# Table of Contents

## Reading 61: Estimating Market Risk Measures: An Introduction and Overview

**After completing this reading you should be able to:**

- Estimate VaR using a historical simulation approach.
- Estimate VaR using a parametric approach for both normal and lognormal return distributions.
- Estimate the expected shortfall given P/L or return data.
- Define coherent risk measures.
- Estimate risk measures by estimating quantiles.
- Evaluate estimators of risk measures by estimating their standard errors.
- Interpret QQ plots to identify the characteristics of a distribution.

# Estimating VaR using a Historical Simulation Approach

One of the simplest approaches to estimating VaR involves historical simulation. The risk manager constructs a distribution of losses by subjecting the current portfolio to the actual changes in the key factors experienced over the last $t$ time periods. After that, the mark-to-market P/L amounts for each period are calculated and recorded in an ordered fashion.

Let's assume we have a list of 100 ordered P/L observations and we would like to determine the VaR at 95% confidence. This implies that we should have a 5% tail, and that there are 5 observations (= 5% × 100) in the tail. In this scenario, the 95% VaR would be the sixth largest P/L observation.

In general, if there are n ordered observations, and a confidence level cl%, the cl% VaR is given by the $[(1-\text{cl}\%)n+1]^{th}$ highest observation. This is the observation that separates the tail from the body of the distribution. For instance, if we have 1,000 observations and a confidence level of 95%, the 95% VaR is given by the $(1-0.95)1,000+1=51^{st}$ observation. There are 50 observations in the tail.

### Example: Computing VaR using the historical simulation approach

Over the last 300 trading days, the five worst daily losses (in millions) were: -30, -27, -23, -21, and -19. If the historical window is these 300 daily P&L observations, what is the 99% daily HS VaR?

### Solution

Since VaR is to be estimated at 99% confidence, this means that 1% (i.e., 3) of the ordered observations would fall in the tail of the distribution.

The 99% VaR would be given by the (1 – 0.99) 300 + 1 highest observation, i.e., the $4^{th}$ highest value. This would be -21.

Note that the $4^{th}$ highest observation would separate the 1% of the largest losses from the remaining 99% of returns.

## Estimating Parametric VaR

When estimating VaR using the historical simulation approach, we do not make any assumption regarding the distribution of returns. In contrast, the parametric approach explicitly assumes a distribution for the underlying observations. We shall be looking at (1) VaR for returns that follow the normal distribution and (II) VaR for returns that follow the lognormal distribution.

### I. Normal VaR

#### a. Profit/Loss data

We have already established that the VaR for a given confidence level indicates the point that separates tail losses from the body of the distribution. Using P/L data, our VaR is:

$$\text{VaR}(\alpha\%) = -\mu_{PL} + \sigma_{PL} \times z_{\alpha}$$

where $\mu_{PL}$and $\sigma_{PL}$ are the mean and standard deviation of P/L, and $z_{\alpha}$ is the standard normal variate corresponding to the chosen confidence level. If we take the confidence level to be cl, then $z_{\alpha}$ is the standard normal variate such that 1 – cl of the probability density function lies to the left and cl of the probability density lies to the right. In most cases, you will be required to calculate the VaR when cl = 95%, in which case the standard normal variate is -1.645.

Notably, the VaR cutoff will be on the left side of the distribution. As such, the VaR is usually negative but is typically reported as positive since it is the value that is at risk (the negative amount is implied).

## Example: Computing VaR (Normal distribution)

Let P/L for ABC limited over a specified period be normally distributed with a mean of $12 million and a standard deviation of $24 million. Calculate the 95% VaR and the corresponding 99% VaR.

## Solution

$$\begin{aligned} 95\%\ \text{VaR}: \quad & \alpha = 95 \\ \text{VaR}(95\%) &= -\mu_{PL} + \sigma_{PL} \times z_{95} \\ &= -12 + 24 \times 1.645 = 27.48 \end{aligned}$$

How do we interpret this? ABC expects to lose at most $27.48 million over the next year with 95% confidence. Equivalently, ABC expects to lose more than $27.48 million with a 5% probability.

$$\begin{aligned} 99\%\ \text{VaR}: \quad & \alpha = 99 \\ \text{VaR}(99\%) &= -\mu_{PL} + \sigma_{PL} \times z_{99} \\ &= -12 + 24 \times 2.33 = 43.824 \end{aligned}$$

The 99% VaR can be interpreted in a similar fashion. However, note that the VaR at 99% confidence is significantly higher than the VaR at 95% confidence. Generally, the VaR increases as the confidence level increases.

b. **Arithmetic data**

When using arithmetic data rather than P/L data, VaR calculation follows a similar format.

Assuming the arithmetic returns follow a normal distribution,

$$r_t = \frac{p_t + D_t - p_{t-1}}{p_{t-1}}$$

Where (\text p_\text t): asset price at the end of periods; (\text D_{\text t}): interim payments

The VaR is:

$$\text{VaR}(\alpha\%) = [-\mu_r + \sigma_r \times z_\alpha] p_{(t-1)}$$

(\textbf{Example 1}: \textbf{Computing VaR given arithmetic data})

The arithmetic returns $r_t$, over some period of time, are normally distributed with a mean of 1.34 and a standard deviation 1.96. The portfolio is currently worth $1 million. Calculate the 95% VaR and 99% VaR.

## Solution

$$\text{VaR}(\alpha\%) = [-\mu_r + \sigma_r \times z_\alpha] p_{t-1}$$
$$95\%\text{VaR}: \quad \alpha = 95$$
$$\text{VaR}(95\%) = [-1.34 + 1.96 \times 1.645] \times \$1 = \$1.8842 \text{ million}$$
$$99\% \text{ VaR}: \quad \alpha = 99$$
$$\text{VaR}(99\%) \text{amp;} = [-1.34 + 1.96 \times 2.33] \times \$1 = \$3.2190 \text{ million}$$

Again, note that as the confidence level increases, so does the VaR

### Example 2: Computing VaR given arithmetic data

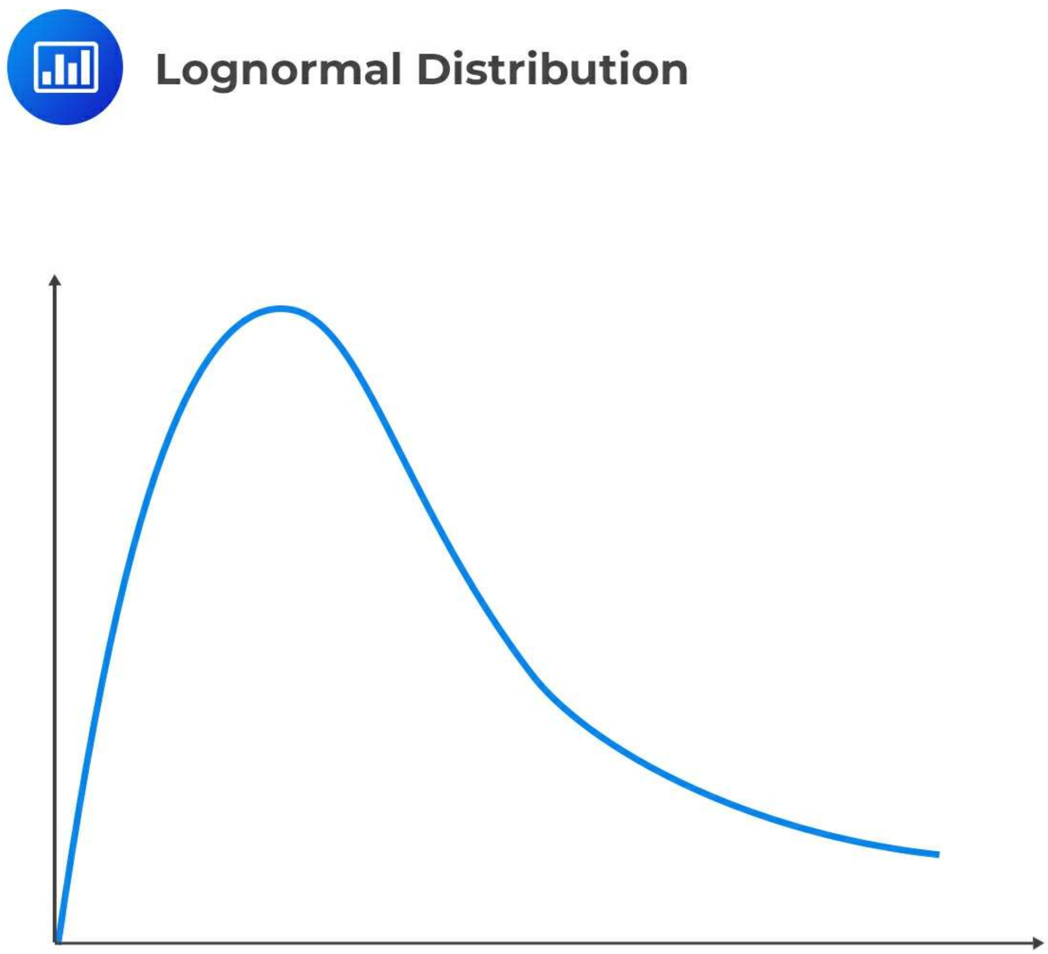
A portfolio has a beginning period value of $200. The arithmetic returns follow a normal distribution with a mean of 15% and a standard deviation of 20%. Determine the VaR at both the 95% and 99% confidence levels.

### Solution

$$\text{VaR}(\alpha\%) = [-\mu_r + \sigma_r \times z_\alpha]p_{t-1}$$
$$95\%\ \text{VaR}: \quad \alpha = 95$$
$$\text{VaR}(95\%) = [-0.15 + 0.2 \times 1.645] \times \$200 = \$35.8\ \text{million}$$
$$99\%\ \text{VaR}: \quad \alpha = 99$$
$$\text{VaR}(99\%) = [-0.15 + 0.2 \times 2.33] \times \$200 = \$63.2\ \text{million}$$

## II. Lognormal VaR

Unlike the normal distribution, the lognormal distribution is bounded by zero and is also skewed to the right.




This explains why this is the favored distribution when modeling the prices of assets such as stocks (which can never be negative). We also ditch the arithmetic returns in favor of geometric ones. As earlier established, the geometric return is:

$$R_t = \ln\left[\frac{p_t + D_t}{p_{t-1}}\right]$$

If we assume that geometric returns follow a normal distribution $(\mu_R, \sigma_R)$, then the natural logarithm of asset prices follows a normal distribution and $p_t$ follows a lognormal distribution

It can be shown that:

$$VaR(\alpha\%) = (1 - e^{\mu_R - \sigma_R \times z_\alpha})p_{t-1}$$

### Example: Computing lognormal VaR

Suppose geometric returns over some period are distributed as normal with mean 0.1 and standard deviation 0.15, and we have a portfolio currently valued at $20 million. Calculate the VaR at both 95% and 99% confidence.

### Solution

$$\begin{aligned}
VaR(\alpha\%)amp; &= (1 - e^{\mu_R - \sigma_R \times z_\alpha})p_{(t-1)} \\
95\%\ VaR: &\quad \alpha = 95 \\
VaR(95\%) &= (1 - e^{0.1 - 0.15 \times 1.645})20 = \$2.7298 \text{ million} \\
99\%\ VaR: &\quad \alpha = 99 \\
VaR(99\%) &= (1 - e^{0.1 - 0.15 \times 2.33})20 = \$4.4162 \text{ million}
\end{aligned}$$

### Example 2: Computing lognormal VaR

Let's assume that the geometric returns Rt, are distributed as normal with a mean 0.06 and standard a deviation 0.30. The portfolio is currently worth $1 million. Calculate the 95% and 99% lognormal VaR.

### Solution

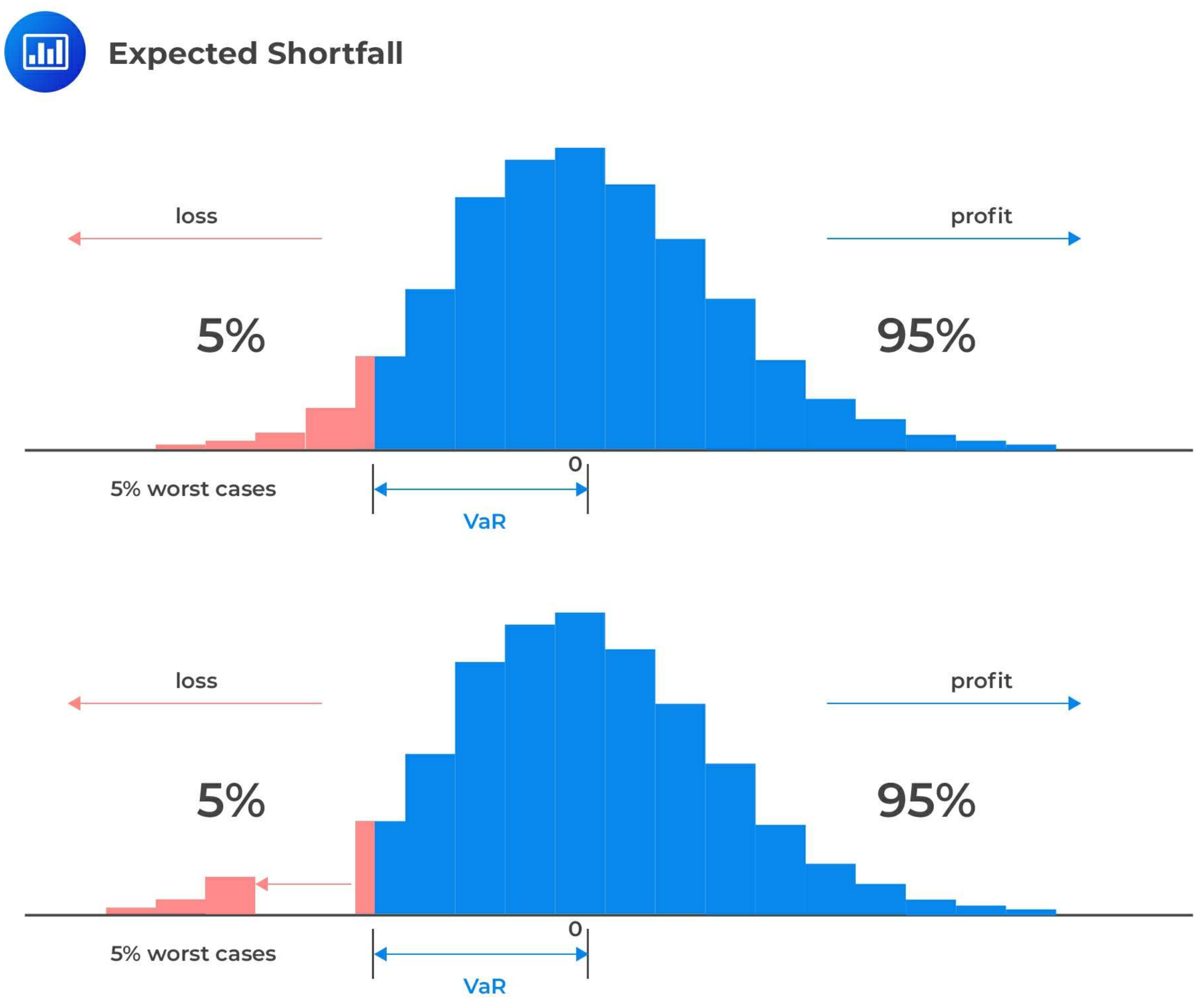
$$\begin{aligned}
VaR(\alpha\%) &= (1 - e^{\mu_R - \sigma_R \times z_\alpha})p_{(t-1)} \\
95\%\ VaR: &\quad \alpha = 95 \\
VaR(95\%) &= (1 - e^{0.06 - 0.30 \times 1.645})1 = \$0.3518 \text{ million} \\
99\%\ VaR: &\quad \alpha = 99 \\
VaR(99\%) &= (1 - e^{0.06 - 0.30 \times 2.33})1 = \$0.4689 \text{ million}
\end{aligned}$$

## Estimating the Expected Shortfall Given P/L or Return Data

Despite the significant role VaR plays in risk management, it stops short of telling us the amount or magnitude of the actual loss. What it tells us is the maximum value we stand to lose for a given confidence level. If the 95% VaR is, say, $2 million, we would expect to lose not more than $2 million with 95% confidence but we do not know what amount the actual loss would be. To have an idea of the magnitude of expected loss, we need to compute the expected shortfall.

**Expected shortfall** (ES) is the expected loss given that the portfolio return already lies below the pre-specified worst-case quantile return, e.g., below the 5th percentile return. Put different, expected shortfall is the mean percent loss among the returns found below the q-quantile (q is usually 5%). It helps answer the question: If we experience a catastrophic event, what is the expected loss in our financial position?

The expected shortfall (ES) provides an estimate of the tail loss by averaging the VaRs for increasing confidence levels in the tail. It is also called the expected tail loss (ETL) or the conditional VaR.



To determine the ES, the tail mass is divided into n equal slices and the corresponding n - 1 VaRs are computed.

## Illustration

Suppose that we wish to estimate a 95% ES on the assumption that losses are normally distributed with mean 0 and standard deviation 1. Ideally, we would need a large value of n to improve accuracy and reliability and then use the appropriate computer software. For illustration purposes, however, let's assume that n = 10.

This value gives us 9 tail VaRs (i.e., 10 – 1) at confidence levels in excess of 95%. These vaRs are listed below. The estimated ES is the average of these VaRs.

| Confidence level | Tail VaR |
|---|---|
| 95.5% | 1.6954 |
| 96.0% | 1.7507 |
| 96.5% | 1.8119 |
| 97.0% | 1.8808 |
| 97.5% | 1.9600 |
| 98.0% | 2.0537 |
| 98.5% | 2.1701 |
| 99.0% | 2.3263 |
| 99.5% | 2.5738 |
| | Average of tail VaRs = ES = 2.0250 |

The theoretical true value of the ES can be worked out by using different values of n. This can be easily achieved with the help of computer software.

| Number of tail slices (n) | ES |
|---|---|
| 10 | 2.0250 |
| 25 | 2.0433 |
| 50 | 2.0513 |
| 100 | 2.0562 |
| 250 | 2.0597 |
| 500 | 2.0610 |
| 1,000 | 2.0618 |
| 2,500 | 2.0623 |
| 5,000 | 2.0625 |
| 10,000 | 2.0626 |
| | Theoretical average = 2.063 |

These results show that the estimated ES rises with n and gradually converges to the theoretical true value of 2.063.

### Example: Computing the expected shortfall given P/L data

A market risk manager uses historical information on 300 days of profit/loss information and calculates a VaR, at the 95th percentile, of CAD30 million. Loss observations beyond the 95th percentile are then used to estimate the expected shortfall. These losses(in millions) are CAD23, 24, 26, 28, 30, 33, 37, 42, 46, and 47. What is the conditional VaR?

### Solution

Expected shortfall is the average of tail losses.

$$\begin{aligned} ES &= \left(\frac{23+24+26+28+30+33+37+42+46+47}{10}\right) \\ &= \text{CAD } 33.6 \text{ million} \end{aligned}$$

## Coherent Risk Measures

If X and Y are the future values of two risky positions, a risk measure $\rho(\bullet)$ is said to be coherent if it satisfies the following properties:

I. **Sub-additivity:** $\rho(X+Y) \le \rho(X) + \rho(Y)$

Interpretation: if we add two portfolios together with the total risk, the risk measure can't get any worse than adding the two risks separately.

II. **Homogeneity:** $\rho(XX) = X\rho(X)$

Interpretation: doubling a portfolio results in double the risk.

III. **Monotonicity:** $\rho(X) \ge \rho(Y),\ \text{if } X \le Y$

Interpretation: If one portfolio has better values than another under all scenarios then its risk will be better.

IV. **Translation invariance:** $\rho(X+n) = \rho(X) - n$

Interpretation: the addition of a sure amount n (cash) to our position will decrease our risk by the same amount because it will increase the value of our end-of-period portfolio.

VaR is not a coherent risk measure because it can be shown that it fails to satisfy the sub-additivity property. The expected shortfall (ES), however, does satisfy this property and is a coherent risk measure. If we combine two portfolios, the total ES would usually decrease to reflect the benefits of diversification. The ES would certainly never increase because it takes correlations into account. By contrast, the total VAR can – and in fact occasionally does – increase. As a result, the ES does not discourage risk diversification, but the VaR sometimes does.

The ES tells us what to expect in bad (i.e., tail) states—it gives an idea of how bad might be, whilst the VaR tells us nothing other than to expect a loss higher than the VaR itself.

The ES is also better justified than VAR in terms of decision theory.

Suppose you are faced with a choice between two portfolios A and B with different distributions. Under first – order stochastic dominance (FSD) which is a rather strict decision rule, A has first-order stochastic dominance over random variable B if, for any outcome x, A gives at least as high a probability of receiving at least x as does B, and for some x, A gives a higher probability of receiving at least x. And under second – order stochastic dominance (SSD), a more realistic decision rule, portfolio A would dominate B if it has a higher mean and lower risk. Using the ES as a risk measure is consistent with SSD, whereas VAR requires FSD, which is less realistic.

Overall, the ES dominates the VaR and presents a stronger case for use in risk management.

## Estimating Risk Measures by Estimating Quantiles

It is possible to estimate coherent risk measures by manipulating the "average VaR" method. A coherent risk measure is a weighted average of the quantiles (denoted by $q_p$ of the loss distribution):

$$M_{\emptyset} = \int_0^1 \emptyset(p)\, q_p dp$$

where the weighting function $\emptyset(p)$ is specified by the user, depending on their risk aversion. The ES gives all tail-loss quantiles an equal weight of [1/(1 – cl)] and other quantiles a weight of 0. Thus the ES is a special case of $M_{\emptyset}$.

Under the more general coherent risk measure, the entire distribution is divided into equal probability slices weighted by the more general risk aversion (weighting) function.

We could illustrate this procedure for n = 10. The first step is to divide the entire return distribution into nine (10 – 1) equal probability mass slices (loss quantiles) as shown below. Each breakpoint indicates a different quantile.

For example, the 10% quantile(confidence level = 10%) relates to -1.2816, the 30% quantile (confidence level = 30%) relates to -0.5244, the 50% quantile (confidence level = 50%) relates to 0.0,and the 90% quantile (confidence level = 90%) relates to 1.2816. After that, each quantile is weighted by the specific risk aversion function and then averaged to arrive at the value of the coherent risk measure.

| Confidence level | Normal deviate (A) | Weight (B) | A × B |
|---|---|---|---|
| 10% | −1.2816 | 0 | |
| 20% | −0.8416 | 0 | |
| 30% | −0.5244 | 0 | |
| 40% | −0.2533 | x | |
| 50% | 0.0 | xx | |
| 60% | 0.2533 | xxx | |
| 70% | 0.5244 | xxxx | |
| 80% | 0.8416 | xxxxx | |
| 90% | 1.2816 | xxxxxx | |
| | | | Average risk measure = sum of A × B |

The x's in the third column represent some weight that depends on the investor's risk aversion.

Compared to the expected shortfall, such a coherent risk measure is more sensitive to the choice of n. However, as n increases, the risk measure converges to its true value. Remember that increasing the value of n takes us farther into some very extreme values at the tail.

## Evaluating Estimators of Risk Measures by Estimating their Standard Errors

Crucially, bear in mind that any risk measure estimates are only as useful as their precision. The true value of any risk measure is unknown and thus it is important to come up with estimates in a precise manner. Why? Only then can we be confident that the true value is fairly close to the estimate. Hence, it is important to supplement risk measure estimates with some indicator that gauges their precision. The standard error is a useful indicator of precision. More generally, confidence intervals (built with the help of standard errors) can be used.

The big question is: how do we go about determining standard errors and establishing confidence intervals? Let's start with a sample size of n and arbitrary bin width of h around quantile, q. Bin width refers to the width of the intervals, or what we usually call "bins," in a (statistical) histogram. The square root of the variance of the quantile is equal to the standard error of the quantile. Once the standard error has been specified, a confidence interval for a risk measure can be constructed:

$$[q + se(q) \times z_\alpha] > VaR > [q - se(q) \times z_\alpha]$$

## Example: Computing the standard error for a risk measure

Construct a 90% confidence interval for 5% VaR (the 95% quantile) drawn from a standard normal distribution. Assume bin width = 0.1 and that the sample size is equal to 1,000.

## Solution

**Step 1:** determine the value of q

The quantile value, q, corresponds to the 5% VaR. For the normal distribution, the 5% VaR occurs at 1.645 (implying that q = 1.645). So in crude form, the confidence interval will take the following shape:

$$[q + se(q) \times z_\alpha] > VaR > [q - se(q) \times z_\alpha]$$

**Step 2:** determine the range of q

For the bin width of 0.1, we know that q falls in the bin spanning 1.645 ± 0.1/2 = [1.595,1.695].

Note: the left tail probability, p, is the area to the left of 1.695 for a standard normal distribution.

**Step 3:** determine the probability mass f(q)

We wish to calculate the probability mass between 1.595 and 1.695, represented as f(q). From the normal distribution table, the probability of a loss exceeding 1.695 is 4.5% (which is also equal to p) and the probability of profit or a loss less than 1.595 is 94.46%. Hence, f(q) = 1 – 0.045 – 0.9446 = 1.032%

**Step 4:** calculate the standard error of the quantile from the variance approximation of q.

$$se(q) = \frac{\sqrt{\frac{p(1-p)}{n}}}{f(q)}$$

In this case,

$$se(q) = \frac{\sqrt{\frac{0.045 \times 0.955}{1000}}}{0.01032} = 0.63523$$

Thus, the following gives us the required CI:

$$[1.645 + 0.63523 \times 1.645] > \text{VaR} > [1.645 - 0.63523 \times 1.645]$$
$$= 2.69 > \text{VaR} > 0.6$$

Important:

Unlike the VaR, confidence intervals are two-sided, so for a 90% CI, there will be 5% in each tail.This is equivalent to the 5% significance level of VaR, and therefore the critical values are ± 1.645.

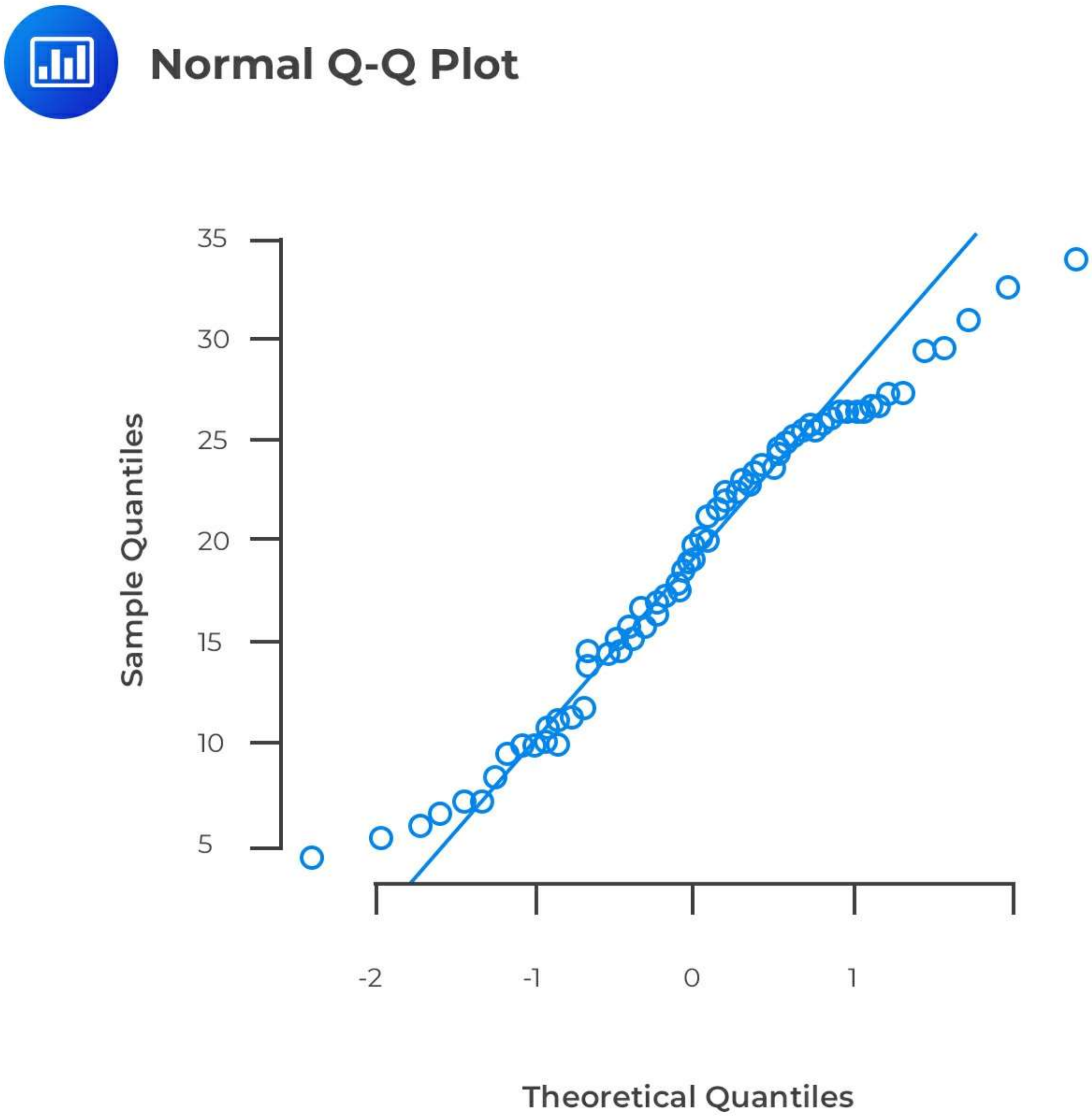
The larger the sample size the smaller the standard error and the narrower the confidence interval.

Increasing the bin size, h, holding all else constant, will increase the probability mass f(q) and reduce p, the probability in the left tail. Subsequently, the standard error will decrease and the confidence interval will again narrow.

Increasing p implies that tail probabilities are more likely. When that happens, the estimator becomes less precise, and standard errors increase, widening the confidence interval. Note that the expression p(1- p) will be maximized at p = 0.5.

## Interpreting QQ Plots to Identify the Characteristics of a Distribution

The quantile-quantile plot, more commonly called the Q-Q plot, is a graphical tool we can use to assess if a set of data plausibly came from some theoretical distribution such as a Normal or exponential.




For example, if we conduct risk analysis assuming that the underlying data is normally distributed, we can use a normal QQ plot to check whether that assumption is valid. We would need to plot the quantiles of our data set against the quantiles of the normal distribution. It's **not** a perfect air-tight check but a visual proof that can be quite subjective.

Remember that by a quantile, we mean the fraction (or percent) of points below the given value. For example, the 0.1 (or 10%) quantile is the point at which 10% percent of the data fall below and 90% fall above that value.

## Why are QQ plots important?

First, we can use a QQ plot to form a tentative view of the distribution from which our data might be drawn: we specify a variety of alternative distributions and construct QQ plots for each. If the data are drawn from the reference population, then the QQ plot should be linear. Any reference distributions that produce non-linear QQ plots can then be dismissed, and any distribution that produces a linear QQ plot is a good candidate distribution for our data.

Second, since a linear transformation in one of the distributions in a QQ plot changes the intercept and slope of the QQ plot we can use the intercept and slope of a linear QQ plot to give us a rough idea of the location and scale parameters of our sample data.

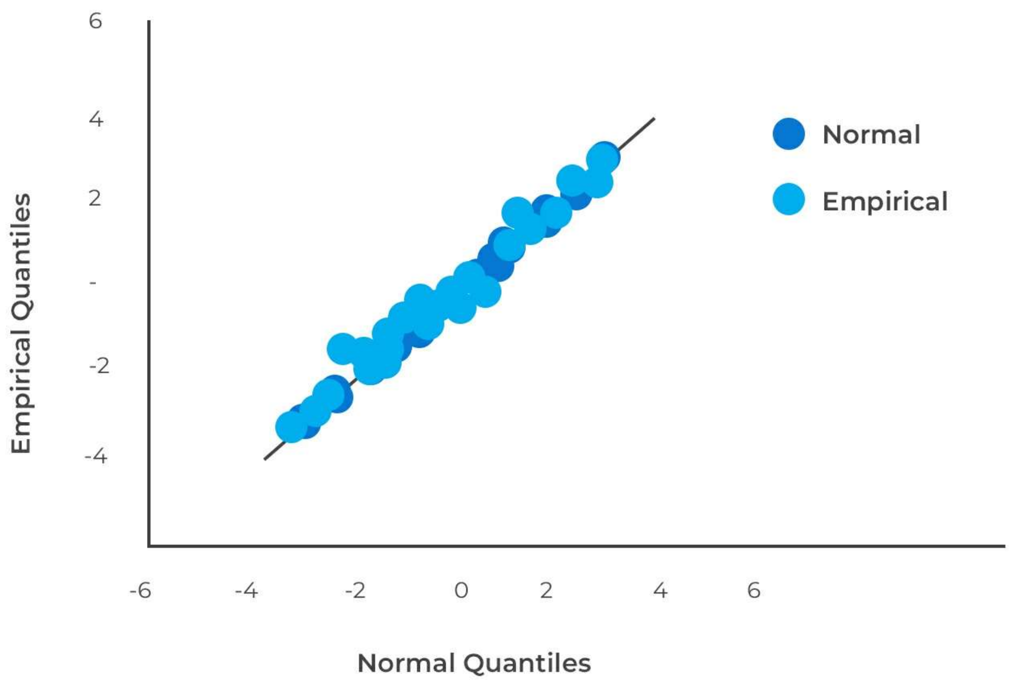
Third, if the empirical distribution has heavier tails than the reference distribution, the QQ plot will have steeper slopes at its tails, even if the central mass of the empirical observations are approximately linear.

Finally, a QQ plot is good for identifying outliers (e.g., observations contaminated by large errors).

## Illustration

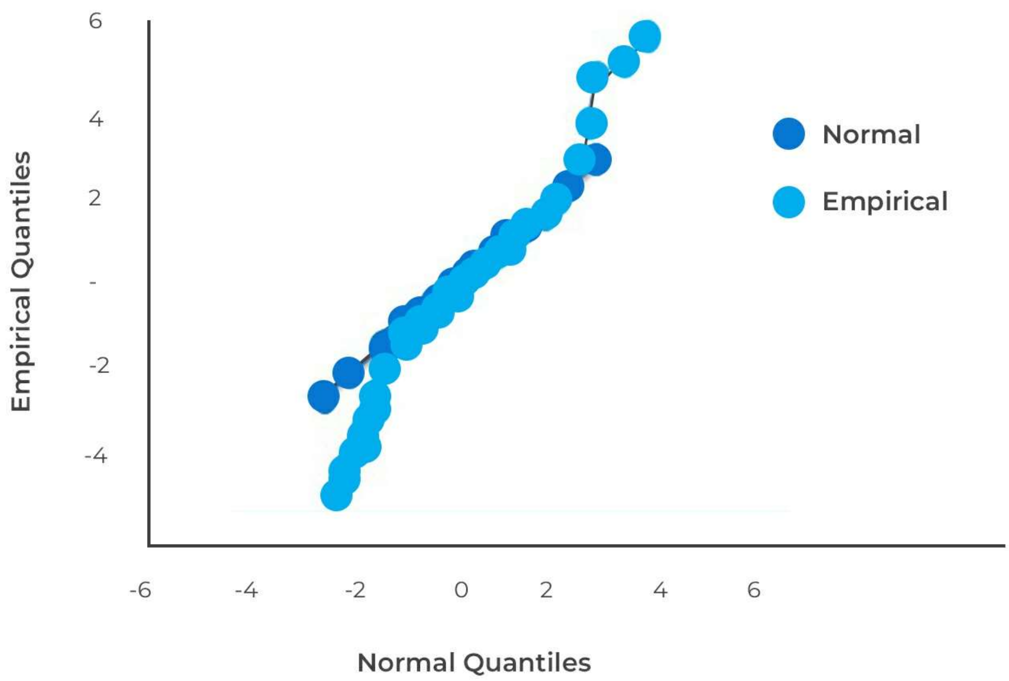
The chart here shows a QQ plot for a data sample drawn from a normal distribution, compared to a reference distribution that is also normal. The central mass observations fit a linear QQ plot very closely while the observations at the tail are a bit spread out. In this case, the empirical distribution matches the reference population.

## Q-Q Plot (Example 1)




The next one shows a QQ plot for a data sample drawn from a normal distribution, compared to a reference distribution that is not normal. Notably, it has a very heavy left tail and light right tail. In this case, the empirical distribution does not match the reference population.

## Q-Q Plot (Example 2)



## Practice Questions

## Question 1

If the arithmetic returns $r_t$ over some period of time are 0.075, what are the geometric returns?

A. -2.59
B. 0.0723
C. 0.0779
D. 0.0895

The correct answer is **B**.

Recall that, $R_t = \ln(1 + r_t)$

Therefore, $\ln(1 + .075) = \ln(1.075) = 0.0723$

## Question 2

Let the geometric returns $r_t$ be 0.020, over a specified period, calculate the arithmetic returns.

A. 0.1823
B. 0.0182
C. 0.0202
D. 0.2020

The correct answer is **C**.

Recall, $1 + r_t = \exp(R_t)$, then: $r_t = \exp(R_t) - 1$

Implying that: $\exp(0.020) – 1 = 0.0202$

## Question 3

Assuming that P/L over a specified period is normally distributed and has a mean of 13.9 and a standard deviation of 23.1. What is the 95% VaR?

A. 24.22
B. 25.10
C. 51.90
D. 59.18

The correct answer is **A**.

Recall that: $\alpha VaR = -\mu_{P/L} + \sigma_{P/L} z_{\alpha,}$

95% VaR is: $-13.9 + 23.1 Z_{0.95} = -13.9 + 23.1 \times 1.65 = 24.22$

## Question 4

The arithmetic returns $r_t$, over some period of time, are normally distributed with a mean of 1.89 and a standard deviation 0.98. The portfolio is currently worth $1. Calculate the 99% VaR.

A. 0.3895
B. 4.1695
C. 3.8108
D. 0.0308

The correct answer is **A**.

Recall, $\alpha VaR = -(\mu_r - \sigma_r z_\alpha) P_{t-1}$,

99% VaR : $-1.89 + 0.98 \times 2.326 = 0.3895$

## Question 5

Assuming we make the empirical assumption that the mean and volatility of annualized returns are 0.24 and 0.67. Assuming there are 250 trading days in the years, calculate the normal 95% VaR and lognormal 95% VaR at the 1-day holding period for a portfolio worth $1.

A. 95% VaR: 6.89%; lognormal 95% VaR: 6.11%.

B. 95% VaR: 6.89%; lognormal 95% VaR: 6.65%.

C. 95% VaR: 6.65%; lognormal 95% VaR: 6.11%.

D. 95% VaR: 6.65%; lognormal 95% VaR: 6.04%.

The correct answer is **B**.

The daily return has a mean $0.24/250 = 0.00096$ and standard deviation $0.67/\sqrt{250} = 0.0424$.

Then, the normal 95% VaR is $-0.00096 + 0.0424 \times 1.645 = 0.0689$.

Therefore, the normal VaR is 6.89%.

If we assume a lognormal, then the 95% VaR is $1 - \exp(0.00096 - 0.0424 \times 1.645) = 0.0665$.

This implies that the lognormal VaR is 6.65% of the value of the portfolio.

## Reading 62: Non-parametric Approaches

**After completing this reading you should be able to:**

- Apply the bootstrap historical simulation approach to estimate coherent risk measures.
- Describe historical simulation using non-parametric density estimation.
- Compare and contrast the age-weighted, the volatility-weighted, the correlation-weighted, and the filtered historical simulation approaches.
- Identify the advantages and disadvantages of nonparametric estimation methods.

# The Bootstrap Historical Simulation Approach to Estimating Coherent Risk Measures

Bootstrapping presents a simple but powerful improvement over basic Historical Simulation is to estimate VaR and ES. Crucially, it assumes that the distribution of returns will remain the same in the past and in the future, justifying the use of historical returns to forecast the VaR.

A bootstrap procedure involves resampling from our existing data set with replacement. A sample is drawn from the data set, its VaR recorded, and the data “returned.” This procedure is repeated over and over. The final VaR estimate from the full data set is taken to be the average of all sample VaRs. In fact, bootstrapped VaR estimates are often more accurate than a ‘raw’ sample estimates.

There are three key points to note regarding a basic bootstrap exercise:

1. We start with a given original sample of size n. We then draw a new random sample of the same size from this original sample, “returning” each chosen observation back in the sampling pool after it has been drawn.
2. Sampling with replacement implies that some observations get chosen more than once, and others don’t get chosen at all. In other words, a new sample, known as a resample, may contain multiple instances of a given observation or leave out the observation

completely, making the resample different from both the original sample and other resamples. From each resample, therefore, we get a different estimate of our parameter of interest.

3. The resampling process is repeated many times over, resulting in a set of resampled parameter estimates. In the end, the average of all the resample parameter estimates gives us the final bootstrap estimate of the parameter. The bootstrapped parameter estimates can also be used to estimate a confidence interval for our parameter of interest.

Equally as important is the possibility to extend the key tenets of bootstraps to the estimation of the expected shortfall. Each drawn sample will have its own ES. First, the tail region is sliced up into n slices and the VaR for each of the resulting n – 1 quantiles is determined. The final VaR estimate is taken to be the average of all the tail VaRs.We then estimate the ESas the average of losses in excess of the final VaR.

As in the case of the VaR, the best estimate of the expected shortfall given the original data set is the average of all of the sample expected shortfalls.

In general, this bootstrapping technique consistently provides more precise estimates of **coherent risk measures** than historical simulation on raw data alone.

## Bootstrapped Confidence Intervals

For a start, we know that thanks to the central limit theorem, the distribution of $\hat{\theta}$often approaches normality as the number of samples gets large. In these circumstances, it would be reasonable to estimate a confidence interval for $\theta$ assuming $\hat{\theta}$ is approximately normal.

Given that $\hat{\theta}$ is our estimate of $\theta$ and $\hat{\sigma}$ is the estimate of the standard error of $\hat{\theta}$, the confidence interval at 95% is:

$$[\hat{\theta} - 1.96\hat{\sigma}, \hat{\theta} + 1.96\hat{\sigma}]$$

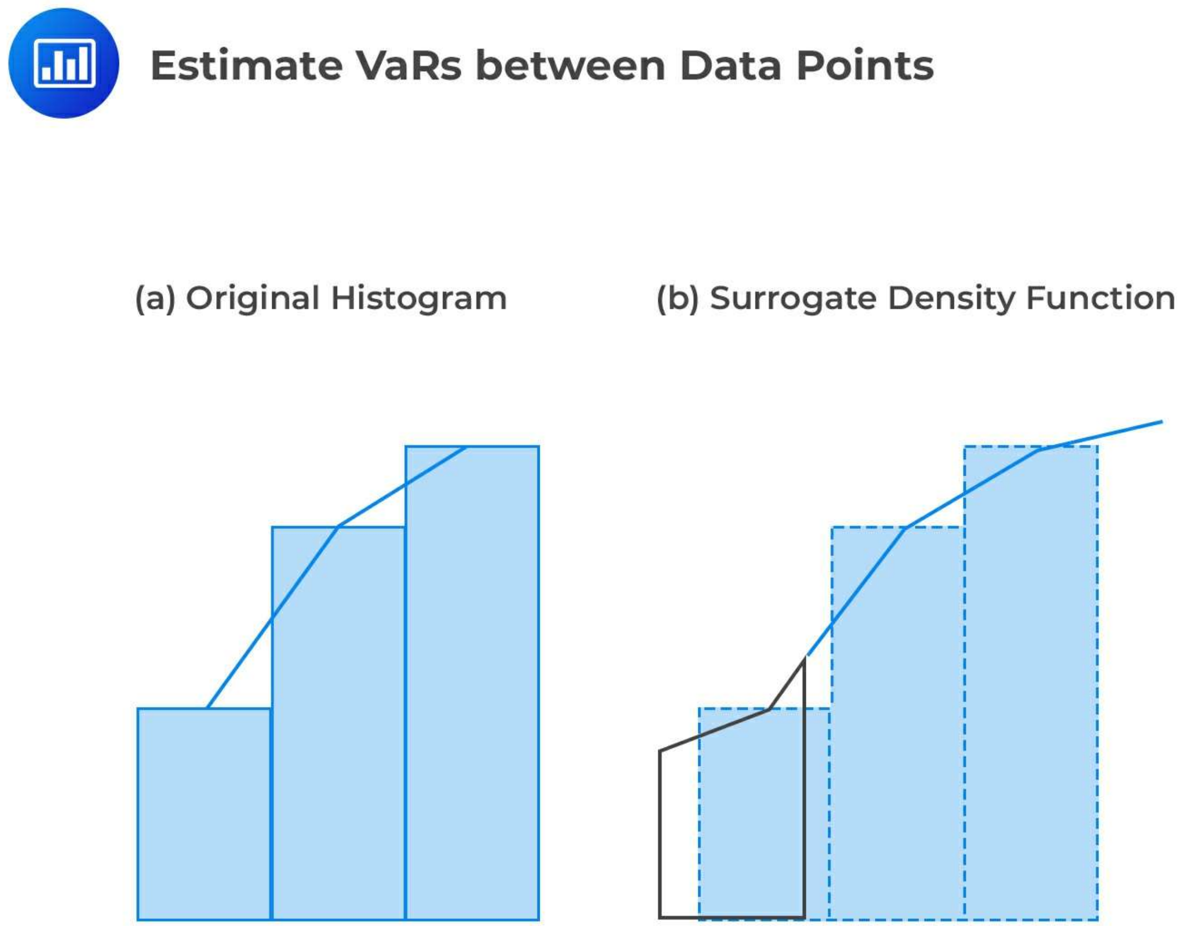
It is also possible to work out confidence intervals using percentiles of the sample distribution.

The upper and lower bounds of the confidence interval are given the percentile points (or quantiles)of the sample distribution of parameter estimates.

## Historical Simulation using Non-parametric Density Estimation

A huge selling point about the traditional historical approach has much to do with its simplicity. However, there's one major drawback: due to the discrete nature of the data, it is impossible to estimate VaRs between data points. For example, if there are 100 historical observations, it would be easy to estimate the VaR at 95% or even 99%. But what about the VaR at, say, 96.5%? It would be impossible to incorporate a level of confidence of 96.5%. The point here is that with n observations, the historical simulation method only allows for n different confidence levels. Luckily, Non-parametric density estimation offers a potential solution to this problem.

So what happens? We treat our data as drawings that are free from the "shackles" of some specified distribution. The idea is to make the data "speak for itself" without making any strong assumptions about its distribution. To enable us estimate VaRs and ESs for any confidence levels, we simply draw straight lines connecting the mid-points at the top of each histogram bar (in the original data set's distribution) and treat the area under the lines as if it were a pdf. By so doing, don't we lose part of the data? No: by connecting the midpoints, the lower bar "receives" some area from the upper bar, which "loses" or cedes an equal amount of area. In the end, no area is lost, only the displacement occurs. We still end up with a probability distribution function. The shaded area in the figure below represents a possible confidence interval that can be utilized regardless of the size of the data set.




## Weighted Historical Simulation Approaches

Recall that under the historical method of estimating VaR, all of the past n observations are weighted equally, where each observation has a weight of 1/n. In other words, our HS P/L series is constructed in a way that gives any observation n periods old or less the same weight in our VaR, and no weight (i.e., a zero weight) to all observations that come after that. While simple in construction, this weighting scheme has several flaws.

First, it seems hard to justify giving each observation the same weight without taking into account its age, market volatility at the time it was observed, or the value it takes. For instance, it's an open secret that gas prices are more volatile in the winter than in the summer, so if the sample period cuts across the two seasons of the year, the resulting VaR estimate will not reflect the true risk facing the firm. As a matter of fact, equal weights will tend to underestimate true risks in the winter and overestimate them in the summer.

Second, equal weights make the resulting risk estimates unresponsive to major events. For instance, we all know that risk increases significantly following a major destabilizing event such as a stock market crash or the start of a trade war involving one or more economies (US and China would be perfect examples). Unless a very high level of confidence is used, HS VaR estimates would not capture the increased risk following such events. The increase in risk would only reflect in subsequent dates if the market slide continued.

Third, equal weights suggest that each observation in the sample period is equally likely and independent of all the others. That is untrue because, in practice, periods of high or low volatility tend to be clustered together.

Fourth, an unusually high observation will tend to have a major influence on the VaR until n days have passed and the observation has fallen out of the sample period, at which point the VaR will fall again.

Finally, it would be difficult to justify a sudden shift of weight from 1/n on date n to zero on date n+1. In other words, it would be hard to explain why the observation on date n is important, but that on date n+1 is not.

This learning outcome looks at four improvements to the traditional historical simulation method.

## Age-weighted Historical Simulation (Hybrid HS)

Instead of equal weights, we could come up with a weighting structure that discounts the older observations in favor of newer ones.

Let the ratio of consecutive weights be constant at lambda ($\lambda$). If w(1) is the probability weight given to an observation that's 1 day old, then w(2), the probability given to an observation 2 days old, could be $\lambda$w(1); w(3), the probability weight given to an observation 3 days old, could be $\lambda^2$ w(1); w(4) could be $\lambda^3$ w(1,), w(5) could be $\lambda^4$ w(1), and so on. In such a case, lambda would be a term between 0 and 1 and would reflect the exponential rate of decay in the weight as time goes.

Aλ close to 1 signifies a slow rate of decay, and a λ far away from 1 signifies a high rate of decay.

Under age-weighted historical simulation, therefore, the weight given to an observation i days old is given by:

$$w(i) = \frac{\lambda^{i-1}(1-\lambda)}{(1-\lambda^{n})}$$

w(1) is set such that the sum of the weights is 1.

Example

$$\lambda = 0.96; n = 100$$

Initial date

listing only the worst 6 returns

| | | Simple | HS | Hybrid | (Exp) |
|---|---|---|---|---|---|
| Return | periods ago(i) | Weight | Cumul. | Weight | Cumul. |
| −3.50% | 6 | 1.00% | 1.00% | 3.32% | 3.32% |
| −3.20% | 4 | 1.00% | 2.00% | 3.60% | 6.92% |
| −2.90% | 55 | 1.00% | 3.00% | 0.45% | 7.37% |
| −2.70% | 35 | 1.00% | 4.00% | 1.02% | 8.39% |
| −2.60% | 8 | 1.00% | 5.00% | 3.06% | 11.45% |
| −2.40% | 24 | 1.00% | 6.00% | 1.60% | 13.05% |

$$\lambda = 0.96; n = 100$$

20 days later

Notice: Only the 6th worst return is recent, others are same

| | | Simple | HS | Hybrid | (Exp) |
|---|---|---|---|---|---|
| Return | periods ago(i) | Weight | Cumul. | Weight | Cumul. |
| −3.50% | 26 | 1.00% | 1.00% | 1.47% | 1.47% |
| −3.200% | 24 | 1.00% | 2.00% | 1.59% | 3.06% |
| −2.90% | 75 | 1.00% | 3.00% | 0.20% | 3.26% |
| −2.70% | 55 | 1.00% | 4.00% | 0.45% | 3.71% |
| −2.60% | 28 | 1.00% | 5.00% | 1.35% | 5.06% |
| −2.50% | 14 | 1.00% | 6.00% | 2.39% | 7.45% |

$$w(i) = \frac{\lambda^{i-1}(1-\lambda)}{(1-\lambda^n)} \qquad \text{e.g. } w(6) = \frac{0.96^{6-1}(1-0.96)}{(1-0.96^{100})} = 3.32\%$$

Advantages of the age-weighted HS method include:

- It generalizes standard historical simulation (HS) because “we can regard traditional HS as a special case with zero decay, where λ is essentially equal to 1.
- Choosing lambda appropriately will make VaR/ES estimates more responsive to large loss observations. A suitable choice of lambda will award a large loss event a higher weight than under traditional HS, making the resulting next day VaR higher than it would otherwise have been.
- It helps to reduce distortions caused by events that are unlikely to recur and helps to reduce ghost effects. An unusually large loss will have its weight gradually reduced as time goes until it is “kicked out” of the historical sample size.
- Age-weighting can be modified in a way that renders VaR and ES more efficient.

## Volatility-weighted Historical Simulation

Instead of weighting individual observations by proximity to the current date, we can also weight data by relative volatility. This idea was originally put forth by Hull and White to incorporate changing volatility in risk estimation. The underlying argument is that if volatility has been on the rise in the recent past, then using historical data will underestimate the current risk level. Similarly, if current volatility has significantly reduced, then using historical data will overstate the current risk level.

If $r_t$,i, is the historical return in asset i on day t in our historical sample, $\sigma_{t,i}$, the historical GARCH (or EWMA) forecast of the volatility of the return on asset i for day t, and $\sigma_{T,i}$, the most recent forecast of the volatility of asset i, then the volatility-adjusted return is:

$$r_{t,i}^* = \frac{\sigma_{T,i}}{\sigma_{t,i}} r_{t,i}$$

Actual returns in any period t will therefore increase (or decrease), depending on whether the current forecast of volatility is greater (or less than) the estimated volatility for period t.

Advantages of the volatility-weighted approach relative to equal-weighted or age-weighted approaches include:

- The approach explicitly incorporates volatility into the estimation procedure. The equal-weighted HS completely ignores volatility changes. Although the age-weighted approach recognizes volatility, its treatment is rather arbitrary and restrictive.
- The method produces near-term VaR estimates that are likely to be more sensitive to current market conditions.
- Volatility-adjusted returns allow for VaR and ES estimates that that can exceed the maximum loss in our historical data set. Under traditional HS, VaR or ES cannot be bigger than the losses in our historical data set.
- Empirical evidence indicates that this approach produces VaR estimates that are superior to the VaR estimates under the age-weighted approach.

## Correlation-weighted Historical Simulation

Historical returns can also be adjusted to reflect changes between historical and current correlations. In other words, this method incorporates updated correlations between asset pairs. In essence, the historical correlation (or equivalently variance-covariance) matrix is adjusted to the new information environment by "multiplying" the historic returns by the revised correlation matrix to yield updated correlation-adjusted returns.

## Filtered Historical Simulation

The filtered historical simulation is undoubtedly the most comprehensive, and hence most complicated, of the non-parametric estimators. The method aims to combine the benefits of historical simulation with the power/flexibility of conditional volatility models(like GARCH or asymmetric GARCH).

Steps involved:

1. A conditional volatility model (e.g., GARCH) is fitted to our portfolio-return data.
2. Actual returns are translated into standardized returns
3. The conditional volatility model is used to forecast volatility for each of the days in a sample period.
4. These volatility forecasts are then divided into the realized returns to produce a set of standardized returns that are iid (independent and identically distributed)
5. A bootstrapping exercise is performed assuming a 1-day VaR holding period.
6. The VaR is computed

## Advantages and Disadvantages of Non-Parametric Methods

### Advantages

i. They are instinctive and conceptually simple.
ii. They can accommodate fat tails, skewness, and other abnormal features to parametric approaches.
iii. They can accommodate any type of position including derivative positions.
iv. HS works quite well empirically.
v. In varying degrees, they are quite easy to implement on a spreadsheet.
vi. They are free of operational problems.
vii. They use readily available data.
viii. Results provided are easily reported and communicated to seniors.
ix. Confidence intervals for nonparametric VaR and ES are easily produced.
x. When combined with add-ons they are capable of refinement and potential improvement.

### Disadvantages

i. For unusually quiet data periods, VaR and ES estimates are too low for actual risks faced.
ii. For unusually volatile data periods the estimates for VaR or ES produced are too high.
iii. Difficulty in handling shifts during sample periods.

iv. An extreme loss in the data set dominates non-parametric risk estimates.
v. Subject to the phenomenon of ghost effect or shadow effects.
vi. They are constrained by the largest loss in historical data.

## Question 1

Assume that Mrs. Barnwell a risk manager has a portfolio with only 2 positions with a historical correlation between them being 0.5. She wishes to adjust her historical returns R to reflect a current correlation of 0.8. Which of the following best reflects the 0.8 current correlation?

A. $\begin{pmatrix} 1 & 0.3464 \\ 0 & 0.6928 \end{pmatrix} R$

B. $\begin{pmatrix} 0 & 0.3464 \\ 1 & 0.6928 \end{pmatrix} R$

C. $\begin{pmatrix} 1 & 0 \\ 0.3464 & 0.6928 \end{pmatrix} R$

D. 0.96R

Correct answer is **C**

Recall if $a_{i,j}$ is the i, jth element of the 2 x 2 matrix A, then by applying Choleski decomposition, $a_{11} = 1$, $a_{12} = 0$, $a_{21} = \rho$, $a_{22} = \sqrt{1-\rho^2}$. From Our data, $\rho = 0.5$, Matrix $\bar{A}$ is similar but has a $\rho = 0.8$.

Therefore:

$$A^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}$$

Substituting

$$\hat{R} = \bar{A}A^{-1}R$$

We get

$$\begin{pmatrix} 1 & 0 \\ 0.8 & \sqrt{1-0.8^2} \end{pmatrix} \frac{1}{\sqrt{1-0.5^2}} \begin{pmatrix} \sqrt{1-0.5^2} & 0 \\ -0.5 & 1 \end{pmatrix} R$$

$$= \begin{pmatrix} 1 & 0 \\ 0.3464 & 0.6925 \end{pmatrix} R$$

## Question 2

Given that the mean return from a dataset has been pre-calculated and is given as 0.04. The standard deviation has also been given as 0.32. With 90% confidence, what will be our maximum percentage loss? Assume that from our dataset, Z= -0.28 and N (Z) = 0.10 since you are to locate the value at the 10 percentile.

A. 36.96%
B. 11.27%
C. 11.32%
D. 36.72%

The correct answer is **A**

Recall that

$$Z = \frac{X - \mu}{\sigma}$$

From the data we are given that : $\mu = 0.04$, $\sigma = 0.32$ and $Z = -1.28$

Therefore:

$$-1.28 = \frac{X - 0.04}{0.32} \Rightarrow X = -1.28\,(0.32) + 0.04 = -0.3696$$

$$X = -0.3696 = 36.96\% \text{ loss}$$

This means that we are 90% confident that the maximum loss will not exceed 36.96%

## Question 3

A dataset is given such that, the kurtosis in its distribution is 8, $x_\tau$ is 1.57 and a chosen bound on the percentage deviation given as 30.24. What is the required number of the resamples?

A. 54

B. 30

C. 34

D. 47

The correct answer is **D**.

Recall that from Standard Errors of Bootstrap Estimators:

$$Pr\left[100\left|\frac{\dot{S}_{B}-\dot{S}-}{\dot{S}_{B}}\right| \leq \text{bound}\right] = \tau$$

we have:

$$B = \frac{2500\,(k-1)\,x_{\tau}^{2}}{\text{bound}^{2}} \qquad \text{(a)}$$

We are given that $k = 8$, $x_{\tau} = 1.57$, $\text{bound} = 22.7$. Applying these values in the equation (a) gives:

$$B = \frac{2500\,(8-1) \times 1.57^{2}}{30.24^{2}} = 47.17$$

$$\approx 47$$