MOODY'S ANALYTICS

B&H RESEARCH
NOVEMBER 2013

Efficient Statistical Estimation of 1-year VaR Economic Capital

Authors
Steven Morrison PhD
Laura Tadrowski PhD

Contact Us

Americas
+1.212.553.1658
clientservices@moodys.com

Europe
+44.20.7772.5454
clientservices.emea@moodys.com

Asia (Excluding Japan)
+85 2 2916 1121
clientservices.asia@moodys.com

Japan
+81 3 5408 4100
clientservices.japan@moodys.com

Overview

Estimation of 1-year Value-at-Risk capital requirements using simulation involves the statistical estimation of a tail (e.g. 99.5th) percentile. Much work has been done in academia and banking on how to efficiently estimate percentile statistics when using simulation. Interestingly, this work has generally not found its way into insurers' economic capital model implementations (yet). In this paper we survey some of the most popular statistical methods that have been developed and implemented for percentile estimation, and we use an insurance case study to investigate the potential effectiveness of these techniques in the specific context of a 99.5% 1-year VaR capital assessment.

Contents

1. Introduction

Value at Risk (VaR) is a commonly used measure of financial risk, in particular forming the basis of the many insurance firms' economic and regulatory capital requirements. VaR is defined as a *quantile* (equivalently, *percentile*) of the distribution of potential financial losses at some future time horizon, typically at an 'extreme' probability level. For example, under Solvency II the Solvency Capital Requirement is defined as the 99.5% worst change in net assets at a one-year time horizon.

Calculation of VaR-based capital requirements thus involves the statistical estimation of quantiles. In an Internal Model approach, this estimation is performed using simulation. Losses are estimated under a large number of one year, real-world, economic scenarios and the quantile of interest is estimated from these simulated losses. For example, the 99.5th percentile can be estimated by ranking the losses and picking out the loss in the 99.5% ranked scenario.

The use of simulation gives rise to statistical error in the estimated quantile. For any particular finite number of scenarios, the estimated quantile will not necessarily equal the population quantile (i.e. the quantile that would be obtained in the limit as the number of scenarios tends to infinity) and the estimated quantile will change if we use a different set of economic scenarios (by regenerating scenarios using a different random number seed for example).

The question of statistical error also raises the question of how best to efficiently estimate extreme quantiles. The simple estimator described above, whereby we rank and pick out a single scenario, is simple and intuitive. However, it is just one of a number of possible estimation approaches.

The aim of this paper is to help raise awareness of statistical error in the quantile estimates used in calculation of 1-year VaR economic capital requirements, and some of the alternative estimation techniques that can be used. The paper is structured as follows:

- Section 2 describes various methods for quantile estimation.
- Section 3 presents a case study, where the different methods are applied and compared using an example insurance business.
- Section 4 concludes.

Appendices A and B provide further technical details.

2. Methods for efficient statistical estimation of quantiles

This section presents some statistical methodologies for efficient estimation of quantiles when using simulation. The first method of empirical estimators might be considered as the 'brute force' method, and this approach is prevalent in insurers' current economic capital model implementations. We then go on to present two further more sophisticated methods that may provide a more statistically efficient estimate of the quantile.

Empirical Estimators

Suppose we have simulated n scenarios, with estimated values X_i $(i = 1, \dots, n)$. As discussed in the introduction, a simple and intuitive approach to estimating quantiles is to rank the scenarios in increasing order and read off the single scenario corresponding to the probability of interest.

Formally, we define *order statistics* $X_{(i)}$ by ranking X_i in increasing order i.e. $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$. The *empirical estimator* for the p'th quantile is then the single order statistic:

$$\hat{Q}_{Empirical}(p) = X_{(\lceil np \rceil)}$$

where $\lceil np \rceil$ is the smallest integer not less than np. If np is an integer we pick out the single loss with rank np. For example, with $n = 1{,}000$ scenarios, if we are interested in the $p = 99.5\%$ VaR, we pick out the scenario with the 995th largest loss[1].

Clearly, for any finite number of scenarios n, the empirical quantile will be subject to statistical error. If we repeat the procedure using an independent set of n scenarios (by regenerating economic scenarios using a different random number seed for example) our estimated quantile will change. In the $n = 1{,}000$ scenario example, the quantile is estimated as the 6th worst loss and we would certainly expect the values of the 6 worst losses to change if we generated these using another set of 1,000 scenarios.

L-statistics

In picking a single order statistic to estimate a quantile, we've actually made a subjective choice. As an alternative to choosing the 995th loss when ranked in increasing order, we could pick the 5th loss when ranked in decreasing order (which corresponds to the 996th loss when ranked in increasing order), or we could take some weighted average of the 995th and 996th losses.

More generally, we can consider *L-statistics* (H. Mausser 2001, Sheather and Marron 1990), defined as weighted averages of all order statistics:

$$\hat{Q}_L(p) = \sum_{i=1}^{n} w_i \, X_{(i)}$$

The empirical estimator can then be considered as a particular type of L-statistic with all weight placed on a single order statistic i.e.:

$$w_i = \begin{cases} 1 & i = \lceil np \rceil \\ 0 & otherwise \end{cases}$$

The motivation for adopting more general weighting schemes, which put weight on more than one order statistic, is to smooth out statistical error across different order statistics in the hope of achieving an estimator with a lower statistical error.

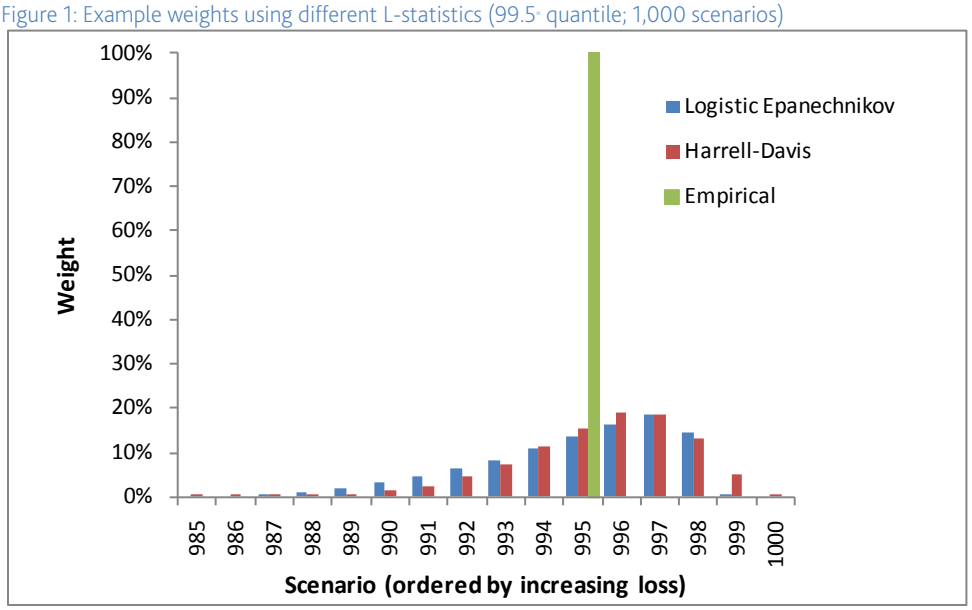
For example, Figure 1 compares weights under three different schemes: Empirical, Harrell-Davis and Logistic Epanechnikov. While the empirical estimator puts all weight on the 995th order statistic, the more general L-estimators put weight on a number of order statistics in the vicinity of the 995th order statistic. For example, the Logistic Epanechnikov estimators puts some weight on all order statistics from the 988th to the 999th in its estimate of the 99.5th percentile, with the greatest weight being put on the 996th order statistic.

[1] More generally, if np isn't an integer, we round up to the next integer and pick out the corresponding loss. For example, with $n = 500$ scenarios we would pick out the 498th largest loss.

Note that in practice, we have some choice in the parameterization of the Logistic Epanechnikov weights – we can spread the weights more widely (thus including an even wider range of order statistics in the quantile estimate) or narrowly (taken to the extreme, the Logistic Epanechnikov estimator is equally weighted on the 995th and 996th order statistics only). Intuitively, by spreading the weights more widely we use more order statistics in the estimate and thus expect a more stable estimate, though potentially at the expense of increased bias (as any individual order statistic will provide a biased estimate of the specific quantile of interest and so, depending on the weighting scheme, the weighted average may increase the overall bias).

Here we have parameterized the Logistic Epanechnikov estimator so that the weight function has exactly the same *width*[2] as the Harrell-Davis estimator. The *mean ranks* for the Harrell-Davis and Logistic Epanechnikov estimators in this case are similar at 995.5 and 995.1 respectively. Thus the Harrell-Davis and Logistic Epanechnikov functions here have the same size and approximately the same location, but differ in their exact shape.

Figure 1: Example weights using different L-statistics (99.5th quantile; 1,000 scenarios)



Further mathematical details of a number of weighting schemes – Harrell-Davis, Epanechnikov and Logistic Epanechnikov – are provided in Appendix A.

Extreme Value Theory

The above estimators can be considered to be *non-parametric*, in the sense that we don't make any particular assumption about the shape of the underlying distribution. This contrasts with parametric approaches where we assume a particular distribution (e.g. Gaussian), find the parameters that best fit the data, and then calculate the quantile(s) of interest. The problem with a parametric approach is of course that, in general, we don't know the shape of the distribution. In fitting a specific distribution in order to estimate a quantile, there is a danger that we fit the wrong distribution and therefore misestimate the quantile.

An alternative *semi-parametric* approach is to use Extreme Value Theory (McNeil, Frey and Embrechts 2005). Extreme Value Theory doesn't make any particularly strong assumptions about the distribution, but appeals to a Theorem that tells us the limiting shape of the distribution in the tail. Specifically, the distribution of losses *in excess of some (high) threshold*, can be approximated by a Generalised Pareto Distribution (GPD), which is parameterized by just two parameters.

We can use this result to estimate quantiles as follows:

1. Pick a ('high') threshold and calculate losses in excess of this threshold.
2. Fit the two parameters of the Generalised Pareto Distribution to these excess losses and use this to estimate the distribution of losses in the tail.
3. Calculate the quantile of interest by inverting the fitted distribution function.

[2] Here we define the width of the function by $\sqrt{\sum_{i=1}^{n} w_i\,(i-\bar{\imath})^2}$ where the mean rank $\bar{\imath}$ is defined $\bar{\imath} = \sum_{i=1}^{n} w_i\, i$.

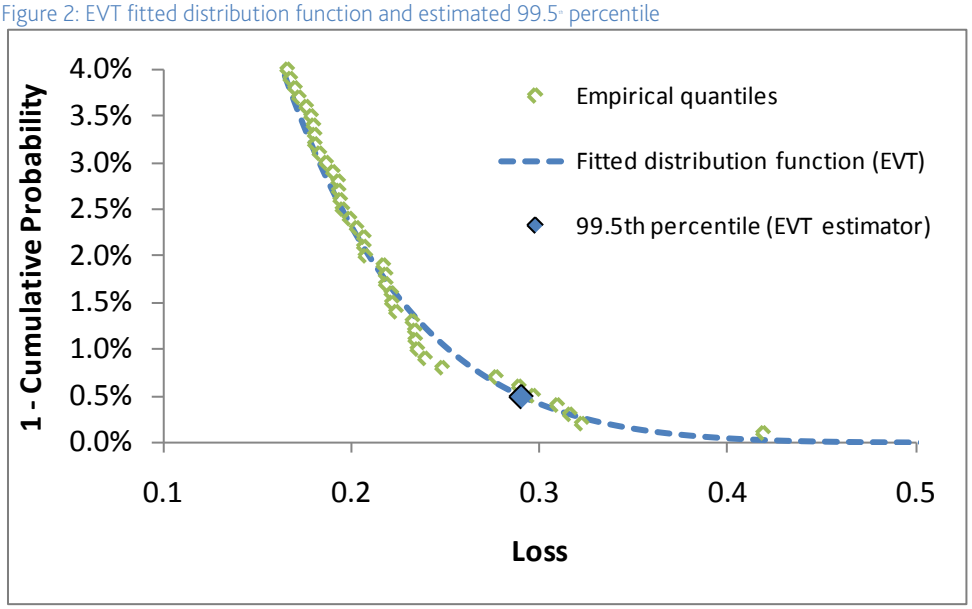
The result of this process is a quantile estimator of the form:

$$\hat{Q}_{EVT}(p) = u + \frac{\beta}{\xi}\left(\left(\frac{n(1-p)}{\sum_{i=1}^{n} 1_{\{X_i>u\}}}\right)^{-\xi} - 1\right)$$

where u is the chosen threshold (thus $\sum_{i=1}^{n} 1_{\{X_i>u\}}$ calculates the number of scenarios exceeding the threshold), and β, ξ are the two parameters of the Generelized Pareto Distribution, estimated by fitting the distribution to observations beyond the threshold. Further technical details of the approach are described in Appendix B.

Figure 2 shows an example fit to simulated losses, with the fitted 99.5th percentile highlighted by the blue diamond.

Figure 2: EVT fitted distribution function and estimated 99.5th percentile



Note that the fitted GPD allows us to estimate the entire distribution (beyond the chosen threshold) and not just the 99.5th percentile. Indeed a potential application of the fitted GPD is to extrapolate to more extreme percentiles where data is even scarcer.

The main drawback of the EVT estimator is that it relies on the assumption that the threshold is sufficiently large that the distribution is well described by a Generalised Pareto Distribution. The use of a theoretical result based on limiting assumptions (in this case the limit as the threshold is progressively increased) means that EVT estimators are likely to be harder to communicate, and gain trust in, than empirical (or L) estimators.

3. A 1-year VaR Economic Capital Case study

For illustration, we consider the estimation of the 99.5% 1-year VaR economic capital for a simple example insurance business. The liabilities of the business consist of a single policy type, with a payout linked to the performance of a corporate bond portfolio, with a guarantee applied annually. The main assumptions are summarized below:

» Assume an annual return of max (fund return – 1.5%, 2%) is credited to the policy account,

i.e. Policy Account(t) = Policy Account(t-1) x (1 + max (fund return(t-1,t) – 1.5%, 2%))

and Policy Account(0) = Fund Value(0)

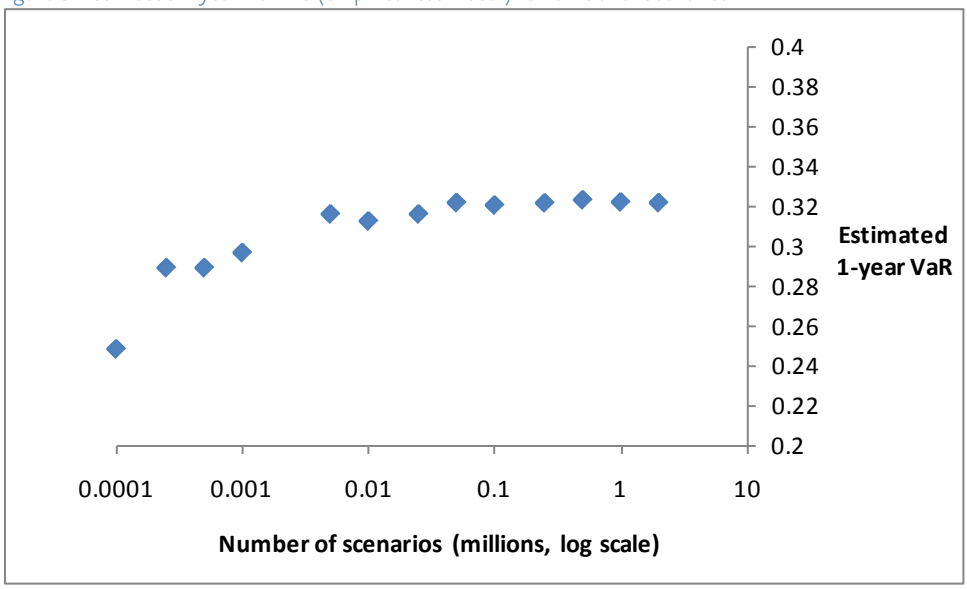
» The underlying fund is a diversified portfolio of US corporate bonds. The bonds are assumed to be invested with a credit mix of 70% A-rated and 30% BBB-rated, and with a term of 8 years. The bonds' credit rating and term are assumed to be re-balanced annually.

» The policyholder is assumed to exit the policy after ten years, and will receive the value of the policy account at that point.

In order to evaluate the liability value in each scenario, we have fitted a proxy function using the Least Squares Monte Carlo technique (Morrison, Turnbull and Vysniauskas 2013).

The economic capital requirements for this business cannot be calculated analytically, and must be estimated. Below we will consider estimating the 1-year VaR economic capital using different types of estimator and sample sizes. In order to quantify accuracy of each method, we have estimated the 'population' 1-year VaR EC using the empirical estimator with a very large number of scenarios. To decide on an appropriately large number of scenarios, we investigated convergence of the empirical quantile estimator, for varying numbers of scenarios, as shown in Figure 3. We observe that the estimated 1-year VaR EC appears to be relatively stable (on the scale shown) beyond around 500,000 scenarios.

Figure 3: Estimated 1-year VaR EC (empirical estimator) vs number of scenarios



Using 2 million scenarios[3], the resulting 1-year VaR EC estimate is 0.3221. To further check stability of this estimate, we repeated for a further 3 independent samples of 2 million scenarios and observed economic capital estimates in the range 0.3220 to 0.3223 (with an average of 0.3221 over all 4 independent samples). Given these results, we have assumed that the population 1-year VaR EC is 0.3221, and this is the benchmark to which we will compare different estimators below.

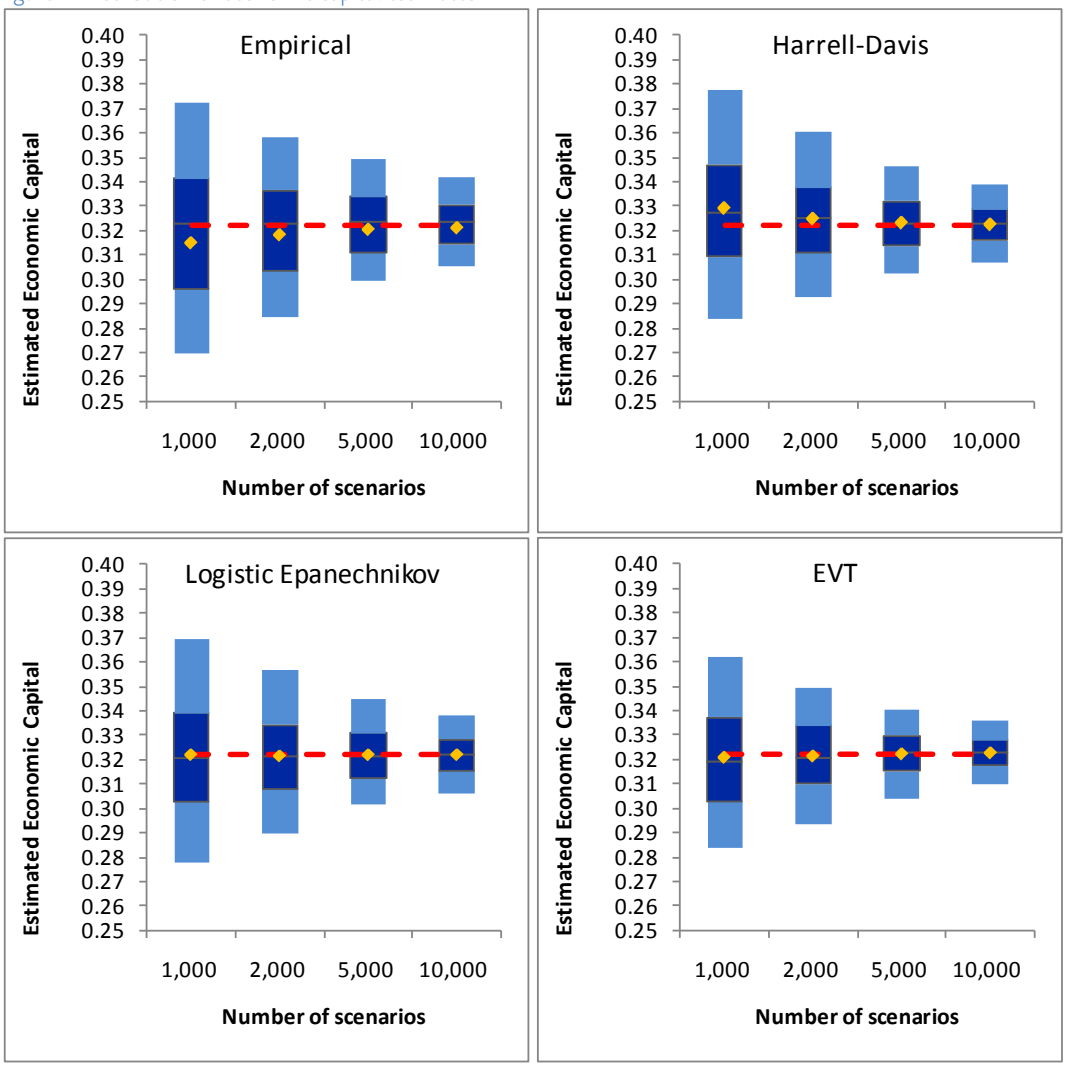
[3] Note that in this case the empirical estimator corresponds to the 1,990,000'th order statistic i.e. 10,000 scenarios have losses that exceed the empirical estimator.

We now estimate the 1-year VaR EC using far smaller sample sizes, ranging from 1,000 to 10,000 scenarios, using four different estimators:

1. Empirical estimator
2. L-estimator with Harrell-Davis weights
3. L-estimator with Logistic Epanechnikov weights[4]
4. Extreme Value Theory (EVT) estimator

In order to quantify the statistical properties of each estimator, in each case we have repeated the estimation 800 times. Figure 4 shows the mean of each estimator (shown by the orange symbols), 5th, 25th, 50th, 75th and 95th quantiles, and compares with the population economic capital (red dashed line). For example, using 1,000 scenarios, on average the empirical estimator is 0.3153 (and therefore biased low compared to the population economic capital of 0.3221), while the inter-quartile range is (0.2961, 0.3327).

Figure 4: Distribution of economic capital estimates

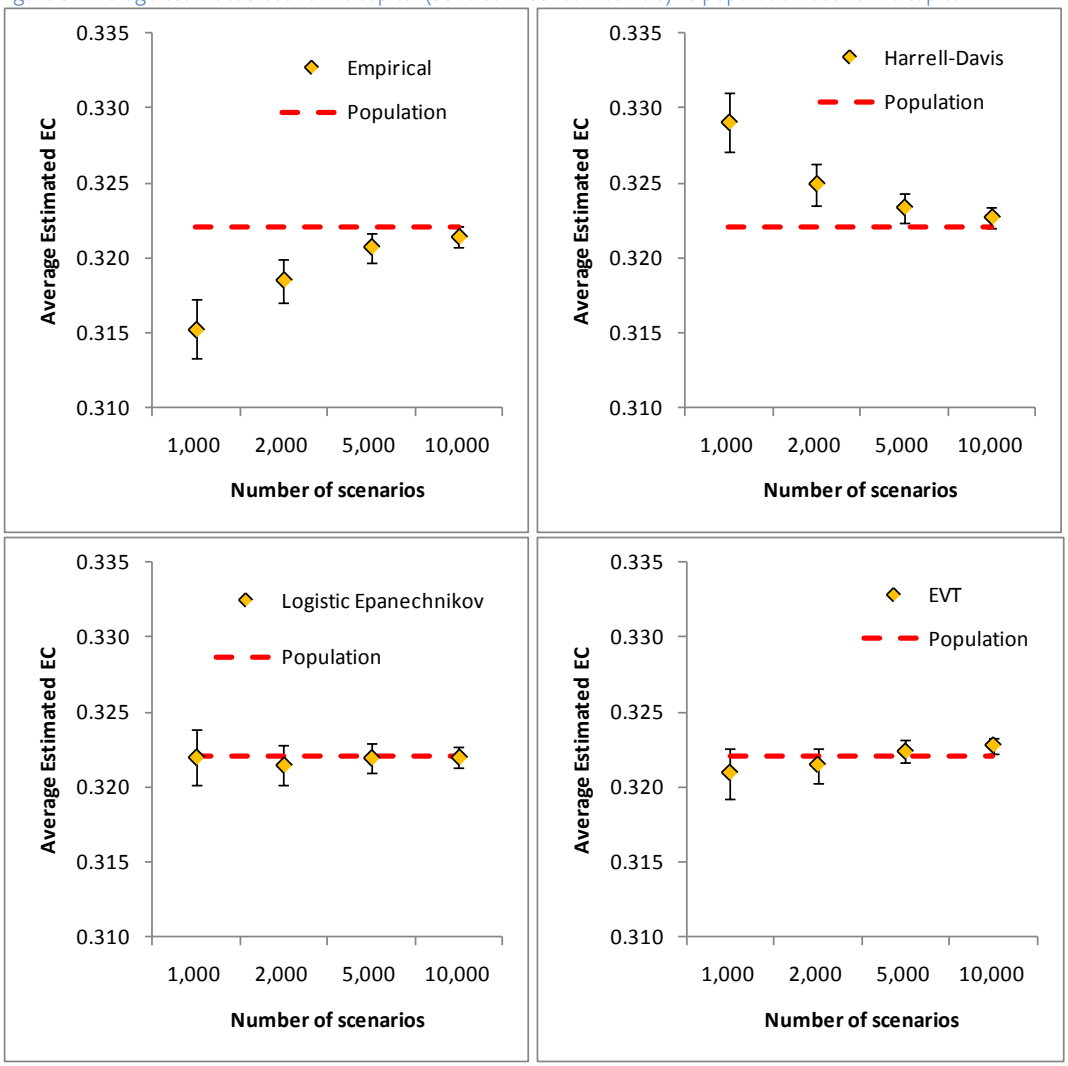


[4] In all cases, we have chosen the Logistic Epanechnikov bandwidths so that the weight functions have the same width as the corresponding Harrell-Davis estimators.

Of course, the means and quantiles shown above are themselves estimates based on 800 underlying economic capital estimates, and so subject to statistical error. To quantify the statistical uncertainty in the mean, we appeal to the Central Limit Theorem and estimate confidence intervals under the assumption that the mean is approximately normally distributed.

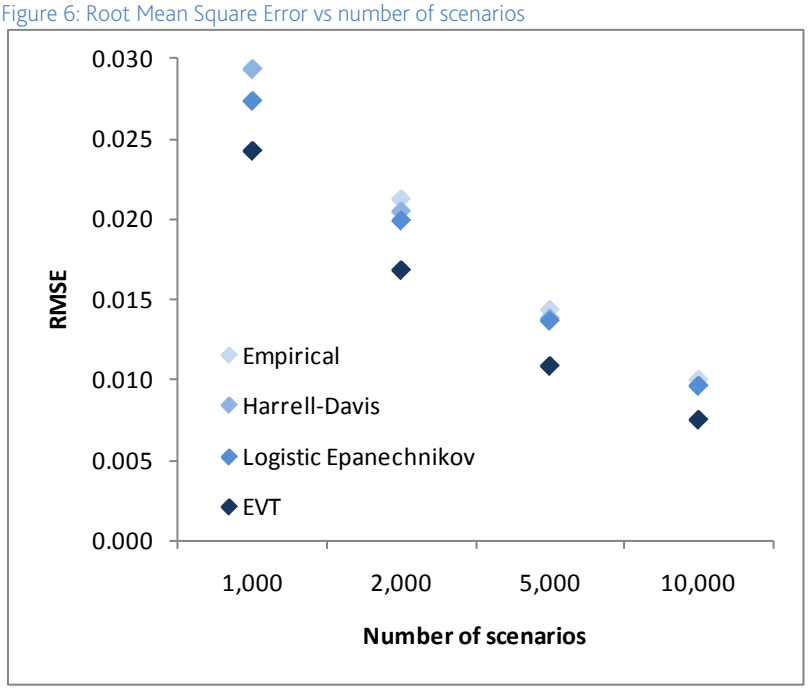
Figure 5 shows estimated 95% confidence intervals for the mean economic capital. This indicates that, for 2,000 scenarios or less, the empirical economic capital estimator is biased low (i.e. underestimated the economic capital on average), while the Harrell-Davis economic capital estimator is biased high (i.e. overestimates the economic capital on average). In contrast, any bias in the Logistic Epanechnikov and Extreme Value Theory economic capital estimators is relatively small, even at low sample sizes (1,000 scenarios). It is worth noting that even the largest bias observed here is small compared to statistical noise. For example, using 1,000 scenarios, the median empirical estimate is close to the population estimate i.e. approximately half of the estimates are larger than the population despite the fact that the estimator is biased low.

Figure 5: Average estimated economic capital (95% confidence intervals) vs population economic capital



To summarise the overall effectiveness of the different quantile estimates, we can use Root Mean Square Error (RMSE)[5]. Figure 6 compares RMSEs for all estimators and sample sizes considered. Using 1,000 scenarios, the Empirical and Harrell-Davis estimators show the highest RMSEs, due to the bias already noted above. RMSEs for the empirical, Harrell-Davis and Logistic Epanechnikov estimators converge as we increase the number of scenarios to 10,000, as all estimators are approximately unbiased but have similar variance. For all sample sizes considered, the EVT estimator has the lowest variance and RMSE.

Figure 6: Root Mean Square Error vs number of scenarios



The 'best' performing estimator here (in the lowest RMSE sense), is the EVT estimator. Using this estimator, with 10,000 scenarios, the inter-quartile range is (0.3172, 0.3277) compared to the population 1-year VaR EC of 0.3221. In this best case, given a random sample of 10,000 scenarios, there is a 25% chance of overestimating the 1-year VaR EC by 1.5% or more and a 25% chance of underestimating the 1-year VaR EC by 1.5% or more.

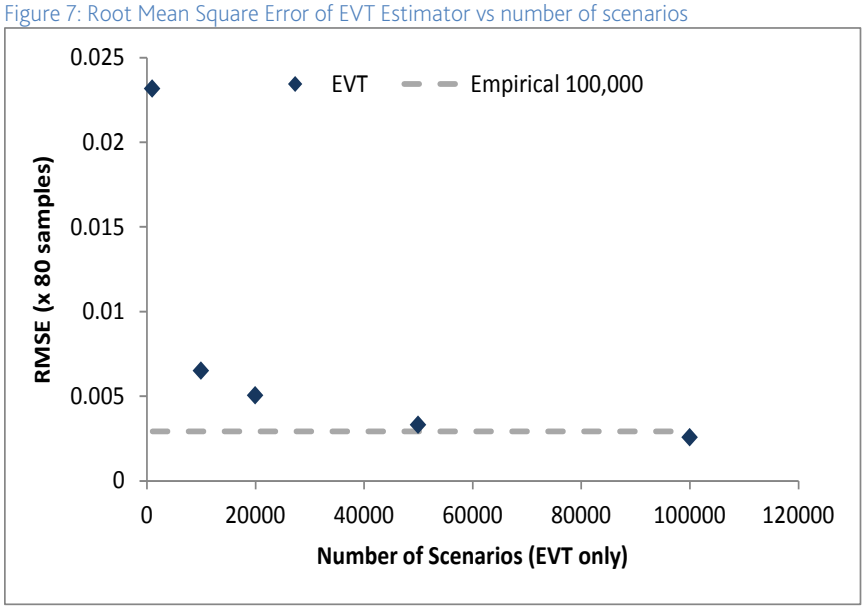
In order to further quantify the relative outperformance (as measured by RMSE) of the EVT estimation technique over the standard empirical estimator the following investigation was performed:

- The total scenario budget was held fixed at 8 million trials
- The empirical estimator was evaluated across 80 subsamples each containing 100,000 trials. This number is typical of industry practice.
- The RMSE was evaluated (0.002915)
- The number of subsamples was then held fixed at 80 to ensure reliable comparison of statistical metrics and EVT was then used to estimate the SCR for a number of scenario sets {1000, 10000, 20000, 50000, 100000}

The results are presented in Figure 7.

[5] The Root Mean Square Error is defined $RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{VaR_i} - VaR_{population}\right)}$ where n is the total number of estimates ($n = 800$ in this case), $\widehat{VaR_i}$ is the i'th estimate, and $VaR_{population}$ is the population VaR.

Figure 7: Root Mean Square Error of EVT Estimator vs number of scenarios



This result confirms, qualitatively, what we first observe in Figure 6: *the EVT estimator can obtain the same RMSE as the empirical method with half the number of scenarios* (i.e. 50,000 instead of 100,000).

4. Summary and conclusions

Estimation of 1-year VaR economic capital requirements using simulation involves the statistical estimation of quantiles. In this paper we have outlined some alternative approaches to quantile estimation, and used these to estimate the economic capital requirements of an example insurance business under different numbers of scenarios.

In the example considered here, Extreme Value Theory (EVT) performs most effectively amongst the quantile estimation methods considered in the paper. Specifically, the EVT implementation halved the number of 1-year scenarios required to obtain a given target level of estimation accuracy relative to the 'brute force' empirical method, i.e. 50,000 simulations with EVT to obtain the same level of sampling error as is obtained by using 100,000 with the empirical method.

Note that the relative performance of difference quantile estimators is likely to be sensitive to the underlying distribution of net assets and one should be careful in drawing general conclusions from the results presented here for a particular example insurance business. Nonetheless, we hope that the paper helps raise awareness of statistical error in 1-year VaR estimates, and some of the alternative estimation techniques that can be used to obtain more efficient estimation of economic capital.

Appendix A: L-statistics weights

Recall that the empirical quantile estimator is defined as the single order statistic:

$$\hat{Q}_{Empirical}(p) = X_{(\lceil np \rceil)}$$

where $\lceil np \rceil$ is the smallest integer not less than np. $\hat{Q}_{Empirical}(p)$ is a step function.

L-statistics generalise this by kernel smoothing over the empirical quantile function:

$$\hat{Q}_L(p) = \int_{t=0}^{1} X_{(\lceil nt \rceil)} k(t,p) dt$$

where $k(t,p)$ is a kernel function.

Since the empirical quantile is a step function this integral can be written as a sum:

$$\hat{Q}_L(p) = \sum_{i=1}^{n} \left(\int_{t=\frac{i-1}{n}}^{\frac{i}{n}} k(t,p) dt \right) X_{(i)}$$

which can be written as a weighted average of order statistics:

$$\hat{Q}_L(p) = \sum_{i=1}^{n} w_i X_{(i)}$$

with weights:

$$w_i = \int_{t=\frac{i-1}{n}}^{\frac{i}{n}} k(t,p) dt = K\left(\frac{i}{n}, p\right) - K\left(\frac{i-1}{n}, p\right)$$

where:

$$K(x,p) = \int_{t=-\infty}^{x} k(t,p) dt$$

Specific choices of kernel and resulting weights are described below.

1. Harrell-Davis

The Harrell-Davis kernel is the probability density function of the beta distribution with shape parameters $(n+1)p$ and $(n+1)(1-p)$:

$$k(t,p) = \frac{\Gamma(n+1)}{\Gamma\big((n+1)p\big)\Gamma\big((n+1)q\big)} t^{(n+1)p-1}(1-t)^{(n+1)q-1}, t \in [0,1]$$

where $q = 1 - p$ and Γ is the gamma function.

This gives rise to weights $w_i = K\left(\frac{i}{n}, p\right) - K\left(\frac{i-1}{n}, p\right)$ with:

$$K(x,p) = \frac{\Gamma(n+1)}{\Gamma\big((n+1)p\big)\Gamma\big((n+1)q\big)} \int_{t=0}^{x} t^{(n+1)p-1}(1-t)^{(n+1)q-1} dt$$

2. Epanechnikov

The Epanechnikov kernel is defined:

$$k(t,p) = \begin{cases} \frac{3}{4h}\left(1 - \left(\frac{t-p}{h}\right)^2\right) & |t-p| \le h \\ 0 & otherwise \end{cases}$$

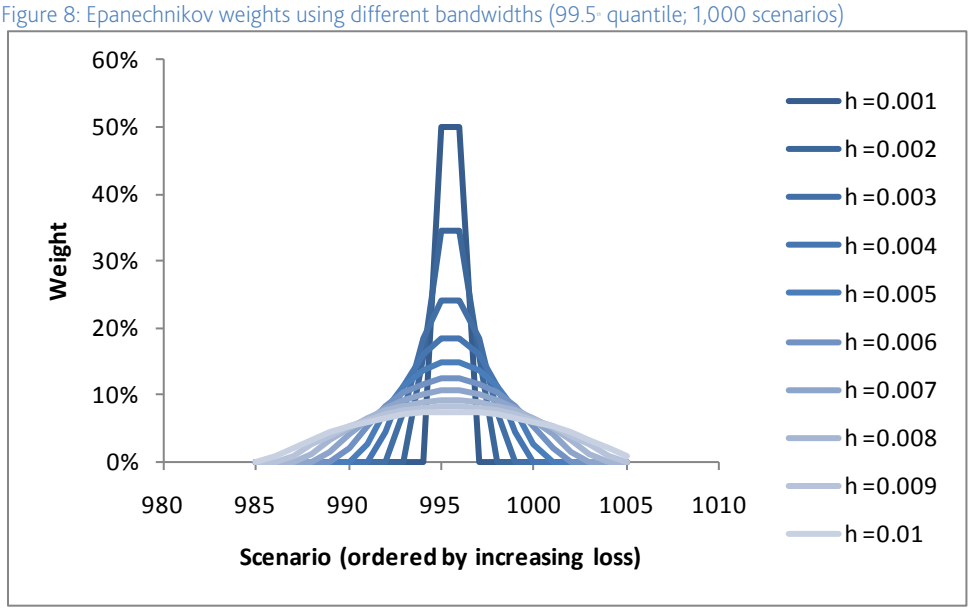
where h is a 'bandwidth' parameter.

This gives rise to weights $w_i = K\left(\frac{i}{n}, p\right) - K\left(\frac{i-1}{n}, p\right)$ with:

$$K(x,p) = \begin{cases} 0 & x \le p-h \\ \frac{1}{2} + \frac{3}{4}\left(\frac{x-p}{h}\right) - \frac{1}{4}\left(\frac{x-p}{h}\right)^3 & p-h \le x \le p+h \\ 1 & x \ge p+h \end{cases}$$

The bandwidth h controls the width of the weight function, as demonstrated in Figure 7.

Figure 8: Epanechnikov weights using different bandwidths (99.5· quantile; 1,000 scenarios)



Note that Epanechnikov kernel is defined on the range $(-\infty, +\infty)$. As a result, for sufficiently large bandwidths, the weights defined above do not add to one over all available order statistics (note for example that the weight function in Figure 7 assigns positive weight to scenarios with order greater than the total number of scenarios). To get around this problem, we can change variables before smoothing as described below.

3. Logistic Epanechnikov

Let f be a distribution function. We can then define an estimator as:

$$\hat{Q}_{kernel}(p) = \int_{z=-\infty}^{\infty} X_{(\lceil nf(z) \rceil)} k\left(z, f^{-1}(p)\right) dz$$

This leads to weights:

$$w_i = \int_{z=f^{-1}\left(\frac{i-1}{n}\right)}^{f^{-1}\left(\frac{i}{n}\right)} k\left(z, f^{-1}(p)\right) dz = K\left(f^{-1}\left(\frac{i}{n}\right), f^{-1}(p)\right) - K\left(f^{-1}\left(\frac{i-1}{n}\right), f^{-1}(p)\right)$$

The logistic Epanechnikov weights (Institute and faculty of Actuaries Extreme Events Working Party 2013) are defined as a particular example with k being the Epanechnikov kernel and $f(z) = \frac{1}{1+e^{-z}}$ (the distribution function of the logistic distribution).

Appendix B: Extreme Value Theory and the Generalised Pareto Distribution

The Generalised Pareto Distribution (GPD) is a two parameter distribution with cumulative distribution function:

$$G_{\xi,\beta}(x) = \begin{cases} 1 - \left(1 + \xi \frac{x}{\beta}\right)^{-1/\xi} & \xi \neq 0 \\ 1 - e^{-\frac{x}{\beta}} & \xi = 0 \end{cases}$$

Let $F(x)$ denote the cumulative distribution function for a random variable X. We define the excess distribution over some (high) threshold u is:

$$F_u(x) = P(X - u | X > u) = \frac{F(x+u) - F(u)}{1 - F(u)}$$

The idea of Extreme Value Theory is that, if we choose a large enough threshold u, we can approximate the distribution function $F_u(x)$ by a GPD $G_{\xi,\beta}(x)$ for some choice of the parameters ξ, β.

Having fitted the GPD $G_{\xi,\beta}(x)$, we can estimate the cumulative distribution function $F(x)$ for $x \geq u$ using:

$$\frac{\hat{F}_{EVT}(x) - \hat{F}(x)}{1 - \hat{F}(x)} = G_{\xi,\beta}(x - u)$$

Hence:

$$\frac{1 - \hat{F}_{EVT}(x)}{1 - \hat{F}(u)} \cong 1 - G_{\xi,\beta}(x - u) = \left(1 + \xi \frac{(x-u)}{\beta}\right)^{-1/\xi}$$

$1 - \hat{F}(u)$ can be estimated as the proportion of observations exceeding the threshold, and hence an estimate of the cumulative distribution (for $x \geq u$) is:

$$\hat{F}_{EVT}(x) = 1 - \frac{\sum_{i=1}^{n} 1_{\{X_i > u\}}}{n} \times \left(1 + \xi \frac{(x-u)}{\beta}\right)^{-1/\xi}$$

Finally, we can invert to estimate quantiles:

$$\hat{Q}_{EVT}(p) = u + \frac{\beta}{\xi}\left(\left(\frac{n(1-p)}{\sum_{i=1}^{n} 1_{\{X_i > u\}}}\right)^{-\xi} - 1\right)$$

The use of Extreme Value Theory to estimate quantiles relies on as assumption that the excess distribution beyond the chosen threshold is accurately described by a Generalised Pareto Distribution. We know that this result holds *asymptotically* i.e. as we let the threshold tend to infinity) for a very broad class of distributions. However it should be stressed that, for any finite choice of threshold, this assumption is an approximation. Guidance for choosing appropriate thresholds is described in (McNeil, Frey and Embrechts 2005).

References

Institute and faculty of Actuaries Extreme Events Working Party. "Difficult Risks and Capital Models." 2013.
Mausser, Helmut. "Calculating Quantile-based Risk Analytics with L-estimators." *Algo Research Quarterly* (Algo), 2001.
McNeil, A J, R Frey, and P Embrechts. *Quantitative Risk Management.* Princeton University Press, 2005.
Morrison, Steven, Craig Turnbull, and Naglis Vysniauskas. *Multi-year projection of 1-yr VaR capital requirements and free surplus.* Moody's Analytics, 2013.
Sheather, S J, and J S Marron. "Kernel quantile estimators." *Journal of the American Statistical Association*, 1990: 85 (410): 410-416.