/08/1932	0.0869	08/06/1915	0.1041	08/08/1930	0.0118
	2	25/08/1932	0.0854	25/05/1915	0.1022	28/10/1928	0.0095
	3	26/08/1932	0.0812	03/03/1915	0.1002	12/12/1929	0.0086
	4	02/02/1932	0.0427	09/06/1915	0.0999	21/07/1930	0.0084
Cornish Fisher	1	06/11/1929	0.0140	28/10/1930	0.0760	01/12/2008	0.0975
	2	29/10/1929	0.0130	29/10/1929	0.0750	08/12/2008	0.0951
	3	10/02/1930	0.0120	14/12/1914	0.0540	29/12/2008	0.0933
	4	28/10/1929	0.0110	19/10/1987	0.0280	20/11/2008	0.0915

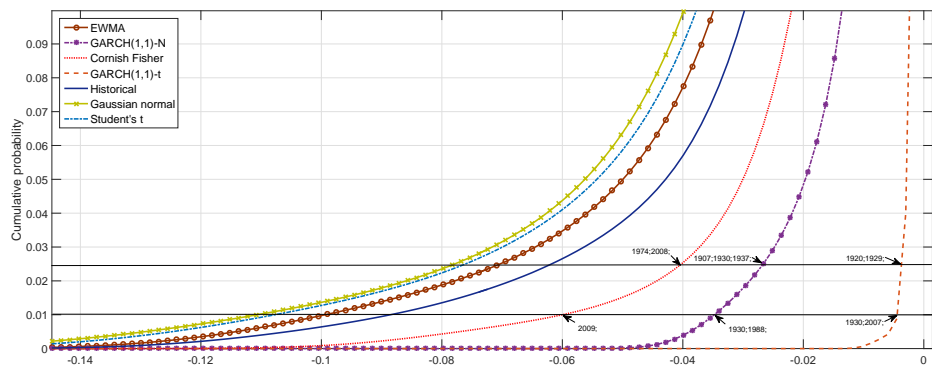
Note: The results are based on the DJIA index daily returns from 1st January 1900 to 23rd May 2017, downloaded from DataStream. Assuming several models, we re-estimate the parameters with a four-year moving window and backtest ES estimates in the evaluation period of 250 days. The variables C_1 , C_2 and C_3 denote the optimal corrections required to pass the unconditional coverage test (*UC* test), the conditional coverage test (*CC* test) and the magnitude test (Z_2 test), respectively.



(a) UC test for ES



(b) CC test for ES



(c) Z_2 test

Figure 8: The left tail of the cumulative distribution (using a Gaussian Kernel smoothing density) of the negative of the optimal corrections obtained by passing the UC , CC , and Z_2 tests, respectively.

Table 7: Standard deviations of the absolute and relative corrections made to the daily ES ($\alpha = 2.5\%$), as well as standard deviations of the absolute corrections made to the daily ES ($\alpha = 2.5\%$) after VaR model risk is first removed, based on several models and different backtests.

Method	Std. dev of C_1	Std. dev of C_2	Std. dev of C_3
<i>Panel A: Standard deviations of the absolute optimal corrections made to the daily ES ($\alpha = 2.5\%$).</i>			
Historical	0.0039	0.0108	0.0157
EWMA($\lambda=0.94$)	0.0133	0.0108	0.0179
Gaussian Normal	0.0135	0.0111	0.0200
Student's t	0.0125	0.0098	0.0186
GARCH(1,1)-N	0.0039	0.0038	0.0067
GARCH(1,1)-t	0.0051	0.0063	0.0006
<i>Panel B: Standard deviations of the relative optimal corrections made to the daily ES ($\alpha = 2.5\%$).</i>			
Historical	0.1215	0.3050	0.5010
EWMA($\lambda=0.94$)	0.4263	0.3034	0.6769
Gaussian Normal	0.4339	0.3095	0.7991
Student's t	0.3823	0.2167	0.5933
GARCH(1,1)-N	0.1530	0.1471	0.2415
GARCH(1,1)-t	0.1430	0.1556	0.0138
<i>Panel C: Standard deviations of the absolute optimal corrections made to the daily ES ($\alpha = 2.5\%$), after VaR model risk is first removed.</i>			
Historical	0.0024	0.0100	0.0077
EWMA($\lambda=0.94$)	0.0121	0.0110	0.0093
Gaussian Normal	0.0122	0.0113	0.0099
Student's t	0.0107	0.0097	0.0089
GARCH(1,1)-N	0.0031	0.0046	0.0052
GARCH(1,1)-t	0.0057	0.0063	0.0006

Note: The results are based on the DJIA index daily returns from 1st January 1900 to 23rd May 2017, downloaded from DataStream. Assuming several models, we re-estimate the parameters with a four-year moving window and backtest ES estimates in the evaluation period of 250 days. The variables C_1 , C_2 and C_3 denote the optimal corrections required to pass the unconditional coverage test (UC test), the conditional coverage test (CC test) and the magnitude test (Z_2 test), respectively.

References

- Acerbi, C., Szekely, B., 2014. Backtesting expected shortfall .
- Acerbi, C., Tasche, D., 2002. On the coherence of expected shortfall. *Journal of Banking & Finance* 26, 1487–1503.
- Alexander, C., Sarabia, J. M., 2010. Endogenizing model risk to quantile estimates. Tech. rep., Henley Business School, Reading University.
- Alexander, C., Sarabia, J. M., 2012. Quantile uncertainty and value-at-risk model risk. *Risk Anal* 32, 1293–1308.
- Angelidis, T., Degiannakis, S. A., 2006. Backtesting var models: An expected shortfall approach .
- Artzner, P., Delbaen, F., Eber, J.-M., Heath, D., 1999. Coherent measures of risk. *Mathematical finance* 9, 203–228.
- Barrieu, P., Ravanelli, C., 2015. Robust capital requirements with model risk. *Economic Notes* 44, 1–28.
- Basel, 2011. Basel iii: A global regulatory framework for more resilient banks and banking systems.
- Basel, 2012. Fundamental review of the trading book. <http://www.bis.org/publ/bcbs219.pdf>.
- Basel, 2013. Basel committee on banking supervision (2013): Fundamental review of the trading book: A revised market risk framework consultative document. basel: Bank for international settlements.
- Berkowitz, J., 2001. Testing density forecasts, with applications to risk management. *Journal of Business & Economic Statistics* 19, 465–474.
- Berkowitz, J., Christoffersen, P., Pelletier, D., 2011. Evaluating value-at-risk models with desk-level data. *Management Science* 57, 2213–2227.
- Berkowitz, J., O'Brien, J., 2002. How accurate are value-at-risk models at commercial banks? *The journal of finance* 57, 1093–1111.

- Boucher, C. M., Daníelsson, J., Kouontchou, P. S., Maillet, B. B., 2014. Risk models-at-risk. *Journal of Banking & Finance* 44, 72–92.
- Broda, S., Paoletta, M., 2009. Calculating expected shortfall for distributions in finance. Tech. rep., Mimeo, Swiss Banking Institute, University of Zurich.
- Campbell, S. D., 2006. A review of backtesting and backtesting procedures. *The Journal of Risk* 9, 1.
- Christoffersen, P., 2009. Backtesting. *Encyclopedia of Quantitative Finance* .
- Christoffersen, P. F., 1998. Evaluating interval forecasts. *International economic review* pp. 841–862.
- Clift, S. S., Costanzino, N., Curran, M., 2016. Empirical performance of backtesting methods for expected shortfall .
- Colletaz, G., Hurlin, C., Pérignon, C., 2013. The risk map: A new tool for validating risk models. *Journal of Banking & Finance* 37, 3843–3854.
- Cont, R., 2006. Model uncertainty and its impact on the pricing of derivative instruments. *Mathematical Finance* 16, 519–547.
- Costanzino, N., Curran, M., 2015a. Backtesting general spectral risk measures with application to expected shortfall. Available at SSRN 2514403 .
- Costanzino, N., Curran, M., 2015b. A simple traffic light approach to backtesting expected shortfall. *Red* 9, 99–97.
- Daníelsson, J., James, K., Valenzuela, M., Zer, I., 2016. Model risk of risk models. *Journal of Financial Stability* forthcoming, forthcoming: *Journal of Financial Stability*.
- Danielsson, J., Zhou, C., 2015. Why risk is so hard to measure .
- Du, Z., Escanciano, J. C., 2016. Backtesting expected shortfall: accounting for tail risk. *Management Science* .
- Embrechts, P., Puccetti, G., Rüschendorf, L., Wang, R., Beleraj, A., 2014. An academic response to basel 3.5. *Risks* 2, 25–48.

- Emmer, S., Kratz, M., Tasche, D., 2015. What is the best risk measure in practice? a comparison of standard measures. *Journal of Risk* 18, 31–60.
- Engle, R. F., Manganelli, S., 2004. Caviar: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics* 22, 367–381.
- Escanciano, J. C., Olmo, J., 2010. Robust backtesting tests for value-at-risk models. *Journal of Financial Econometrics* p. nbq021.
- Escanciano, J. C., Pei, P., 2012. Pitfalls in backtesting historical simulation var models. *Journal of Banking & Finance* 36, 2233–2244.
- Farkas, W., Fringuellotti, F., Tunaru, R., 2016. Regulatory capital requirements: Saving too much for rainy days? EFMA annual meeting.
- Hartz, C., Mittnik, S., Paoella, M., 2006. Accurate value-at-risk forecasting based on the normal-garch model. *Computational Statistics & Data Analysis* 51, 2295–2312.
- Kerkhof, J., Melenberg, B., Schumacher, H., 2010. Model risk and capital reserves. *Journal of Banking & Finance* 34, 267–279.
- Kratz, M., Lok, Y. H., McNeil, A. J., 2016. Multinomial VaR backtests: A simple implicit approach to backtesting expected shortfall. ESSEC Working paper. Document de Recherche ESSEC / Centre de recherche de IESSEC. ISSN : 1291-9616. WP 1617. 2016.
- Kupiec, P. H., 1995. Techniques for verifying the accuracy of risk measurement models. *The J. of Derivatives* 3.
- Leccadito, A., Boffelli, S., Urga, G., 2014. Evaluating the accuracy of value-at-risk forecasts: New multilevel tests. *International Journal of Forecasting* 30, 206–216.
- McNeil, A. J., Frey, R., 2000. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of empirical finance* 7, 271–300.
- So, M. K., Wong, C.-M., 2012. Estimation of multiple period expected shortfall and median shortfall for risk management. *Quantitative Finance* 12, 739–754.

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

Wong, W. K., 2010. Backtesting value-at-risk based on tail losses. *Journal of Empirical Finance* 17, 526–538.