You Can't Always Get What You Want — Estimating the Value at Risk from Historical Data with Limited Statistics

Björn Naundorf, Jörg Raufeisen, Christoph Bennemann*
Frederik Herzberg, d-fine Ltd 28 King Street London EC2V 8EH

Abstract

Historical simulation is a method widely used to calculate the VaR in the context of market risk, where a key challenge is that the VaR level of a portfolio must be estimated from a limited amount of data. We analyse the properties of different VaR-estimators by simulating different return distributions. Our analysis shows that the predictive power of a VaR estimate at high confidence levels with a limited amount of data should only be used with caution, because it can deviate substantially from the "real" VaR. Our results also confirm the theoretical results derived by Herzberg and Bennemann in a recent paper.

Keywords

value-at-risk, historical simulation, Monte Carlo simulation, order statistics

1 Introduction

The Value at Risk (VaR) is a measure to quantify the total risk of a single asset or a portfolio of assets. It is widely used by corporate treasures and fund managers as well as financial institutions. Especially in the framework of CAD 2 it has gained particular importance. VaR is aimed at making a statement of the form: "We are x% certain that we will not lose more than V over a given time horizon T", where the value V is given by the VaR of the portfolio.

Historical simulation is a method widely used to calculate the VaR in the context of market risk (see for instance Deutsch (2004) for a detailed discussion of the method). A problem often faced when estimating the VaR via historical simulation is the limited amount of available data. Consider the problem of estimating the 99% VaR using a historical time series of 250 days, which is a typical problem many financial institutions face. In this case, the estimated VaR would lie between the 2nd and 3rd smallest value of this time series (in theory at the 2.5 smallest value). It is not a priori clear which value—the 2nd, 3rd, or maybe even the average of both—does provide a suitable estimate for the 99% VaR.

In Herzberg and Bennemann (2006) it was demonstrated that order statistics can be used to uncover various properties of the VaR estimated by using the 2nd and 3rd (or more generally the kth) smallest value. Most of these properties, e.g. the expected quantile or its error, are independent of the portfolio or the return functions of the underlying assets[1]. However, if a linear combination of the 2nd and 3rd smallest value is used to estimate the VaR, few general statements can be made. The purpose of this paper is to bridge this gap by using computer simulations to compute the VaR of different portfolios using different estimators.

To this end we assessed the implied error by choosing the (i) second (ii) third or (iii) the mean of the second and third smallest values of the initial time series for an estimate of the VaR. We did this for three different portfolio scenarios: (i) A Gaussian distributed portfolio, (ii) a portfolio distribution with fat tails (modeled by a Pareto distribution), and (iii) a two asset portfolio composed of a stock and a put option on that stock, where we used a log-normal distribution for the price increments of the underlying stock.

It turned out, as expected, that the mean between the second and the third smallest value of the initial time series provided the most accurate estimator for the 99% VaR. For all three values, however, the error due to the limited amount of data was not negligible. The error was largest when taking the 3rd smallest value. In this case, one effectively measures a 98.5% VaR in 25%–30%—depending on the scenario—of all

*Inquiries regarding this work should be directed to Christoph Bennemann (christoph.bennemann@d-fine.de).

Dr Christoph Bennemann is a Senior Manager at d-fine, the consulting firm. In the past 8 years he has worked for clients across Europe, including major international banks and hedge funds, which he advises on quantitative risk management, regulatory compliance and system integration. His co-authors Dr Jörg Raufeisen and Dr Björn Naundorf both work as consultants at d-fine, while Dr Frederik Herzberg is a researcher at the University of Bonn.

estimations (which is in good agreement with Herzberg and Bennemann (2006), which predicts 27.5 %) which corresponds to an error of 50%. *But also for the case when the mean between the* 2nd and the 3rd smallest value was taken, the probability of effectively measuring a 98.5% VaR still was approximately 15%–20%.

To summarize, our analysis showed that an estimation of the VaR, in particular at large confidence levels > 95%, should be used with caution, as the implied errors can be substantial due to the limited amount of data.

2 Methods

To assess the 99% VaR, we considered three different scenarios:

(a) A Gaussian distributed portfolio (more precisely, a portfolio whose increments, i.e. its relative changes in value, all have a Gaussian distribution) with mean 0 and a standard deviation of 1. In this case the probability density function is given by:

$$P(V) = (\sqrt{2\pi})^{-1}\exp(-V^2/2)$$

(b) A portfolio with fat tails (more precisely, a portfolio whose increments in value all have distribution with fat tails). We assumed its increments to follow, up to a positive additive constant which we may assume to be zero (since we are only interested in the ranking of the losses, and adding a constant does not change the picture), a Pareto distribution:

$$P(V) = \frac{kV_m}{(-V)^{k+1}},$$

where the parameters V_m and k denote an offset value and the exponent of the power law decay of the distribution, respectively. For the following analysis, the value of V_m was fixed to -1 and the parameter k was changed between 1 and 2. The support of this distribution is thus restricted to $(-\infty\ , V_m]$.

(c) A Portfolio composed of a stock S with weight Δ and a European put option on that stock V,

$$\Pi = V + \Delta S$$

The stock price was assumed to be log-normally distributed,

$$S = S_0\exp(\sigma N(0,1)),$$

and the put option was priced using the Black-Scholes model. Parameters chosen were $S_0 = 1$, $\sigma = 0.2$ p.a., time to maturity $T = 250$ days, risk-less rate $r = 0.05$ p.a. and strike price $K = \exp(rT)$. The particular choice of r ensured that the option was at the money at expiry. The weight Δ was chosen such that the portfolio was Δ-hedged.

In all three cases, we simulated 250 portfolio values, where for each portfolio the 99% VaR was estimated using

(i) the second smallest value (E_2^{VaR}),
(ii) the third smallest value (E_3^{VaR}), and
(iii) the average between the second smallest and the third smallest value ($E_{2.5}^{VaR}$).

Based on this estimate, the **implied VaR level** was assessed. It was computed using the following scheme:

(i) 250 random numbers were drawn from the distributions according to the scenarios (a)–(c)
(ii) The estimates $E^{VaR}_{\{2,3,2.5\}}$ were assessed
(iii) Using the inverse cumulative density $F^{-1}_{\{a,b,c\}}$ of the distributions in (a)–(c), *the implied VaR level for each estimate was computed (in scenario (c),* F_c^{-1} was determined numerically):

$$VaR_{\text{implied}} = 1 - F^{-1}_{\{a,b,c\}}\left(E^{VaR}_{\{2,3,2.5\}}\right)$$

By repeating this scheme 5000 times, the cumulative distribution and a histogram of implied VaR levels was estimated.

3 Results

3.1 Gaussian Distributed Portfolio:

The following table summarizes the mean values and standard deviations of the VaR estimators $E^{VaR}_{\{2,3,2.5\}}$, computed using the implied VaR level distribution:

	2nd value	3rd value	(2nd + 3rd)/2
Mean	99.2	98.8	99.0
Std Dev	0.57	0.69	0.60

The results for the 2nd and 3rd smallest value are in excellent agreement with the theoretical results from Herzberg and Bennemann (2006). According to these results the expected VaR level should be 99.2 and 98.8 for the the 2nd and 3rd smallest value respectively. The theoretical results

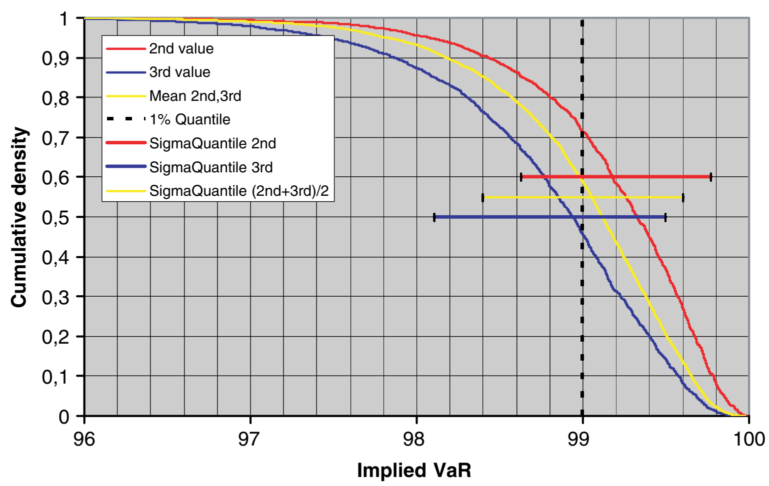

Figure 1.1: Cumulative distribution of implied VaR for the three estimators for a $E^{VaR}_{\{2,3,2.5\}}$ Gaussian distributed portfolio. The horizontal bars denote their mean ± standard deviation. The 99% level is shown as a dashed line.

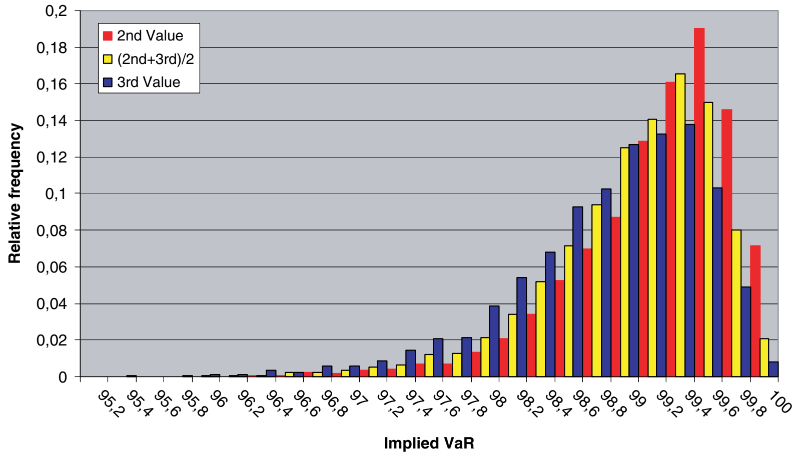

Figure 1.2: Histogram of the implied VaR values for a Gaussian distributed portfolio. Notice the asymmetric shape which gives rise to a large probability of large deviations in the implied VaR. Color code: See legend.

also extend to the standard deviation, predicted to be 0.56 and 0.69, again in good agreement with our simulation results.

As apparent from the cumulative density and the table above, once more the mean between the 2nd and 3rd smallest value is the best estimate for the 99% level. The probability of implicitly assuming a confidence level smaller than 98.5% and 98%—corresponding to a relative error of 50% and 100%, respectively—are given by:

	2nd value	3rd value	Mean 2nd + 3rd
98.5%	11.3% (11.0%)	27.5%(27.5%)	18.0%
98%	3.9% (3.9%)	12.9%(12.2%)	6.6%

The value in brackets show the values computed using the theoretical results from Herzberg and Bennemann (2006).

3.2 Fat Tailed Portfolio (Pareto Distribution):

The following table summarizes the mean values and standard deviations of the VaR estimators $E^{VaR}_{\{2,3,2.5\}}$, computed using the implied VaR level distribution:

	2nd value	3rd value	Mean 2nd + 3rd
Mean	99.2	98.8	99.0
Std Dev	0.56	0.68	0.59

As apparent from the cumulative density and the table above, the mean between the 2nd and 3rd smallest value is the best estimate for the 99% level. The probability of implicitly assuming a level smaller than 98.5% and 98%—corresponding to a relative error of 50% and 100%, respectively—are given by:

	2nd value	3rd value	Mean 2nd + 3rd
98.5%	12.0%	30.3%	18.1%
98%	4.3%	13.1%	7.4%

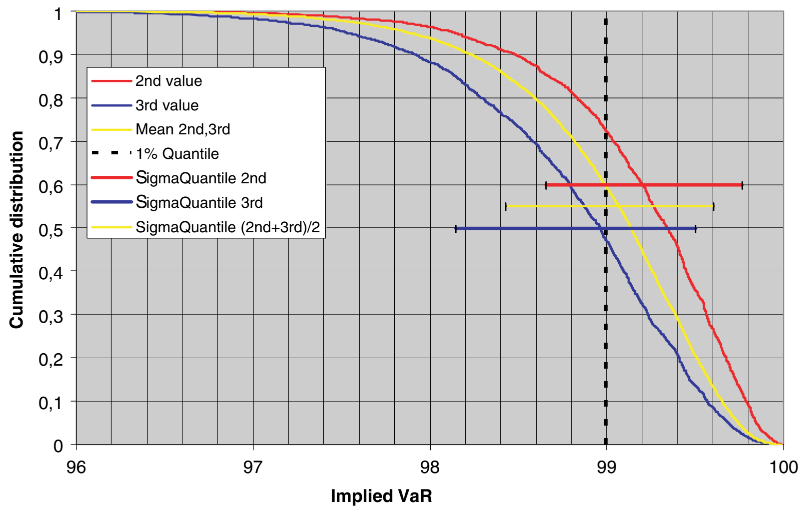

Figure 2.1: Cumulative distribution of implied VaR for the three estimators $E^{VaR}_{\{2,3,2.5\}}$ for a Pareto distributed portfolio The horizontal bars denote their mean $\pm$ standard deviation. The 99% level is shown as a dashed line.

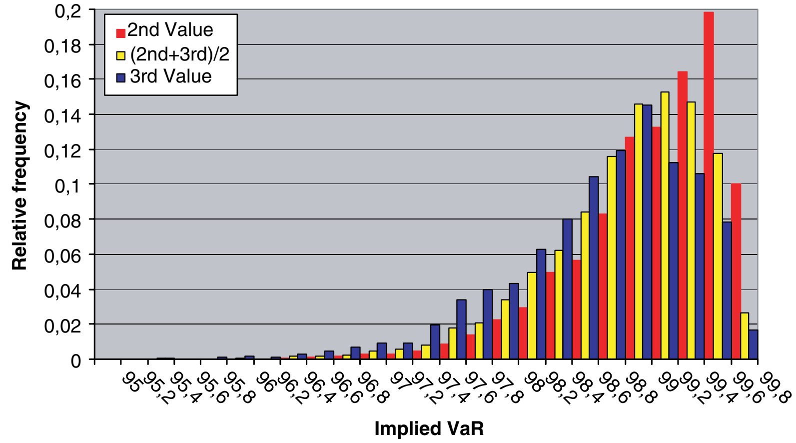

Figure 2.2: Histogram of the implied VaR values for a Pareto distributed portfolio. Notice the asymmetric shape which gives rise to a large probability of large deviations in the implied VaR. Color code: See legend.

3.3 Portfolio Composed of a Stock and an European Put Option (Δ Hedged)

The following table summarizes the mean values and standard deviations of the VaR estimators $E^{VaR}_{\{2,3,2.5\}}$, computed using the implied VaR level distribution:

	2nd value	3rd value	Mean 2nd + 3rd
Mean	99.2	98.8	99.0
Std Dev	0.55	0.66	0.58

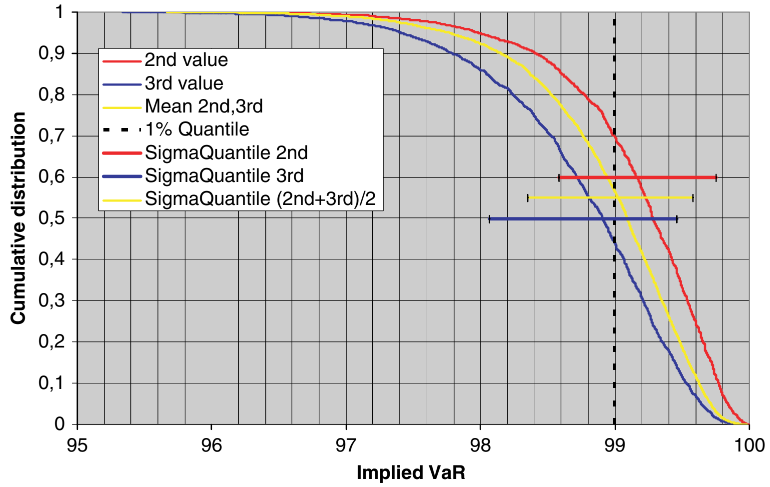

Figure 3.1: Cumulative distribution of implied VaR for the three estimators $E^{VaR}_{\{2,3,2.5\}}$ for a portfolio composed of a stoch and a European put option. The horizontal bars denote their mean ± standard deviation. The 99% level is shown as a dashed line.

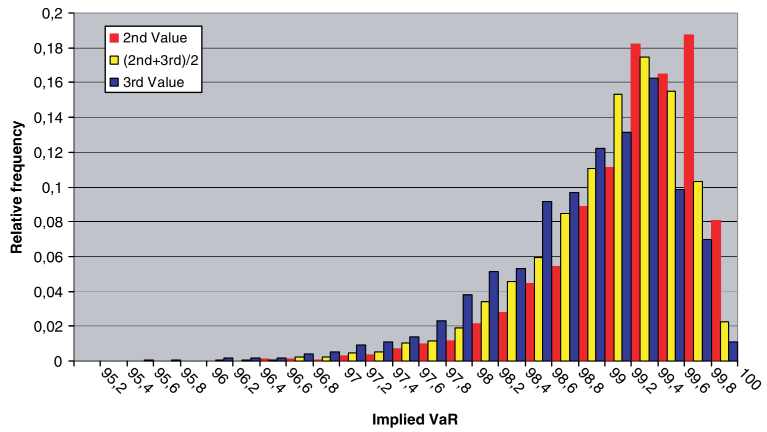

Figure 3.2: Histogram of the implied VaR values for a portfolio composed of a stoch and an European put option. Notice the asymmetric shape which gives rise to a large probability of large deviations in the implied VaR. Color code: See legend.

As apparent from the cumulative density and the table above, the mean between the 2nd and 3rd smallest value is the best estimate for the 99% level. The probability of implicitly assuming a level smaller than

	2nd value	3rd value	Mean 2nd + 3rd
98.5%	24.9%	14.9%	99.0
98%	3.1%	9.6%	5.2%

98.5% and 98%—corresponding to a relative error of 50% and 100%, respectively—are given by:

4 Discussion

For all three scenarios, the mean between the second and third value provided the best estimate for the 99% VaR. Consistently, the standard deviation of the implied VaR distribution was approximately 0.06. Altogether our results for the 2nd and 3rd smallest value were in good agreement with the theoretical predictions from Herzberg and Bennemann (2006). In particular, the VaR level associated with these values (and its standard deviation) proved to be independent of the type of distribution function and portfolio used in the simulations.

Many practitioners tend to use the third heaviest loss as an estimate for the 99% VaR. Our analysis and the theoretical results from Herzberg and Bennemann (2006) reveal, however, that this choice is particularly bad for two reasons: (i) The mean of the estimator is systematically smaller than the 99% level and (ii) the width of the distribution is very large. In particular the latter finding provides a serious issue of concern: For approximately 30% of all cases, the implied error by taking the third smallest value was even larger than 50% (hence, in about 30% of all cases one measured a 98.5% VaR level or worse, significantly underestimating the true risk of the portfolio). This reduced to approximately 15%-20% and 10%-15% of all cases when the mean between the second and the third, or the second smallest value is taken, respectively.

Taken together, our analysis showed that the predictive power of a VaR estimate at high levels (here: 99%) with a limited amount of data (here: 250 values), should only be used with extreme caution. The results further demonstrate that the implied VaR can deviate substantially from the "real" VaR which was intended to be measured in the first place.

5 Acknowledgements

We would like to thank our employer d-fine for supporting this work, and Hans-Peter Deutsch and Andreas Werner for valuable discussions.

FOOTNOTES & REFERENCES

1. In this paper we use a slightly different notation to Herzberg and Bennemann (2006), what is called 1 %-quantile in Herzberg and Bennemann (2006) is what we mean by 99% VaR, etc.

■ H.-P. Deutsch, "Derivatives and Internal Models" (third edition), Palgrave Macmillian 2004
■ F. S. Herzberg, Ch. Bennemann: "Order statistics for Value at Risk estimation and option pricing", Wilmott 26 (2006).