

On the role of the estimation error in prediction of expected shortfall

Carl Lönnbark*

Department of Economics

Umeå School of Business and Economics, Umeå University

Umeå Economics Studies 844

Abstract

In the estimation of risk measures such as Value at Risk and Expected shortfall relatively short estimation windows are typically used rendering the estimation error a possibly non-negligible component. In this paper we build upon previous results for the Value at Risk and discuss how the estimation error comes into play for the Expected Shortfall. We identify two important aspects where it may be of importance. On the one hand there is in the evaluation of predictors of the measure. On the other there is in the interpretation and communication of it. We illustrate magnitudes numerically and emphasize the practical importance of the latter aspect in an empirical application with stock market index data.

Key Words: Backtesting, Delta method, Finance, GARCH, Risk Management.

JEL Classification: G19, C52, C53, C58, G10.

*Correspondence to carl.lonnbark@econ.umu.se. The financial support from the Jan Wallander and Tom Hedelius Foundation is gratefully acknowledged. Samuel Samama is thanked for commenting on a previous version of the paper.

1 Introduction

The recent financial crisis has highlighted the need of properly understanding and measuring financial risks and in particular of evaluating the means of doing so. When it comes to measuring financial risk the Value-at-Risk (VaR) has during the past two decades or so emerged as the standard approach and it is today extensively employed by financial institutions over the world. This popularity is at least partially due to the fact that regulators have adopted the measure as a base for capital adequacy calculations. This was first stipulated in the 1996 Amendment to the first Basel Accord on banking supervision and later further detailed and reinforced in the second Accord (see Basel Committee on Banking Supervision, 2005, 2006). In the aftermath of the financial crises new regulations have been developed to further strengthen capital requirement calculations (see Basel Committee on Banking Supervision, 2012a). Consequently, the measure has been given due attention in the literature (see Jorion, 2007, for an extensive overview).

The VaR gives a potential portfolio loss that will only be exceeded with some (small) probability over a given horizon. As such it is conceptually simple. However, critique has been directed at the VaR measure both from the academia and from the industry. A complaint from the latter is that the VaR is silent about the size of the loss when it exceeds the VaR . Furthermore, the VaR may fail to acknowledge so-called tail risk. That is, two portfolios may have the same risk in terms of VaR , but their outcome in case of VaR exceedence may be substantially different (e.g. Yamai and Yoshida, 2005). In an important paper Artzner, Delbaen, Eber, and Heath (1999) give a formal discussion of what constitutes a good risk measure and establish some properties of coherence that should be satisfied. In particular, a risk measure should acknowledge the principle of diversification. However, it is possible to find perverse cases, where the VaR does not satisfy this property. A measure that fares better in these respects is the Expected Shortfall (ES) that gives the expected loss given that the loss exceeds the VaR . As

the name implies it says something about the size of the loss when disaster strikes, and it also acknowledges tail risk in a better way than VaR . The measure also possesses the desirable property of coherence. In fact, in a recent report the Basel Committee on Banking Supervision suggests a move towards the ES as the risk measure of choice for capital adequacy calculations (see Basel Committee on Banking Supervision, 2012b).

In computing the VaR and the ES a model for the joint movements of the risk factors of the portfolio is typically postulated and the parameters of that model are estimated based on a data-set containing past observations. Thus, uncertainty in the predictors of VaR and ES arises from two primary sources. First of all, the true data generating process is not known, which gives rise to model risk. Secondly, the fact that the parameters of the hypothesized model must be estimated gives rise to estimation risk. Here, the focus is on the estimation risk. This source of error is often referred to as a second order issue and neglected though. Consequently, it is relatively understudied. In fact, Lan, Hu, and Johnson (2007) report that the research on the uncertainty of VaR only amounts to about 2.5 percent of the VaR literature. In practise though, relatively short estimation windows of one or two years are typically used rendering the estimation error a non-negligible component. Indeed, the importance of estimation risk in this context has previously been emphasized by Jorion (1996), Christoffersen and Gonçalves (2005) and others. In fact, Lönnbark (2010) demonstrates that the estimation error in VaR predictors may cause underestimation of portfolio risk in the sense that the probability of exceeding the estimated VaR is higher than the chosen level. Thus, the estimation error affects the interpretation of the VaR . In addition, when it comes to assessing the adequacy of a VaR model the conventional way is to compare a time series of historical VaR predictions to the corresponding portfolio returns. This procedure is commonly referred to as backtesting (e.g. Christoffersen, 2003, Ch. 8). A good VaR model should have a proportion of VaR exceedences (days when the loss exceeds the VaR) close to the chosen probability level. Consequently, as discussed in Escanciano and Olmo (2010) the

estimation error also affects the backtesting procedure and may bias the breach frequency. Of obvious interest is what the picture looks like for the ES measure, which is the focus of this paper.

2 ES and VaR predictors

We assume that portfolio returns are generated in discrete time by

$$y_t = \mu(\boldsymbol{\theta}_{10}, I_{t-1}) + \sigma(\boldsymbol{\theta}_{20}, I_{t-1})\varepsilon_t, \quad (1)$$

where we take ε_t to be a standard normally distributed random variable. The $\mu(\cdot)$ and the $\sigma(\cdot)$ are the conditional mean and standard deviation functions, respectively. The vectors $\boldsymbol{\theta}_{10}$ and $\boldsymbol{\theta}_{20}$ contain true parameters and the set I_t contains the information available at time t . Typically, $\sigma(\cdot)$ is postulated indirectly in terms of the conditional variance (cf. the workhorse GARCH(1,1) specification of Bollerslev (1986) that parameterizes the conditional variance by $\sigma_t^2 = \beta_0 + \beta_1 y_{t-1}^2 + \beta_2 \sigma_{t-1}^2$). For a portfolio with returns generated by (1) the one period ahead conditional VaR , VaR_t^α , satisfies $\Pr_{t-1}(y_t \leq -VaR_t^\alpha) = \alpha$, where the subscript $t-1$ indicates that the probability is conditional on I_{t-1} , and is in this case given explicitly by

$$VaR_t^\alpha = -\mu(\boldsymbol{\theta}_{10}, I_{t-1}) - \sigma(\boldsymbol{\theta}_{20}, I_{t-1})\Phi_\alpha^{-1}, \quad (2)$$

where Φ_α^{-1} is the inverse of the cdf of the standard normal distribution evaluated at α . The associated ES is given by

$$\begin{aligned} ES_t^\alpha &= -E_{t-1}(y_t \mid y_t \leq -VaR_t^\alpha) \\ &= -\mu(\boldsymbol{\theta}_{10}, I_{t-1}) - \sigma(\boldsymbol{\theta}_{20}, I_{t-1})\phi(\Phi_\alpha^{-1})/\alpha \end{aligned} \quad (3)$$

where $\phi(\cdot)$ is the pdf of the standard normal distribution and where the subscript $t-1$ on the expectation operator indicates that it is conditional on I_{t-1} . The VaR and the ES are conventionally reported as positive numbers. Hence, the minus signs in the definitions above.

When it comes to the estimation of the parameter vector, $\boldsymbol{\theta}_0 = (\boldsymbol{\theta}'_{10}, \boldsymbol{\theta}'_{20})'$, the maximum likelihood estimator is commonly employed. It takes as the estimator the parameter vector, $\boldsymbol{\theta} = (\boldsymbol{\theta}'_1, \boldsymbol{\theta}'_2)'$, that maximizes the (conditional) likelihood function, $L \propto -(1/2) \sum (\ln \sigma_t^2 + (y_t - \mu_t)^2 / \sigma_t^2)$, where $\mu_t = \mu(\boldsymbol{\theta}_1, I_{t-1})$ and $\sigma_t = \sigma(\boldsymbol{\theta}_2, I_{t-1})$. Given some regularity conditions the estimator vector, $\hat{\boldsymbol{\theta}}$, is asymptotically normally distributed with the true parameter vector, $\boldsymbol{\theta}_0$, as its mean and covariance matrix $\boldsymbol{\Sigma} = -[E(\partial^2 \ln L(\boldsymbol{\theta}_0) / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}')]^{-1}$. Predictors of VaR_t^α and ES_t^α are simply obtained by plugging in the estimator vector, $\hat{\boldsymbol{\theta}}$, in the expressions (2) and (3), respectively, to obtain

$$\widehat{VaR}_t^\alpha = -\mu(\hat{\boldsymbol{\theta}}_1, I_{t-1}) - \sigma(\hat{\boldsymbol{\theta}}_2, I_{t-1})\Phi_\alpha^{-1}, \quad (4)$$

and

$$\widehat{ES}_t^\alpha = -\mu(\hat{\boldsymbol{\theta}}_1, I_{t-1}) - \sigma(\hat{\boldsymbol{\theta}}_2, I_{t-1})\phi(\Phi_\alpha^{-1})/\alpha. \quad (5)$$

3 The role of the estimation error

When it comes to quantifying the uncertainty due to the estimation error in the VaR and the ES predictors we may rely on the asymptotic normality of the parameter estimator (cf. Hansen, 2006, and others). Heuristically, asymptotic normality of \widehat{VaR}_t^α and \widehat{ES}_t^α follows from the asymptotic normality of $\hat{\boldsymbol{\theta}}$

$$\widehat{VaR}_t^\alpha \sim N(VaR_t^\alpha, \delta_t^2), \quad (6)$$

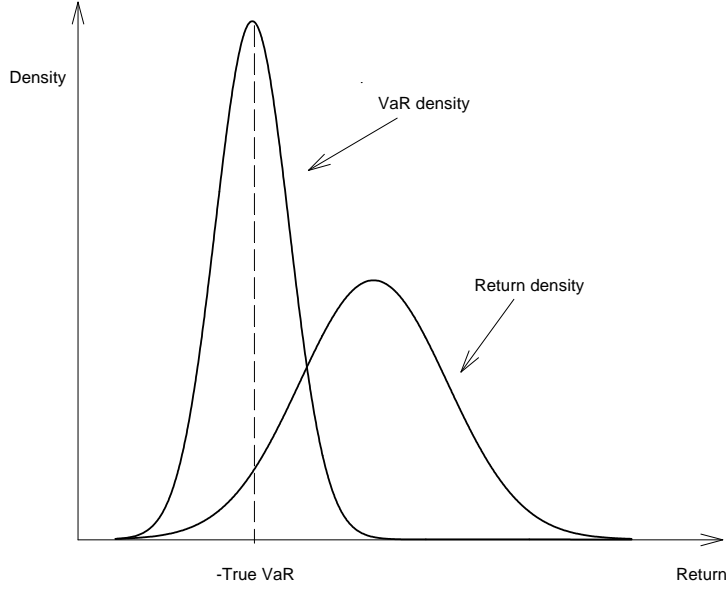


Figure 1: VaR and return densities. VaR density and return density refers to the conditional densities of the VaR predictor and the return, respectively.

$$\widehat{ES}_t^\alpha \sim N(ES_t^\alpha, v_t^2), \quad (7)$$

where the variances, δ_t^2 and v_t^2 , may be obtained by employing the delta method. In the sequel we maintain the assumption that \widehat{VaR}_t^α and \widehat{ES}_t^α are normally distributed. A key insight is that, in practice, we use a random predictor of the true VaR and when it comes to interpreting and communicating the measure the relevant probability is $\Pr_{t-1}\{y_t \leq -\widehat{VaR}_t^\alpha\}$. Clearly, this probability does not necessarily equal α and may in fact equal some $\alpha^* > \alpha$ implying an underestimation of portfolio risk. Indeed, statements such as "the probability that the portfolio loss is less than the VaR is 100α %" may be quite misleading. In Figure 1 we depict a situation with an unbiased VaR predictor.

For a VaR "draw" to the left of (minus) the true VaR the probability of exceedence is smaller than α . For a draw to the right the opposite is true. As the return density

is positively sloped through the VaR density the latter will dominate. We note that if the return density were flat through the VaR density there would be no effect on the exceedence probability, i.e. $\alpha^* = \alpha$. Extrapolating on this reasoning we may conjecture that the difference between α^* and α is smaller for fat tailed return distributions.

Essentially, in the backtesting of a VaR predictor we compare draws from the VaR distribution to draws from the return distribution. Thus, the discussion above have a bearing on this procedure and for VaR the role of estimation error is essentially the same for evaluation and interpretation. Here, the interest is in the role of the estimation error for ES predictors.

Now, the ES gives the expected loss given VaR exceedence and for the purpose of interpreting and communicating ES figures it is of interest to compare the actual expected loss, i.e. $-E_{t-1}(y_t|y_t \leq -\widehat{VaR}_t^\alpha)$, given exceedence of the (random) VaR to the true expected shortfall, ES_t^α . To this end it is straightforward to show that $\Pr_{t-1}(y_t \leq -\widehat{VaR}_t^\alpha) = \alpha^* = \Phi(\Phi_\alpha^{-1}\sigma_t/\sqrt{\sigma_t^2 + \delta_t^2})$ and in the Mathematical Appendix we show that $E_{t-1}[y_t \mathbf{1}(y_t \leq -\widehat{VaR}_t^\alpha)] = \alpha^*\mu_t + \sigma_t/(2\pi s)\sqrt{\pi/a} \exp[b^2/(4a) - c]$, where $a = (1 + 1/s^2)/2$, $b = -m/s^2$ and $c = m^2/(2s^s)$, and where $m = \Phi_\alpha^{-1}$ and $s = \delta_t/\sigma_t$. We have

$$\begin{aligned} E_{t-1}(y_t|y_t \leq -\widehat{VaR}_t^\alpha) &= \frac{E_{t-1}(y_t \mathbf{1}(y_t \leq -\widehat{VaR}_t^\alpha))}{\Pr(y_t \leq -\widehat{VaR}_t^\alpha)} \\ &= \mu_t + \frac{\sigma_t/(2\pi s)\sqrt{\pi/a} \exp[b^2/(4a) - c]}{\alpha^*}. \end{aligned} \quad (8)$$

In expression (8) both the denominator and the numerator in the final term are "biased" upwards. The latter arise as an implication of Jensen's inequality (the pdf of y_t is convex in the tails). In Figure 2 we plot $-E_{t-1}(y_t|y_t \leq -\widehat{VaR}_t^\alpha)$, ES_t^α and the difference $-E_{t-1}(y_t|y_t \leq -\widehat{VaR}_t^\alpha) - ES_t^\alpha$ as functions of α and δ for three different levels on the return standard deviation: $\sigma_t = 1\%$, 2% and 5% (the conditional mean is set to zero).

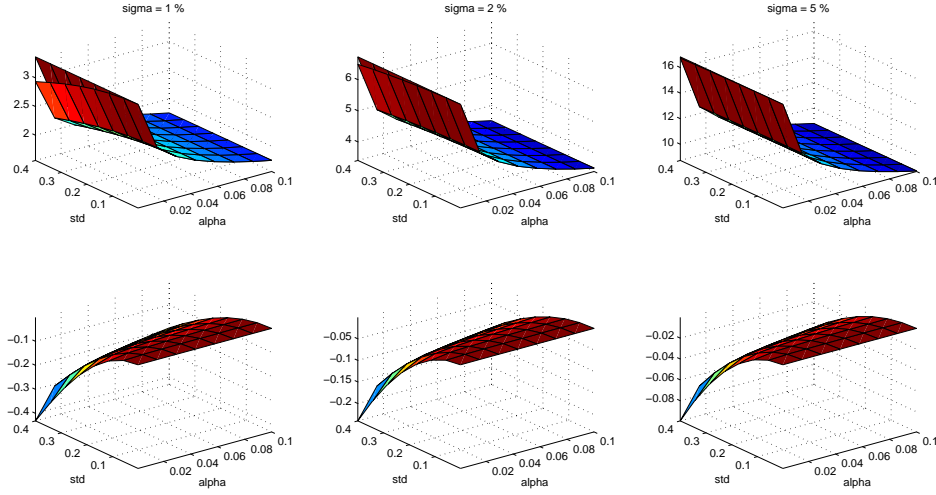


Figure 2: Effect on the interpretation. The top panel gives the levels of $-E_{t-1}(y_t | y_t \leq -\widehat{VaR}_t^\alpha)$ and ES_t^α as a function of the return standard deviation (std) and the probability level (alpha). The bottom panel gives the difference $-E_{t-1}(y_t | y_t \leq -\widehat{VaR}_t^\alpha) - ES_t^\alpha$.

The $-E_{t-1}(y_t | y_t \leq -\widehat{VaR}_t^\alpha)$ is throughout smaller than ES_t^α . The underlying intuition is quite clear cut and essentially the same as for the VaR case. For a VaR draw to the left of (minus) the true VaR exceedences are "large", while they are "small" for VaR draws to the right. Again, as the return density is positively sloped through the VaR density the latter will dominate and $-E_{t-1}(y_t | y_t \leq -\widehat{VaR}_t^\alpha)$ is smaller than the true ES . Again somewhat speculatively we expect smaller magnitudes for fat tailed return distributions. Indeed, the difference decreases with the return variance. We also note that the difference decreases with the probability level, whereas it increases with the estimation error (std in the figure).

When it comes to the backtesting of an ES predictor the direct corresponding way to the VaR case is to compare the average portfolio return on days of VaR exceedence to the corresponding average ES prediction these days (for more sophisticated ways see Berkowitz, 2001; Kerkhof and Melenberg, 2004; Wong, 2008). Thus, of interest is how $E_{t-1}(y_t | y_t \leq \widehat{VaR}_t^\alpha)$ compares to $E_{t-1}(\widehat{ES}_t^\alpha | y_t \leq \widehat{VaR}_t^\alpha)$. The $E_{t-1}(y_t | y_t \leq \widehat{VaR}_t^\alpha)$ was given in eq. (8) above and in the Mathematical Appendix we show that $E_{t-1}(\widehat{ES}_t^\alpha \mathbf{1}(y_t \leq \widehat{VaR}_t^\alpha) =$

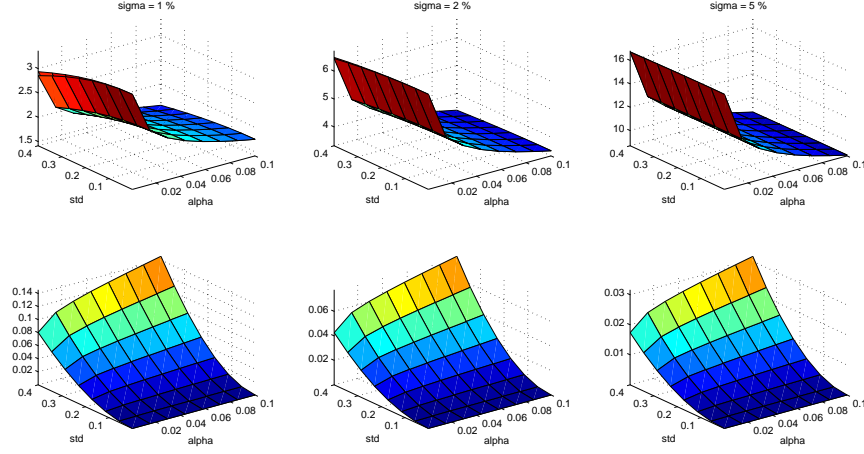


Figure 3: Effect on the back-testing. The top panel gives the levels of $-E_{t-1}(y_t | y_t \leq -\widehat{VaR}_t^\alpha)$ and $E_{t-1}(\widehat{ES}_t^\alpha | y_t \leq -\widehat{VaR}_t^\alpha)$ as a function of the return standard deviation (std) and the probability level (alpha). The bottom panel gives the difference $-E_{t-1}(y_t | y_t \leq -\widehat{VaR}_t^\alpha) - E_{t-1}(\widehat{ES}_t^\alpha | y_t \leq -\widehat{VaR}_t^\alpha)$.

$\sigma_t \alpha^* \phi(m)/\alpha - \alpha^* \mu_t + \delta_t/(2\pi s^*) \sqrt{\pi/a^*} \exp[b^{*2}/(4a^*) - c^*]$, where $a^* = (1 + 1/s^{*2})/2$, $b^* = -m/s^{*2}$ and $c^* = m^{*2}/(2s^{*s})$, and where $m^* = m\sigma_t/\delta_t$ and $s^* = \sigma_t/\delta$. Thus,

$$\begin{aligned} E_{t-1}(\widehat{ES}_t^\alpha | y_t \leq -\widehat{VaR}_t^\alpha) &= \frac{E_{t-1}(\widehat{ES}_t^\alpha \mathbf{1}(y_t \leq -\widehat{VaR}_t^\alpha))}{\Pr(y_t \leq -\widehat{VaR}_t^\alpha)} \\ &= ES_t^\alpha + \frac{\delta_t/(2\pi s^*) \sqrt{\pi/a^*} \exp[b^{*2}/(4a^*) - c^*]}{\alpha^*}. \end{aligned}$$

In Figure 3 we plot $-E_{t-1}(y_t | y_t \leq -\widehat{VaR}_t^\alpha)$, $E_{t-1}(\widehat{ES}_t^\alpha | y_t \leq -\widehat{VaR}_t^\alpha)$ and the difference $-E_{t-1}(y_t | y_t \leq -\widehat{VaR}_t^\alpha) - E_{t-1}(\widehat{ES}_t^\alpha | y_t \leq -\widehat{VaR}_t^\alpha)$ as functions of α and δ for three different levels on the return standard deviation: $\sigma_t = 1\%$, 2% and 5% (the conditional mean is again set to zero).

The $-E_{t-1}(y_t | y_t \leq -\widehat{VaR}_t^\alpha)$ is throughout larger than $E_{t-1}(\widehat{ES}_t^\alpha | y_t \leq -\widehat{VaR}_t^\alpha)$. It is interesting to see that the difference is of opposite sign as compared to the case above. Due to the estimation error $E_{t-1}(\widehat{ES}_t^\alpha | y_t \leq -\widehat{VaR}_t^\alpha) \neq ES_t^\alpha$ and since \widehat{ES}_t^α is simply a constant times \widehat{VaR}_t^α it will be "small" when \widehat{VaR}_t^α is and vice versa. Again, the

Table 1: Descriptive statistics for the index return series. Vce, Skew and Kurt are the sample variance, skewness and kurtosis, respectively. JB is the p-value in the Jarque-Bera normality test. $LB10$ and $LB10^2$ are p-values in the Ljung-Box test for autocorrelation in returns and squared returns, respectively. The Ljung-Box statistics were evaluated using 10 lags. Asy. is the p-value of the t-statistic for $\mathbf{1}(y_{t-1} < 0)y_{t-1}^2$ in the regression of y_t^2 on a constant, ten lags of y_t^2 and $y_{t-1}\min(y_{t-1}, 0)$.

Index	Obs.	Mean	Vce	Min	Max	Skew	Kurt	JB	LB ₁₀	LB ₁₀ ²	Asy.
CAC40	1278	-0.018	2.731	-9.472	10.595	0.168	9.719	0.000	0.000	0.000	0.000
DAX	1275	0.018	2.430	-7.433	10.797	0.167	10.406	0.000	0.007	0.000	0.000
FTSE 100	1261	0.001	2.193	-9.265	9.384	-0.080	9.764	0.000	0.000	0.000	0.000
Hang Seng	1261	0.028	3.852	-13.582	13.407	0.087	10.198	0.000	0.025	0.000	0.000
NIKKEI 225	1222	-0.042	3.392	-12.111	13.235	-0.528	11.000	0.000	0.172	0.000	0.000
S&P 500	1258	0.003	2.495	-9.470	10.957	-0.238	11.449	0.000	0.000	0.000	0.000

difference increases with the size of the estimation error and decreases with the return variance. The effect of the probability level is the opposite though.

4 Empirical illustration

To get an idea of the economic relevance of our findings we compute ES predictions along with estimates of the difference $-E_{t-1}(y_t|y_t \leq -\widehat{VaR}_t^\alpha) - ES_t^\alpha$ for six major stock market indices; CAC 40 (France), DAX (Germany), FTSE 100 (United Kingdom), Hang Seng (Hong Kong), Nikkei 225 (Japan) and S&P 500 (United States). Five years of daily index data was downloaded from Yahoo Finance covering the period May 16, 2006 to May 13, 2011. We calculate returns as $y_t = 100 \ln(I_t/I_{t-1})$, where I_t is the value of the index at time point t , and consider $\alpha = 0.01$. In Table 1 we give some descriptive statistics for the return series.

There is skewness and excess kurtosis in all return series. Consequently, the Jarque-Bera test rejects unconditional normality throughout. The Ljung-Box test indicates serial correlation in all series except Nikkei 225 and Straits Times, while ARCH effects with possible asymmetry are present in all series. As a reasonable specification for all series

we take

$$\begin{aligned} y_t &= \alpha_0 + \alpha_1 y_{t-1} + u_t, & u_t &= \sigma_t \varepsilon_t, & \varepsilon_t &\sim \text{nid}(0, 1), \\ \sigma_t^2 &= \beta_0 + \beta_1 u_{t-1}^2 + \beta_2 \sigma_{t-1}^2 + \beta_3 u_{t-1} \min(u_{t-1}, 0), \end{aligned} \quad (9)$$

The asymmetric conditional variance specification is due to Glosten, Jagannathan, and Runkle (1993) and gives negative return shocks an extra boost in the effect on future conditional variances. To guarantee a positive variance at all times we require that $\beta_0 \geq 0$, $\beta_1, \beta_2, \beta_3 > 0$. In the estimation of the model we employed maximum likelihood. Thus, with observations up to time T the log likelihood function takes the form.

$$\ln L \propto -\frac{1}{2} \sum_{t=1}^T \ln(h_t) - \frac{1}{2} \sum_{t=1}^T u_t^2 / h_t. \quad (10)$$

All estimations were carried out in the RATS 7.3 package using the built in BFGS algorithm for the maximization of (10)¹. We use robust standard errors throughout. To fulfill parameter restrictions we occasionally considered re-parameterizations. In particular, we set $\beta_i = \exp(\beta_i^*)$, $i = 0, 1, 3$ and $\beta_2 = 1/[1 + \exp(\beta_2^*)]$. The model (9) was estimated based on rolling estimation windows of, respectively, 250, 500, 750 and 1000 observations for all series. In Figures 4 and 5 we give the implied *ES* predictions in percentage points along with the estimated differences for the case of 500 observations. Details on how to compute the variance of the predictors are provided in the Appendix.

There is a quite similar pattern among the series regarding the *ES* predictions. All exhibit a strong time variation and sharply rise during the financial crises. The magnitudes of the estimated differences roughly track those of the corresponding *ES* predictions. Noteworthy is that they are quite substantial at times. In Table 2 we give some summariz-

¹In the few cases the BFGS-algorithm did not converge we estimated the parameters in the mean and variance specification separately. For the former we considered least squares, while for the latter we employed the simplex routine with ten iterations.

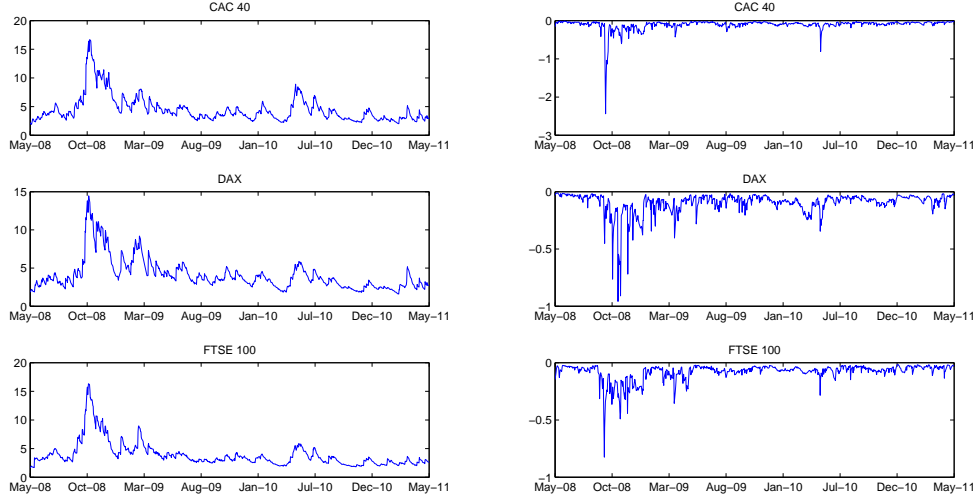


Figure 4: Estimates of ES_t^α (left panel) and the difference $-E_{t-1}(y_t | y_t \leq -\widehat{VaR}_t^\alpha) - ES_t^\alpha$ in percentage units.

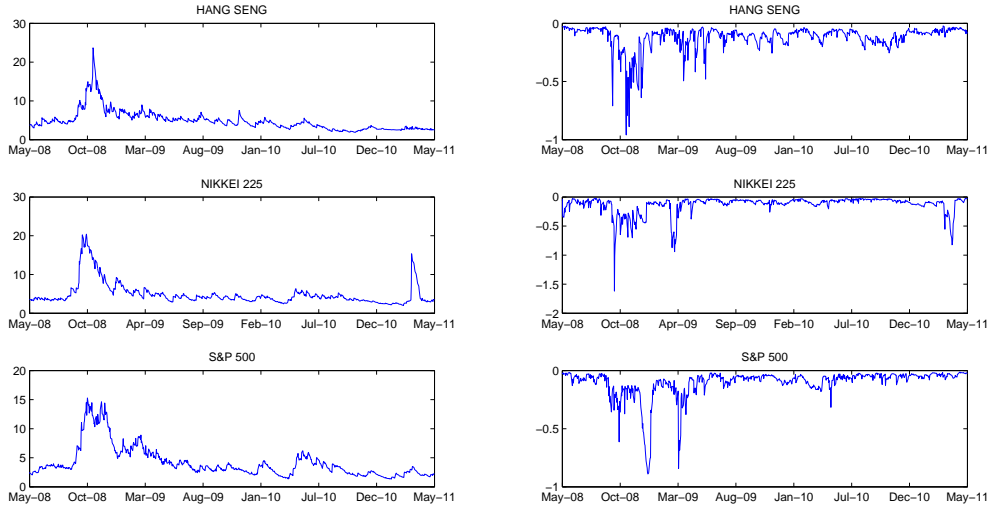


Figure 5: Estimates of ES_t^α (left panel) and the difference $-E_{t-1}(y_t | y_t \leq -\widehat{VaR}_t^\alpha) - ES_t^\alpha$ in percentage units.

Table 2: Descriptive statistics for estimated differences in percentage units. Vce is short for variance.

Index	$T = 250$					$T = 500$				
	Obs.	Mean	Vce	Min	Max	Obs.	Mean	Vce	Min	Max
CAC40	1029	-0.2460	0.0697	-3.2782	-0.0191	779	-0.1051	0.0222	-2.4388	-0.0137
DAX	1026	-0.2234	0.0408	-1.9495	-0.0144	776	-0.0965	0.0090	-0.9584	-0.0096
FTSE 100	1012	-0.2277	0.0494	-2.7240	-0.0201	762	-0.0808	0.0055	-0.8244	-0.0165
Hang Seng	1012	-0.3157	0.1093	-3.1564	-0.0207	762	-0.1218	0.0112	-0.9596	-0.0231
NIKKEI 225	973	-0.2785	0.0976	-2.8593	-0.0222	723	-0.1322	0.0242	-1.6185	-0.0246
S&P 500	1009	-0.2531	0.0904	-1.9646	-0.0148	759	-0.1019	0.0147	-0.8893	-0.0121

Index	$T = 750$					$T = 1000$				
	Obs.	Mean	Vce	Min	Max	Obs.	Mean	Vce	Min	Max
CAC40	529	-0.0472	0.0016	-0.6022	-0.0109	279	-0.0348	0.0014	-0.4620	-0.0078
DAX	526	-0.0408	0.0006	-0.2203	-0.0078	276	-0.0244	0.0003	-0.1595	-0.0054
FTSE 100	512	-0.0328	0.0004	-0.2224	-0.0096	262	-0.0224	0.0002	-0.1574	-0.0077
Hang Seng	512	-0.0548	0.0011	-0.2394	-0.0150	262	-0.0272	0.0001	-0.0675	-0.0110
NIKKEI 225	473	-0.0549	0.0028	-0.3745	-0.0115	223	-0.0455	0.0036	-0.2914	-0.0086
S&P 500	509	-0.0380	0.0005	-0.2145	-0.0064	259	-0.0260	0.0003	-0.1608	-0.0068

ing descriptives for the estimated differences for all sample sizes and indeces. As expected the sample size has a considerable impact on the size of the difference. In practice, the two smaller sample sizes are the ones that would typically be used. In particular, for a sample size of one year the differences are quite large even on average. As the sample size increases the magnitudes becomes smaller.

5 Conclusion

We studied the role of the estimation error in predictors of the ES and identified two important aspects where it may be of importance. On the one hand there is in the way the measure is interpreted and communicated and on the other there is in the way it is evaluated, or backtested. Interestingly, we found that while the effect on these are the same in the case of VaR they differ for ES . We found that the ES predictor overestimates that actual expected loss given VaR exceedence. In an empirical illustration we found this to be practically important. To get an ES prediction with the correct interpretation one could simply add the estimated difference to the conventional ES predictor. When

it comes to the backtesting of the measure we found that the average ES predictions are likely to be larger than the average returns on days of VaR exceedence. We leave for future work to incorporate this result in backtesting procedures for ES predictors. Of course, the analysis carried out here hinges on the normality of portfolio returns. Based on the underlying intuition we noted somewhat speculatively that the magnitudes are likely to be smaller for more realistic fat tailed distributions, though.

Mathematical Appendix

The actual expected shortfall

Let $X = -\widehat{VaR}_t^\alpha$. For notational convenience we drop the subscript on the expectations operator. Thus, $E(\cdot)$ should read expectation conditional on the information available in $t - 1$. With notations from the text and $\alpha^* = \Pr\{y_t \leq X\}$ we have

$$\begin{aligned}
E(y_t \mathbf{1}(y_t \leq X)) &= E[(\mu_t + \sigma_t \varepsilon_t) \mathbf{1}(\mu_t + \sigma_t \varepsilon_t \leq X)] \\
&= \mu_t E\left[\left(\frac{X - \mu_t}{\sigma_t}\right)\right] + \sigma_t E\left[\phi\left(\frac{X - \mu_t}{\sigma_t}\right)\right] \\
&= \alpha^* \mu_t + \sigma_t \int_{-\infty}^{\infty} \frac{1}{2\pi s} \exp\left(-\frac{z^2}{2}\right) \exp\left[-\frac{(z - m)^2}{2s^2}\right] dz \\
&= \alpha^* \mu_t + \frac{\sigma_t}{2\pi s} \int_{-\infty}^{\infty} \exp\left[-\frac{z^2 - (z - m)^2/s^2}{2}\right] dz \\
&= \alpha^* \mu_t + \frac{\sigma_t}{2\pi s} \int_{-\infty}^{\infty} \exp\left[-\frac{(1 + 1/s^2)}{2} z^2 + \frac{m}{s^2} z - \frac{m^2}{2s^2}\right] dz \\
&= \alpha^* \mu_t + \frac{\sigma_t}{2\pi s} \int_{-\infty}^{\infty} \exp[-az^2 - bz - c] dz \\
&= \alpha^* \mu_t + \frac{\sigma_t}{2\pi s} \sqrt{\frac{\pi}{a}} \exp[b^2/(4a) - c].
\end{aligned}$$

where $m = \Phi_\alpha^{-1}$, $s = \delta/\sigma_t$, $a = (1 + 1/s^2)/2$, $b = -m/s^2$ and $c = m^2/(2s^2)$. See Gradshteyn and Ryzhik (1994) for the final step.

The expected ES predictor given VaR breach

Re-using the notation above we first have $\widehat{ES}_t^\alpha = \widehat{VaR}_t^\alpha + (m + \phi(m)/\alpha)\sigma_t$. Then, $E_{t-1}(\widehat{ES}_t^\alpha | y_t \leq -\widehat{VaR}_t^\alpha) = \alpha^*(m + \phi(m)/\alpha)\sigma_t - E_{t-1}(X | y_t \leq X)$. Let Z be a standard

normally distributed variable. Then

$$\begin{aligned}
E(X\mathbf{1}(y_t \leq X)) &= E[(-VaR_t^\alpha - \delta_t Z)\mathbf{1}(y_t \leq X)] \\
&= \alpha^*(\mu_t + m\sigma_t) - \delta_t E[Z\mathbf{1}(y_t \leq \mu_t + m\sigma_t - \delta_t Z)] \\
&= \alpha^*(\mu_t + m\sigma_t) - \delta_t E\{E[Z\mathbf{1}(y_t \leq \mu_t + m\sigma_t - \delta_t Z)]|y_t\} \\
&= \alpha^*(\mu_t + m\sigma_t) - \delta_t E\{E[Z\mathbf{1}(Z \leq \frac{\mu_t + m\sigma_t - y_t}{\delta_t})]|y_t\} \\
&= \alpha^*(\mu_t + m\sigma_t) - \delta_t E\{E[Z\mathbf{1}(Z \leq y_t^*)]|y_t^*\} \\
&= \alpha^*(\mu_t + m\sigma_t) - \delta_t E[\phi(y_t^*)] \\
&= \alpha^*(\mu_t + m\sigma_t) - \delta_t \int_{-\infty}^{\infty} \frac{1}{2\pi s^*} \exp(-\frac{z^2}{2}) \exp[-\frac{(z - m^*)^2}{2s^{*2}}] dz \\
&= \alpha^*(\mu_t + m\sigma_t) - \frac{\delta_t}{2\pi s^*} \sqrt{\frac{\pi}{a^*}} \exp[b^{*2}/(4a^*) - c^*].
\end{aligned}$$

where $y_t^* = (\mu_t + m\sigma_t - y_t)/\delta_t$, $m^* = m\sigma_t/\delta_t$, $s^* = \sigma_t/\delta$, $a^* = (1 + 1/s^{*2})/2$, $b^* = -m^*/s^{*2}$ and $c^* = m^{*2}/(2s^{*s})$. We have $E_{t-1}(\widehat{ES}_t^\alpha | y_t \leq -\widehat{VaR}_t^\alpha) = \alpha^*(m + \phi(m)/\alpha)\sigma_t - \alpha^*(\mu_t + m\sigma_t) + \delta_t/(2\pi s^*) \sqrt{\pi/a^*} \exp[b^{*2}/(4a^*) - c^*] = \sigma_t \alpha^* \phi(m)/\alpha - \alpha^* \mu_t + \delta_t/(2\pi s^*) \sqrt{\pi/a^*} \exp[b^{*2}/(4a^*) - c^*]$.

Computation of the variance of the VaR and ES predictors

The variance, δ_t^2 , of the VaR predictor may be obtained from the delta method as follows (the corresponding variance of ES is simply a constant times this variance). The VaR predictor is given as in (5) with $\mu_t = \alpha_0 + \alpha_1 y_{t-1}$ and $\sigma_t^2 = \beta_0 + \beta_1 u_{t-1}^2 + \beta_2 \sigma_{t-1}^2 + \beta_3 u_{t-1} \min(u_{t-1}, 0)$. Thus, with Σ denoting the covariance matrix of the parameter estimator we have

$$\delta_t^2 = \frac{\partial VaR_t^\alpha}{\partial \theta'} \Sigma \frac{\partial VaR_t^\alpha}{\partial \theta},$$

where $\partial VaR_t^\alpha / \partial \theta = -\partial \mu_t / \partial \theta - \Phi_\alpha^{-1} \partial \sigma_t / \partial \theta$. In the present case we have $\partial \mu_t / \partial \theta = (1, y_{t-1}, 0, 0, 0)'$ and for $\partial \sigma_t / \partial \theta$ we may conveniently use the decomposition $\theta = (\theta'_1, \theta'_2)'$,

where $\boldsymbol{\theta}_1 = (\alpha_0, \alpha_1)'$ and $\boldsymbol{\theta}_2 = (\beta_0, \beta_1, \beta_2, \beta_3)'$. We have $\partial\sigma_t/\partial\boldsymbol{\theta}_i = (2\sigma_t)^{-1}(\partial\sigma_t^2/\partial\boldsymbol{\theta}_i)$, $i = 1, 2$, with $\partial\sigma_t^2/\partial\boldsymbol{\theta}_1$ and $\partial\sigma_t^2/\partial\boldsymbol{\theta}_2$ respectively following the recursions

$$\begin{aligned}\frac{\partial\sigma_t^2}{\partial\boldsymbol{\theta}_1} &= -2u_{t-1}(\beta_1 + \beta_3u_{t-1}\min(u_{t-1}, 0))(1, y_{t-1})' + \beta_2\frac{\partial\sigma_{t-1}^2}{\partial\boldsymbol{\theta}_1} \\ \frac{\partial\sigma_t^2}{\partial\boldsymbol{\theta}_2} &= (1, u_{t-1}^2, \sigma_{t-1}^2, u_{t-1}\min(u_{t-1}, 0))' + \beta_2\frac{\partial\sigma_{t-1}^2}{\partial\boldsymbol{\theta}_2}.\end{aligned}$$

References

- Artzner, P., Delbaen, F., Eber, J.-M., Heath, D., 1999. Coherent measures of risk. *Mathematical Finance* 9, 203–228.
- Basel Committee on Banking Supervision, 2005. Amendment to the Capital Accord to incorporate market risks. Bank for International Settlements, Basel, Switzerland.
- Basel Committee on Banking Supervision, 2006. International Convergence of Capital Measurement and Capital Standards: A Revised Framework - Comprehensive Version. Bank for International Settlements, Basel.
- Basel Committee on Banking Supervision, 2012a. Fundamental review of the trading book. Bank for International Settlements, Basel.
- Basel Committee on Banking Supervision, 2012b. Revisions to the Basel II market risk framework. Bank for International Settlements, Basel.
- Berkowitz, J., 2001. Testing density forecasts, with applications to risk management. *Journal of Business & Economic Statistics* 19, 465–474.
- Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–327.
- Christoffersen, P., 2003. *Elements of Financial Risk Management*. Academic Press, San Diego.
- Christoffersen, P., Gonçalves, S., 2005. Estimation risk in financial risk management. *Journal of Risk* 7, 1–28.
- Escanciano, J. C., Olmo, J., 2010. Backtesting parametric value-at-risk with estimation risk. *Journal of Business & Economic Statistics* 28 (1), 36–51.

- Glosten, L. R., Jagannathan, R., Runkle, D. E., 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* 48, 1779–1801.
- Gradshteyn, I. S., Ryzhik, I. M., 1994. *Tables of Integrals, Series and Products*, 5th Edition. Academic Press, New York.
- Hansen, B. E., 2006. Interval forecasts and parameter uncertainty. *Journal of Econometrics* 135, 377–398.
- Jorion, P., 1996. Risk²: Measuring the Risk in Value at Risk. *Financial Analysts Journal* 52, 47–56.
- Jorion, P., 2007. *Value at Risk: The New Benchmark for Managing Financial Risk*, 3rd Edition. McGraw-Hill, New York.
- Kerkhof, J., Melenberg, B., 2004. Backtesting for risk-based regulatory capital. *Journal of Banking and Finance* 28, 1845–1865.
- Lan, Y., Hu, Z. M., Johnson, J., 2007. Can value-at-risk be made more valuable? Resampling techniques for risk uncertainty. Working paper, Financial Studies Discipline, UWA Business School, The University of Western Australia.
- Lönnbark, C., 2010. A corrected value at risk predictor. *Applied Economic Letters* 17 (12), 1193 – 1196.
- Wong, W. K., 2008. Backtesting trading risk of commercial banks using expected shortfall. *Journal of Banking & Finance* 32, 1404–1415.
- Yamai, Y., Yoshida, T., 2005. Value-at-risk versus expected shortfall: A practical perspective. *Journal of Banking and Finance* 29, 997–1015.