Using Order Statistics to Estimate Confidence Intervals for Probabilistic Risk Measures

KEVIN DOWD

KEVIN DOWD
is a professor of financial risk management in the Centre for Risk and Insurance Studies at Nottingham University Business School in Nottingham, UK.
kevin.dowd@nottingham.ac.uk

This article shows how the theory of order statistics can be used to estimate confidence intervals for probabilistic risk measures, most particularly, VaR and expected shortfall. The method is very general in that it can be applied to any loss distribution, parametric or nonparametric. The method is straightforward to implement, and has a number of other advantages compared to alternatives methods of estimating confidence intervals for risk measures.

One of the most important tasks faced by risk managers is to ensure that estimates of risk measures are reliable. This requires that risk managers be able to evaluate the precision of their estimated risk measures, and perhaps the best way to evaluate precision is to estimate confidence intervals. However, estimating confidence intervals is widely regarded as problematic, especially when dealing with confidence intervals for some of the more recently developed risk measures, such as coherent risk measures.

This article proposes a partial solution to this problem by suggesting that confidence intervals for probabilistic risk measures can be estimated using the theory of order statistics (OS).[1] An order statistic is the k^{th} highest (or alternatively, lowest) observation in a sample, and the theory of order statistics is well developed in the statistical literature. The solution proposed here is a partial one in so far as it is limited to probabilistic risk measures (i.e., risk measures that can be defined in terms of a confidence interval or probability); however, it is very useful in that the majority of risk measures actually used by financial institutions do, in fact, fall into this class, including VaR and expected shortfall (ES).

The OS approach is well suited for estimating confidence intervals of probabilistic risk measures for a number of reasons:

1. The OS approach is very general, in the sense that it can be applied to *any* loss distribution, whether parametric or nonparametric.
2. The OS approach is very general, in the sense that it can be used to estimate confidence intervals for any probabilistic risk measure; hence, the confidence intervals for the ES can be estimated using a similar approach to that used to estimate confidence intervals for the VaR.
3. The OS approach is applicable even for relatively small samples, because it is not based on asymptotic theory.
4. The OS approach does not force estimated confidence intervals to be symmetric, which can be a problem when dealing with extremes, where confidence intervals are asymmetric.
5. The OS approach is easy to implement on a computer.

These advantages make the OS approach superior to the obvious alternative methods of estimating confidence intervals for risk measures. The OS approach is superior to analytical methods, which are few and far between, limited to restrictive distributional assumptions, and only applicable (where they are applicable at all) to the VaR risk measure. The OS method is superior to the bootstrap in so far as it is free of resampling error (though not sampling error) and of the potential biases of bootstrap methods. It is also superior to methods that estimate confidence intervals using estimates of the standard errors of risk-measure estimators: such methods rely on large sample theory, impose symmetric confidence intervals that are not always appropriate, often assume normality, and can be inaccurate, difficult to implement, and dependent on ancillary assumptions.[2]

This article is organized as follows. The first section explains how the theory of order statistics can be used to estimate confidence intervals for the VaR, the next section explains how OS can be used to estimate confidence intervals for the expected shortfall measure, and the last section concludes.

USING ORDER STATISTICS TO ESTIMATE CONFIDENCE INTERVALS FOR VaR

Suppose we have $x_1, x_2, \ldots, x_n$ observations from some distribution (or cumulative density) function $F(x)$ with r^{th} order statistic $x_{(r)}$ and where $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$. The probability that j of our n observations do not exceed a fixed value x obeys the binomial distribution

$$\Pr\{j \text{ observations} \leq x\} = \binom{n}{j}[F(x)]^j[1-F(x)]^{n-j} \qquad (1)$$

Hence, the probability that at least r observations in the sample do not exceed x is also a binomial,

$$G_r(x) = \sum_{j=r}^{n}\binom{n}{j}[F(x)]^j[1-F(x)]^{n-j} \qquad (2)$$

$G_r(x)$ is, therefore, the distribution function of our order statistic.[3]

However, the VaR is itself an order statistic. Suppose our observations are denominated in units where profits take positive values and losses take negative ones, and suppose we define the α VaR as the negative of the $100\alpha\%$ quantile of $F(x)$, where α is the tail probability.[4] So, for example, if $\alpha = 5\%$ and $n = 1{,}000$, we can take the 5% VaR as the negative of the 950th largest observation.[5] More generally, we can take the αVaR to be equal to the negative of the r^{th} largest observation, where r is equal to $(1-\alpha)n$. Thus, $G_r(x)$ is also the distribution function for (the negative of) the αVaR, where $\alpha = 1 - r/n$.

This VaR distribution function equation given by equation (2) is very useful because once empirically calibrated it gives us estimates of the VaR confidence intervals.[6] Suppose we calibrate $F(x)$ using empirical data and call this calibrated distribution function $\hat{F}(x)$.[7] If we now replace $F(x)$ in equation (2) with $\hat{F}(x)$, we then derive a calibrated $G_r(x)$ function $\hat{G}_r(x)$, as specified in equation (3):

$$\hat{G}_r(x) = \sum_{j=r}^{n}\binom{n}{j}[\hat{F}(x)]^j[1-\hat{F}(x)]^{n-j} \qquad (3)$$

$\hat{F}_r(x)$ and $\hat{G}_r(x)$ can be regarded as estimates of $F(x)$ and $G_r(x)$ based on available data, and we wish to use them to construct an estimate of the confidence interval for our VaR, which will be unknown.[8] The situation is analogous to the textbook situation where we might have a sample mean and a sample variance, and wish to estimate a confidence interval for the "true" unknown mean or variance.

Now suppose that we wish to estimate the 90% confidence interval for a VaR. Applying equation (3) then tells us that the 90% confidence interval for (the negative of) the VaR has a lower bound equal to the 5th-percentile point of $\hat{G}_r(x)$ and an upper bound equal to the 95th-percentile point of $\hat{G}_r(x)$. Thus, the VaR itself has a lower bound equal to the negative of the 95th-pecentile point of $\hat{G}_r(x)$ and an upper bound equal to the negative of the 5th-percentile point of $\hat{G}_r(x)$.

To illustrate, suppose we wish to estimate the 90% confidence interval of a normal 95% VaR for a sample size of $n = 1{,}000$, based on an estimated mean and standard deviation of 0 and 1, respectively. To do so, we first use equation (3) to solve for the $\hat{F}_r(x)$ values that correspond to the 5th- and 95th-percentiles of $\hat{G}_r(x)$. This gives us $\hat{F}(x)$ values of 0.0392 and 0.0618, respectively. We then treat these values of $\hat{F}(x)$ as cumulative probabilities and infer the corresponding quantiles: $\hat{F}(x) = 0.0392$ implies $x = \hat{F}^{-1}(0.0392) = \Phi^{-1}(0.0392) = -1.7600$, where $\Phi(.)$ indicates the value of the standard normal cumulative density function, and $\hat{F}(x) = 0.0618$ implies $x = \hat{F}^{-1}$

$(0.0618) = \Phi^{-1}(0.0618) = -1.5398$. These x values mark the boundaries of the 90% confidence interval for the negative of the VaR. Hence, the 90% confidence interval of the VaR itself is [1.5398, 1.7600].

The calculation process therefore has two main steps: a) we (numerically) solve equation (3) for the two implied values of $\hat{F}(x)$, and then b) we regard these $\hat{F}(x)$ values as confidence levels and obtain the corresponding VaRs. This latter step is carried out by inverting the two $\hat{F}(x)$ values to find their corresponding x values, which we then multiply by minus 1.

Step (a) can be carried out using a suitable numerical search method (e.g., a bisection algorithm[9]) and is always the same, because it is based on the same underlying binomial distribution. Step (b) is elementary. Thus, the calculations are straightforward to program.

It is important to appreciate that this approach is very general and gives us confidence intervals for *any* specified $\hat{F}(x)$: the distribution function might be any type of parametric distribution function (e.g., normal, t, extreme-value, etc.) or nonparametric distribution function (e.g., an empirical cdf based on historical simulation).

USING ORDER STATISTICS TO ESTIMATE CONFIDENCE INTERVALS FOR EXPECTED SHORTFALL

The OS approach can be generalized further to give us the distribution function (and, hence, confidence intervals) for any probabilistic risk measure. An example of such a measure is expected shortfall (ES)[10]

$$ES = -\frac{1}{\alpha}\int_0^{\alpha} q_p dp \tag{4}$$

where q_p is the 100th p-percentile of $F(x)$.

We can estimate the ES confidence intervals using an approach that is very similar to that used for VaR confidence intervals. Step (a) is identical, and the only difference arises in step (b), where we map our $\hat{F}(x)$ values to their ESs rather than to their VaR values.

To illustrate, suppose we wish to estimate the ES confidence interval corresponding to our earlier VaR interval (i.e., so we want the 90% confidence interval, $n = 1{,}000$ and profits/losses are standard normal). Because step (a) is exactly the same as before, this means that we can use the same two $\hat{F}(x)$ values as before (i.e., 0.0392 and 0.0618). We now treat these values of $\hat{F}(x)$ as if they were confidence levels and infer the corresponding ESs, using the relevant formula for the normal ES. If our confidence level is α, the relevant formula for the normal ES is

$$\alpha\, ES = -\mu + \frac{\sigma}{\alpha}\phi(z_{1-\alpha}) \tag{5}$$

where $\phi(.)$refers to the value of the standard normal density function, $z_{1-\alpha}$ refers to the value of the standard normal variate at the cumulative probability α, and μ and σ are equal to 0 and 1, respectively.[11] We now use the ES formula in (5) with α set to 0.0392 and 0.0618 to obtain the bounds of the ES confidence interval,

$$\frac{\phi(z_{1-0.0618})}{0.0618} = \frac{\phi(z_{0.9382})}{0.0618} = 1.9726$$

and (6)

$$\frac{\phi(z_{1-0.0392})}{0.0392} = \frac{\phi(z_{0.9608})}{0.0392} = 2.1625$$

CONCLUSION

The OS approach allows us to obtain estimates of confidence intervals for any probabilistic-based risk measure, most obviously VaR and ES. The method suggested here requires only two calculation ingredients—a bisection algorithm and an engine to estimate the risk measure itself. The method is very general in that it can be applied to any parametric or nonparametric distribution. It also has the attraction that it uses exact statistics and avoids relying on asymptotic theory.[12] And, from a practical point of view, it has the attractions of being straightforward to implement and of giving answers in a fraction of a second.[13]

ENDNOTES

This research was carried out under the auspices of an Economic and Social Research Council (ESRC) research fellowship on "risk measurement in financial institutions," and the author thanks the ESRC for their support. He also thanks John Cotter, Steve Figlewski, and Ghulam Sorwar for helpful comments. The usual caveat applies.

[1]The use of order statistics to estimate confidence intervals for the VaR was suggested by Dowd [2001], based on

earlier work by Kendall and Stuart [1973] and Reiss [1989]. The present article provides a more complete treatment of the OS approach, and shows how it can be applied to estimate confidence intervals for the ES as well as for the VaR.

[2]For example, expressions for the VaR confidence intervals can be derived from statistical theory relating to quantile standard errors (e.g., as in Kendall and Stuart [1972, pp. 251-252]). However, such methods seem to produce quite inaccurate confidence intervals and are dependent on supplementary assumptions about bin width (see Dowd [2005, chapter 3.5.1]).

[3]See, e.g., Kendall and Stuart [1973, p. 348] or Reiss [1989, p. 20].

[4]These assumptions are arbitrary and can easily be altered. For example, we might take our units to be denominated so that losses have positive values and profits have negative ones and/or we might define α to be 1 minus the tail probability. Such changes would alter some of our later calculations and formulas, but make no difference to any substantial results.

[5]Strictly speaking, there is a question over which order-statistic observation best corresponds to the α VaR. The VaR itself corresponds to the quantile that demarcates tail observations from non-tail ones. With 1,000 observations, we therefore have 50 observations in the tail and 950 outside of it. This means that (the negative of) the VaR lies somewhere between the 50th smallest and the next biggest ordered observations (i.e., between the 950th largest and 949th largest observations). We can, therefore, take the VaR to be the negative of either of these observations or the negative of some weighted average of them. However, for convenience, we follow the common practice of taking the VaR to be the negative of the 950th largest observation and note that this might produce a slight upward bias in our VaR estimate.

[6]The reason for empirically calibrating our distribution functions is as follows. If we knew the true distribution functions $F(x)$ and $G_r(x)$, we could use equation (2) to give us the distribution function of random order statistics considered *ex ante*, i.e., before we sample them. In such cases, the distribution functions are treated as known, but the samples and any resulting sample estimates are treated as unknown. The VaR is also implicitly known and so too are the confidence intervals, but these confidence intervals relate to samples of random order statistics. However, in practical risk management contexts the situation is typically the other way round—the sample and any resulting sample estimates are known, but the "true" distribution functions are not. In the latter case, we have to work with calibrated distribution functions (i.e., equation (3)) to estimate the confidence interval of the (now unknown) VaR.

[7]This calibrated distribution function might be parametric or nonparametric. In the former case, it would be the relevant distribution (e.g., normal, t, etc.) calibrated using empirical estimates of the relevant parameters. In the latter case, it might be some kind of empirical distribution function (e.g., such as the type of historical distribution underlying historical simulation approaches to VaR).

[8]Since this latter confidence interval is estimated, it is subject to the vagaries of sampling variation (i.e., it is subject to parameter estimation risk). There will, therefore, always be a concern about how accurate any estimate might be. A possible response to this concern is given in note 12 below.

[9]More particularly, we use the bisection algorithm to find that value of $\hat{F}(x)$ in equation (3), such that $\hat{G}_r(x)$ matches the desired percentile. So, for example, if we want the 5th percentile, we use the bisection search routine to find that value of $\hat{F}(x)$ such that $\hat{G}_r(x) = 0.05$.

[10]The definition of ES given here closely follows that of Acerbi [2004] and can be interpreted as the expected loss conditional on a tail event occurring. It closely related to the Tail Conditional Expectation (TCE), which is given by $-\mathrm{E}[X \mid X \leq q_\alpha$ for loss variable X, and is interpreted as the expected loss, conditional on experiencing a loss exceeding VaR. In the former case, the loss is conditional on a threshold delimited in probability, and in the latter case, the loss is conditional on a threshold specified in terms of loss amount. The two measures will coincide when $F(x)$ is continuous, but can otherwise differ. I prefer to work with ES because it is coherent, whereas the TCE is not coherent, except where $F(x)$ is continuous and, hence, the same as ES. For more on these issues, see Acerbi [2004, pp. 157-165].

[11]Needless to say, we should only use equation (5) if $\hat{F}(x)$ is normal. If $\hat{F}(x)$ is parametric but not normal, then we should use the formula relevant to the appropriate parametric distribution (where it exists) or we should use an appropriate numerical algorithm that estimates ES from the weighted average of (parametric) tail quantiles. Where $\hat{F}(x)$ is nonparametric, we can estimate the ES from the weighted average of (empirical) tail quantiles. For more on these issues, see, e.g., Dowd [2005; chapters 3.4 and 6] and McNeil et al. [2005, pp. 45-46 and 283].

[12]However, as mentioned earlier (see, e.g., note 8 above), our estimates of the VaR or ES confidence intervals are subject to parameter estimation risk and may, therefore, be inaccurate or imprecise. One response to this problem is to use Monte Carlo simulation to gauge the precision of these estimates. For example, if $\hat{F}(x)$ is normal with estimated mean $\hat{\mu}$ and estimated standard deviation $\hat{\sigma}$, then we can simulate a set of new values of μ and σ from the estimated normal distribution. We then use this new set of parameter values which enables us to estimate the bounds of the confidence interval using the approach outlined in the article. If we carry out this exercise m times, say, we then get m estimates of each of the two bounds, we can use these bounds to estimate the standard errors of the bound estimates. However, since estimation errors in the lower and upper bounds of the confidence interval would typically be related, it might be better to focus on their difference (i.e.,

the width of the confidence interval) and use Monte Carlo to estimate the standard error of the width of the confidence interval. Naturally, where $\hat{F}(x)$ is not normal, we should carry out simulations from the particular (i.e., non-normal parametric or nonparametric) distribution we are dealing with.

[13]The method can also be refined further. One way to do so is to apply a Richardson extrapolation method, which is essentially a weighted average of estimates based on alternative sample sizes. For more on such an approach, see Inui and Kijima [2005].

REFERENCES

Acerbi, C. "Coherent Representations of Subjective Risk Aversion." In G. Szegö, ed. *Risk Measures for the 21st Century.* New York: Wiley, 2004, pp. 147–207.

Dowd, K. "Estimating VaR with Order Statistics." *Journal of Derivatives,* Vol. 8, No. 3 (2001), pp. 23–30.

——. *Measuring Market Risk,* 2nd ed. Chichester, U.K., and New York, NY: John Wiley and Sons, 2005.

Inui, K., and M. Kijima "On the Significance of Expected Shortfall as a Coherent Risk Measure." *Journal of Banking and Finance,* Vol. 29 (2004), pp. 853–864.

Kendall, M., and A. Stuart. *The Advanced Theory of Statistics. Vol. 1: Distribution Theory,* 4th ed. London: Charles Griffin and Co. Ltd, 1972.

——. *The Advanced Theory of Statistics. Volume 2: Inference and Relationship,* 3rd ed. London: Griffin, 1973.

Mc Neil, A.J., R. Frey, and P. Embrechts. *Quantitative Risk Measurement.* Princeton, NJ: Princeton University Press, 2005.

Reiss, R.-D. *Approximate Distributions of Order Statistics with Applications to Nonparametric Statistics.* Berlin: Springer Verlag, 1989.