An Introduction to Value at Risk[1]

This chapter provides an introduction to value at risk. We examine five basic models for calculating value at risk, and how to assess the effectiveness of value at risk models through backtesting.

What is Value at Risk?

Value at risk (VaR) is one of the most widely used risk measures in finance. VaR was popularized by J.P. Morgan in the 1990s. The executives at J.P. Morgan wanted their risk managers to generate one statistic at the end of each day, which summarized the risk of the firm's entire portfolio. What they came up with was VaR.

If the 95% VaR of a portfolio is $100, then we expect the portfolio will lose $100 or less in 95% of the scenarios, and lose $100 or more in 5% of the scenarios. We can define VaR for any confidence level, but 95% has become an extremely popular choice in finance. The time horizon also needs to be specified for VaR. On trading desks, with liquid portfolios, it is common to measure the one-day 95% VaR. In other settings, in which less liquid assets may be involved, time frames of up to one year are not uncommon. VaR is decidedly a one-tailed confidence interval.
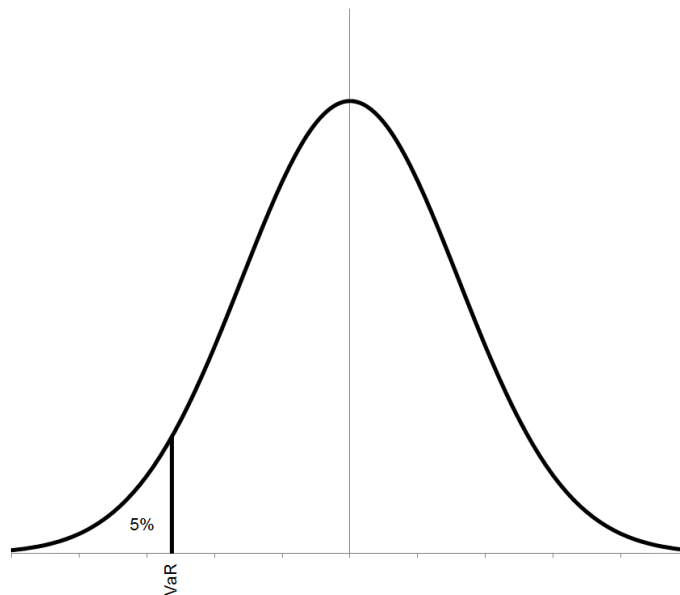

Figure 1

Figure 1 provides a graphical representation of VaR at the 95% confidence level. The figure shows the probability density function for the returns of a portfolio. Because VaR is being measured at the 95% confidence level, 5% of the distribution is to the left of the VaR level, and 95% is to the right.

In order to formally define VaR we begin by defining a random variable L, which represents the loss to our portfolio. L is simply the negative of the return to our portfolio. If the return of our portfolio is −$600, then the loss, L, is +$600. For a given confidence level, γ, then, we can define value at risk as:

$$P\big[L \geq \text{VaR}_\gamma\big] = 1 - \gamma$$

<div align="center">Equation 1</div>

If a risk manager says that the one-day 95% VaR of a portfolio is $400 this means that there is a 5% probability that the portfolio will *lose* $400 or more on any given day, that L will be more than $400.

If an actual loss exceeds the predicted VaR threshold, that event is known as an exceedance. Another assumption of VaR models is that exceedance events are uncorrelated with each other. In other words, if our VaR measure is set at a one-day 95% confidence level, and there is an exceedance event today, then the probability of an exceedance event tomorrow is still 5%. An exceedance event today has no impact on the probability of future exceedance events. More generally, the probability of an exceedance conditional on all available information should equal the unconditional probability of an exceedance. In other words, an exceedance event should be no more likely to occur on a Tuesday, the day after the market is up, or when the level of risk is high. We return to these assumptions at the end of this chapter when we discuss back-testing.

VaR has become extremely popular in risk management. The appeal of VaR is its simplicity. Because VaR can be calculated for any portfolio, it allows us to easily compare the risk of different portfolios. Because it boils risk down to a single number, VaR provides us with a convenient way to track the risk of a portfolio over time. Finally, the concept of VaR is intuitive, even to those not versed in statistics. Because it is so popular, VaR has come under a lot of criticism. As we will see, VaR is not a perfect risk measure, but it is incredibly versatile. While some of the criticism is justified, much of the criticism is misplaced.

Delta-Normal VaR

One of the simplest and easiest ways to calculate VaR is to make what are known as delta-normal assumptions. For any underlying asset, we assume that the log returns are normally distributed, and we approximate the returns of any option based on its delta-adjusted exposure. The delta-normal model includes additional assumptions when multiple securities are involved, which we will cover when we begin to look at portfolio risk measures.

The delta-normal assumptions make it very easy to calculate VaR statistics even with limited computing power. This made delta-normal models a popular choice when VaR models were first introduced. Predictably, the results of such a simple model were often disappointing. A good risk

manager would often be able to compensate for these deficiencies, but the basic model presented an easy target for critics. Delta-normal models are rarely used in practice today, but are still an excellent starting point when learning about VaR models. By understanding the pros and cons of the delta-normal model we will be better able to understand the pros and cons of more complex models. Unfortunately many people outside of risk management believe that delta-normal models are still widely used in practice, or believe that the shortcomings of these simple models are somehow inherent to all risk models.

To calculate the delta-normal VaR of a security, we start by calculating the standard deviation of returns for the security, or, in the case of an option, for the returns of the option's underlying. For regular securities, we then multiply the return standard deviation by the absolute market value or notional of our position to get the position's standard deviation. For options we multiply by the absolute delta-adjusted exposure. The delta adjusted exposure being the value of the underlying multiplied by the option's delta. We then multiply the position's standard deviation by an appropriate factor based on the inverse of the standard normal distribution (e.g. −1.64 for 95% VaR).

Notice that we have not said anything about the expected return. In practice, most VaR models assume that the distribution of return has a mean of zero. This is almost always a very reasonable assumption at short horizons. At longer horizons this assumption may no longer be reasonable. Some practitioners will also assume that the time decay for options is also zero. While this assumption may also be valid in many situations, it can fail even over short time horizons. In what follows, unless stated otherwise, assume security returns have zero mean but include theta in calculating VaR.

Sample Problem

Question

You estimate the standard deviation of daily returns for XYZ Corp.'s stock at 2.00%. You own both the stock and a call option on the stock with a delta of 40% and 1-day theta of −0.01. The underlying price is $100. Calculate the 1-day 95% daily VaR for each security.

Answer

The 95% VaR corresponds to the bottom 5% of returns. For a normal distribution 5% of the distribution is less than 1.64 standard deviations below the mean. We can get this result from a lookup table, from a statistics application, or from a spreadsheet. For example, in Excel *=NORMSINV(0.05)* would give us −1.64, the negative sign indicating that the result is below the mean.

For the stock, the final answer is simply: −1.64 × $100 × 2.00% = −$3.28 (if you use additional decimal places for the number of standard deviation, you might get −3.29%). For the option, the final answer is only slightly more complicated −1.64 × 40% × $100 × 2.00% −0.01 = −$1.32.

The 1-day 95% daily VaR for the stock and option are a loss of $3.28 and $1.32, respectively.

Historical VaR

Another very simple model for estimating VaR is historical simulation or the historical method. In this approach we calculate VaR directly from past returns. For example, suppose we want to calculate the 1-day 95% VaR for an equity using 100 days of data. The 95th percentile corresponds to the least worst of the worst 5% of returns. In this case, because we are using 100 days of data, the VaR simply corresponds to the 5th worst day. If we have the following 100 returns, sorted from lowest to highest:

	Rank	Return
Worst	1	-4.00%
	2	-3.62%
	3	-3.57%
	4	-3.52%
	5	**-3.37%**
	6	-3.24%
	7	-3.14%
	8	-3.07%
	9	-2.92%
	…	…
	99	3.79%
Best	100	4.08%

Table 1

The 95th percentile VaR in this case corresponds to −3.37%, or, dropping the negative sign, we would say that our 1-day 95% VaR is a loss of 3.37%.

For an infinitely-lived underlying asset, the historical approach could not be easier. For derivatives, such as equity options, or other instruments with finite lifespans, such as bonds, it is slightly more complicated. For a derivative, we do not want to know what the actual return series was, we want to know what the return series would have been had we held exactly the same derivative in the past. For example, suppose we own an at-the-money put with two days until expiry. 250 days ago, the option would have had 252 days until expiry, and it may have been far in or out of the money. We do *not* want to know what the return would have been for this option with 252 to expiry, we want to know what the return *would have been* for an at-the-money put with two days to expiry, given conditions in the financial markets 250 days ago. Similarly, for a bond with 30 days to expiry, for risk purposes, we are interested what the return of a bond with 30 days to maturity would have been 250 days ago, *not* what the return of a bond with 280 days to maturity was. These constant maturity series, or back-cast series, are quite common in finance. The easiest way to calculate the back-cast series for an option would be to use a delta approximation. If we currently hold a put with a delta of −30%, and the underlying return 250 days ago was 5%, then our back-cast return for that day would be −1.5% = −30% × 5%. A more accurate approach would be to fully reprice the option, taking into account not just changes in the underlying, but time-decay, changes in implied volatility and changes to the risk-free rate. Just as we could approximate option returns using delta, we could approximate bond returns using DV01, but a

more accurate approach would be to fully reprice the bond based on changes in the relevant interest rates and credit spreads.

One advantage of historical VaR is that it is extremely simple to calculate. Another advantage is that it is easy to explain to non-risk professionals. Most financial professionals will be very used to looking at cumulative return charts. The returns used to create a cumulative return chart are the same returns used to calculate historical VaR. If there is ever a question about the validity of a historical VaR calculation it is easy enough to pull up a chart of historical returns to look for a potential source of error.

The delta-normal approach is an example of what we call a parametric model. We say that the model is parametric because it is based on a mathematically defined, or parametric, distribution (in this case, the normal distribution). By contrast the historical approach is non-parametric. We have not made any assumptions about the distribution of historical returns. There are advantages and disadvantages to both approaches. The historical approach easily reproduces all the quirks that we see in historical data, changing standard deviation, skewness, kurtosis, jumps, etc. Developing a parametric model that reproduces all of the observed features of financial markets can be very difficult. At the same time, models based on distributions often make it easier to draw general conclusions. In the case of the historical approach, it may not be easy to tell if a VaR forecast is the result of a particularly unusual set of input returns.

Monte Carlo Simulation

Monte Carlo simulations are widely used throughout finance, and they can be a very powerful tool for calculating VaR. As an example of how we would calculate VaR using a Monte Carlo simulation, imagine we have a position in gold, and we believe that the daily log returns of gold are normally distributed with a mean of 0.01% and a standard deviation of 1.40%. To calculate the VaR of this position, we could generate 1,000 draws from a normal distribution with a mean of 0.01% and a standard deviation of 1.40%, convert the log returns into standard returns, and then sort the returns from lowest to highest. If we are interested in our 95% VaR, we simply select the 50th worst return from the list. For this set up, the Monte Carlo simulation is very straightforward, but it is also inefficient. Because the log returns are normally distributed, we know that the 5th percentile is −1.64 standard deviations below the mean, corresponding to a log return of −2.29% = 0.01% −1.64 × 1.40%.

The real power of Monte Carlo simulations is in more complex settings, where instruments are non-linear, prices are path dependent, and distributions do not have well defined inverses.

Monte Carlo simulations also make it easy to calculate multi-period VaR. In the preceding example, if instead of being interested in the 1-day VaR, we wanted to know the 4-day VaR, we would simply generate four 1-day log returns, using the same distribution as before, and add them together to get one 4-day return. We could repeat this process 1,000 times, generating a total of 4,000 1-day returns. As with the 1-day example, in this particular situation, there are more efficient ways to calculate the VaR statistic. That said, it is easy to imagine how multiday scenarios could quickly become very complex. What if your policy was to reduce your position by 50% every time you suffered a loss in excess of 3%? What if returns exhibited positive serial correlation, with positive excess returns more likely to be

followed by positive excess returns, and negative excess returns more likely to be followed by negative excess returns?

Monte Carlo simulations are usually based on parametric distributions, but we could also use non-parametric methods, randomly sampling from historical returns. Continuing with our gold example, if we had 500 days of returns for gold, and we wanted to calculate the 4-day VaR, we would randomly pick a number from 1 to 500, and select the corresponding historical return. We would do this four times, to create one 4-day return. We can repeat this process, generating as many 4-day returns as we desire. The basic idea is very simple, but there are some important details to keep in mind. First, generating multi-period returns this way involves what we call sampling with replacement. Pretend that the first draw from our random number generator is a 10, and we select the 10th historical return. We don't remove that return before the next draw. If, on the second draw, our random number generator produces 10 again, then we select the same return. If we end up pulling 10 four time in a row, then our 4-day return will be composed of the same 10th return repeated four times. Even though we only have 500 returns to start out with, there are 500^4, or 62.5 billion, possible 4-day returns that we can generate this way. This method of estimating parameters using sampling with replacement is often referred to as bootstrapping. The second detail that we need to pay attention to is serial correlation. We can only generate multi-period returns in the way just describe if single-period returns are independent of each other. For example, suppose that the standard deviation of gold has gone through long periods of high volatility followed by long periods of low volatility, and we believe our historical data accurately reflects this starting with 250 days of low volatility followed by 250 days of high volatility. If we randomly select returns with replacement, then the probability of getting a draw from the high volatility period is 1/2 each time. If our random numbers are generated independently then there is only $1/16 = (1/2)^4$ chance of drawing four returns in a row from the high period, whereas historically the probability was much closer to 1/2 (except for the transition in the middle of the sample, where we switched from low to high volatility, low volatility days were always followed by low volatility days, and high volatility days were always followed by high volatility days). A simple solution to this problem: instead of generating a random number from 1 to 500, generate a random number from 1 to 497, and then select four successive returns. If our random number generator generates 125, then we create our 4-day return from returns 125, 126, 127, and 128. While this method will capture any serial dependence between periods, it greatly reduces the number of possible returns from 62.5 billion to 497, and effectively reduces the Monte Carlo simulation to the historical simulation method.

Of the three methods we have considered so far, Monte Carlo simulations are generally considered to be the most flexible. Their major drawback is speed. As computers get faster and faster, the speed of Monte Carlo simulations is becoming less of an issue. Still in some situations — a trading desk that require real-time risk number, for example —this speed issue may still rule out the use of Monte Carlo simulations.

Hybrid VaR
For the historical method or historical simulation, all of the data points are given equal weight. In practice, because market risk tends to changes over time, it might make sense to give more weight to more recent data. One very simple way to do this is to apply exponentially decreasing weight to the

historical data. For example, with a decay factor of 0.99, we would apply a weight of 1.00 to the most recent return, 0.99 to the second most recent, 0.99^2 to third, and so on.

This general approach, using historical returns with decreasing weights, is often called the hybrid approach because it combines aspects of standard historical simulation and weighted parametric approaches; see, for example, Allen, Boudoukh, and Saunders (2004).

Suppose we have 100 returns, sorted from worst to best, as before. In the case of the historical simulation, we found the VaR by moving down the table until we got to the fifth data point. For the hybrid approach we simply move down the table until we get to 5% of the total weights:

	Rank	t	Return	Weight	% of Total Weight	Cumulative % Weight
Worst	1	50	-4.00%	0.61	0.95%	0.95%
	2	40	-3.62%	0.55	0.86%	1.82%
	3	9	-3.57%	0.40	0.63%	2.45%
	4	48	-3.52%	0.59	0.94%	3.38%
	5	37	-3.37%	0.53	0.84%	4.22%
	6	**18**	**-3.24%**	**0.44**	**0.69%**	**4.91%**
	7	**54**	**-3.14%**	**0.63**	**0.99%**	**5.91%**
	8	63	-3.07%	0.69	1.09%	7.00%
	9	8	-2.92%	0.40	0.63%	7.62%
	...	...	...	...	...	...
	11	100	-2.80%	1.00	1.58%	10.60%
	...	...	...	...	...	...
	99	41	3.79%	0.55	0.87%	99.32%
Best	100	16	4.08%	0.43	0.68%	100.00%

Table 2

In this case, we get to 5% of the total weight between the sixth and seventh returns. At this point there are two approaches to deciding the VaR. The more conservative approach is to take the sixth return, -3.24%. The alternative is to interpolate between the sixth and seventh returns, to come up with -3.23%. Unless there is a strong justification for choosing the interpolation method, the conservative approach is recommended.

The hybrid approach is fairly easy to implement, computationally efficient, and is underpinned by familiar historical returns. While only slightly more complicated than the delta-normal and historical simulation methods, the hybrid approach often produces much more reliable estimates of VaR.

Cornish-Fisher VaR

The delta-normal VaR model assumes that underlying returns are normally distributed and that option returns can be approximated using their delta-adjusted exposure. The Cornish-Fisher VaR model maintains the first assumption, while trying to improve the approximation for options. The method relies on what is known as the Cornish-Fisher expansion. The Cornish-Fisher expansion is a general

method that allows us to approximate the confidence intervals for a random variable based on the central moments of that variable. As with the delta-normal approach, the Cornish-Fisher approach can easily be extended to portfolios containing multiple securities

To start with, we introduce some notation. Define the value of an option as *V*, and the value of the option's underlying as *U*, we can define the option's exposure-adjusted Black-Scholes Greeks as follows:

$$\widetilde{\Delta} = \frac{dV}{dU} U = \Delta U$$

$$\widetilde{\Gamma} = \frac{d^2V}{dU^2} U^2 = \Gamma U^2$$

$$\theta = \frac{dV}{dt}$$

Given a return on the underlying, *R*, we can then approximate the change in value of the option using the exposure-adjusted Greeks as:

$$dV \approx \widetilde{\Delta}R + \frac{1}{2}\widetilde{\Gamma}R^2 + \theta dt$$

If the returns of the underlying, *R*, are normally distributed with a mean of zero and a standard deviation of σ, then we can calculate the moments of *dV* based on Equation 3. The first three central moments and skewness of of *dV* are:

$$\mu_{dV} = E[dV] = \frac{1}{2}\widetilde{\Gamma}\sigma^2 + \theta dt$$

$$\sigma_{dV}^2 = E[(dV - E[dV])^2] = \widetilde{\Delta}^2\sigma^2 + \frac{1}{2}\widetilde{\Gamma}^2\sigma^4$$

$$\mu_{3,dV} = 3\widetilde{\Delta}^2\widetilde{\Gamma}\sigma^4 + \widetilde{\Gamma}^3\sigma^6$$

$$s_{dV} = \frac{\mu_{3,dV}}{\sigma_{dV}^3}$$

Where μ_{dV} is the mean of *dV*, σ_{dV}^2 is the variance, $\mu_{3,dV}$ is the third central moment, and s_{dV} is the skewness. Notice that even though the distribution of the underlying is symmetric (the underlying distribution is normally distributed), the distribution of the option is skewed ($s_{dV} \neq 0$). This makes sense, given the asymmetric nature of the option payout function. That the Cornish-Fisher model captures the asymmetry of options is an advantage over the delta-normal model, which produces symmetrical distributions for options.

These central moments from Equation 4 can be combined to approximate a confidence interval using a Cornish-Fisher expansion, which can in turn be used to calculate an approximation for VaR. The Cornish-Fisher VaR of the position is given by:

$$-VaR = \mu_{dV} + \sigma_{dV}\left[m + \frac{1}{6}(m^2 - 1)s_{dV}\right]$$

<div align="center">Equation 5</div>

where m corresponds to the distance in standard deviations for our VaR confidence level based on a normal distribution. Unfortunately, this formula is far from intuitive, and its derivation is beyond the scope of this book. The easiest way to understand the approximation is to use it to calculate VaR.

Sample Problem

Question

You are asked to evaluate the risk of a portfolio containing a single call, with a strike price of 110, and three months to expiry. The underlying price is 100, and the risk free rate is 3%. The expected and implied standard deviations are both 20%. Calculate the 1-day 95% VaR using both the delta-normal method and the Cornish-Fisher method. Use 365 days per year for theta and 256 days per year for standard deviation.

Answer

To start with we need to calculate the Black-Scholes delta, gamma, and theta. These can be calculated in a spreadsheet or using other financial applications.

$$\Delta = 0.2038$$
$$\Gamma = 0.0283$$
$$\Theta = -6.2415$$

The one day standard deviation and theta can be found as follows:

$$\sigma_d = \frac{20\%}{\sqrt{256}} = 1.25\%$$
$$\Theta_d = \left(\frac{-6.2415}{365}\right) = -0.0171$$

Using $m = -1.64$ for the normal 5% confidence level, the delta-normal approximation is:

$$-VaR = m\sigma_d S\Delta + \Theta_d = -1.64 \cdot 1.25\% \cdot 100 \cdot 0.2038 - 0.0171 = -0.4361$$

For the Cornish-Fisher, approximation, we first calculate the exposure adjusted Greeks:

$$\tilde{\Delta} = \Delta U = 0.2038 \cdot 100 = 20.3806$$
$$\tilde{\Gamma} = \Gamma U^2 = 0.0283 \cdot 100^2 = 283.1397$$

Next we calculate the mean, standard deviation, and skewness for the change in option value:

$$\mu_{dV} = \frac{1}{2}\tilde{\Gamma}\sigma_d^2 + \Theta_d = \frac{1}{2} \cdot 283.1379 \cdot (1.25\%)^2 - 0.0171 = 0.00503$$

$$\sigma_{dV}^2 = \tilde{\Delta}^2\sigma_d^2 + \frac{1}{2}\tilde{\Gamma}^2\sigma_d^4 = 20.3806^2 1.25\%^2 + \frac{1}{2}283.1397^2 1.25\%^4 = 0.06588$$

$$\sigma_{dv} = \sqrt{0.06588} = 0.256672$$

$$\mu_{3,dV} = 3\tilde{\Delta}^2\tilde{\Gamma}\sigma_d^4 + \tilde{\Gamma}^3\sigma_d^6 = 3 \cdot 20.3806^2 \cdot 283.1379 \cdot 1.25\%^4 + 283.1379^3 1.25\%^6 = 0.0087$$

$$s_{dV} = \frac{\mu_{3,dV}}{\sigma_{dV}^3} = \frac{0.0087}{0.256672^3} = 0.51453$$

We then plug the moments into our Cornish-Fisher approximation, Equation 5:

$$-VaR = \mu_{dV} + \sigma_{dV}\left[m + \frac{1}{6}(m^2 - 1)s_{dV}\right] = 0.00502 + 0.256672\left[-1.64 + \frac{1}{6}(-1.64^2 - 1)0.514528\right]$$

$$= -0.3796$$

The 1-day 95% VaR for the Cornish-Fisher method was a loss of 0.3796, compared to a loss of 0.4316 for the delta-normal approximation. It turns out that in this particular case the exact answer can be found using the Black-Scholes equation. Given the assumptions of this sample problem, the actual 1-day 95% VaR is 0.3759. In this sample problem, the Cornish-Fisher approximation is very close to the actual value, and provides a much better approximation than the delta-normal approximation.

In certain instances, the Cornish-Fisher approximation can be extremely accurate. In practice, it is much more likely to be accurate if we do not go too far out into the tails of the distribution. If we try to calculate the 99.99% VaR using Cornish-Fisher, even for a simple portfolio, the result is unlikely to be as accurate as what we saw in the preceding sample problem. This is because returns are much more likely to be well approximated by a normal distribution closer to the mean of a distribution. As we will see in the next section there are other reasons why we might not want to calculate VaR too far out into the tails.

Quantile Regression

Another approach to calculating VaR is to use what is known as quantile regression. Pretend that we want to model returns using the following simple univariate regression model (See Miller (2012) Chapter 8 for a review of regression analysis):

$$r_t = \alpha + \beta x_t + \varepsilon_t$$

<div align="center">

Equation 6

</div>

where r_t is the return at time t, x_t is an explanatory variable, ε_t is an i.i.d. disturbance term, and α and β are constants. The standard method for estimating α and β using ordinary least squares (OLS) regression is to minimize the sum of squared errors. That is, we want to choose estimators for α and β, $\hat{\alpha}$ and $\hat{\beta}$, such that they minimize the following sum:

$$SSE = \sum_{t=1}^{T} e_t^2 = \sum_{t=1}^{T} \left(r_t - \hat{\alpha} - \hat{\beta} x_t \right)^2$$

Why do we minimize the sum of squared errors? Why not the sum of errors to the fourth power or the sum of absolute errors? It turns out that the choice of sum of squared errors is somewhat arbitrary (not completely arbitrary, but somewhat arbitrary), and under certain circumstances using other methods to calculate the regression parameters can be optimal. Minimizing the sum of absolute errors, for example, can produce more stable parameter estimates when the distribution of returns is fat tailed.

Minimizing the sum of absolute errors doesn't seem so strange, but what would happen if we tried to set the regression parameter so that 10% of the time errors were negative and 90% of the time errors were positive?
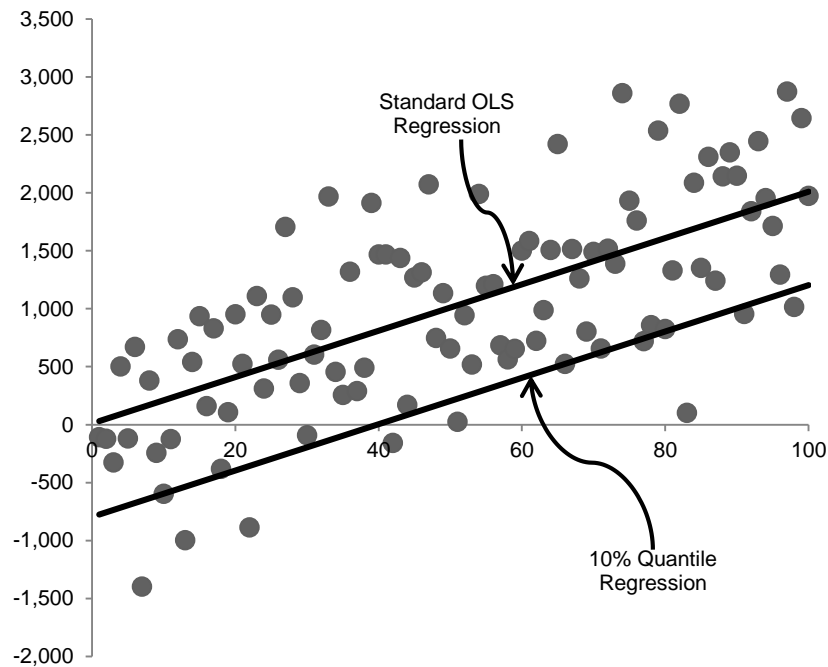

Figure 2

The estimators in this case would produce what is known as a quantile[2] regression. Figure 2 shows some sample data over which has been drawn the standard OLS regression line and a 10% quantile regression line. The 10% quantile regression line in effect draws a border. We expect to find 10% of the points

---

[2] For those not familiar with the term a quantile is the generic term describing equally sized partitions of a data set. The top and bottom deciles of the income distribution refer to the 10% of people who earn the most and the 10% of people who earn the least, respectively. The top quintile refers to the top 20%. Percentiles divide a distribution a dataset into 100 equal parts. Percentiles, deciles and quintiles are all examples of quintiles.

below the line and 90% above the line. In other words, it is the 90% VaR. The advantage of a quantile regression is that it produces a VaR statistic directly, without making any distributional assumptions.

Sample Problem

Question
Another risk analyst at your firm noticed that the firm's VaR was highly correlated with the level of the CBOE Market Volatility Index (VIX). The analyst estimated the following 5% quantile regression for 1-day returns:

$$r_t = 0.01 - 0.10v_t + \varepsilon_t$$

where v_t is the level of the VIX at time t. What is the expected 1-day 95% VaR when the VIX is at 20%? 40%?

Answer
When the VIX is at 20%, the VaR is a loss of 1%, or −1%:

$$E[r_t] = 0.01 - 0.10 \cdot 0.20 + 0.0 = 0.01 - 0.02 = -0.01$$

Similarly, when the VIX is at 40%, the VaR is a loss of 3%.

So how do we actually find the quantile regression parameters? In 1978, Koenker and Basset showed that for the θ quantile, where 0<θ<1, the quantile regression parameters can be found by minimizing the following:

$$\sum_{r_t \geq \hat{\alpha} - \hat{\beta}x_t} \theta|r_t - \hat{\alpha} - \hat{\beta}x_t| + \sum_{r_t < \hat{\alpha} - \hat{\beta}x_t} (1 - \theta)|r_t - \hat{\alpha} - \hat{\beta}x_t|$$

Notice that the first summation only counts those points where r_t is greater than or equal to the expected value, given x_t. Similarly, the second summation only counts those points where r_t is less than expected, given x_t. That minimizing this equation produces the correct quantile regression parameters is not obvious. To get some intuition for this result, it might help to consider what happens at the extremes. If θ is close to zero, then we want r_t to be greater than the quantile regression line most of the time, that is greater than $\hat{\alpha} - \hat{\beta}x_t$ most of the time. In other words, if we choose the parameter correctly, most of the data points will end up in the first summation, where the absolute deviation, $|r_t - \hat{\alpha} - \hat{\beta}x_t|$, will be multiplied by something very close to zero, namely θ, resulting in a very low value for the equation. Similarly, if θ is close to one then most of the data points should end up in the second summation where they will be multiplied by (1- θ), which, in this case, is very close to zero.

In the special case where θ = 0.5, Equation 8 is equivalent to reducing the sum of absolute errors. The 50th percentile is also known as the median, so minimizing the sum of absolute errors produces a quantile regression line that defines the median for the regression. Be careful not to confuse this with the median of the dependent variable. Using our notation from Equation 6, the quantile regression line for θ = 0.5 defines the median of r_t *given* x_t, not the median of r_t.

Quantile regression is a very generic method. In the last sample problem we showed just one possible way in which quantile regression could be used to estimate VaR based on a univariate regression model. In practice we can apply quantile regression to many other types of models. For example we could use quantile regression to estimate the parameters of a standard GARCH(1,1) time series model.

No matter what form the model takes, the advantages of quantile regression are that it focuses on the quantile directly, it is robust to outliers, and it is non-parametric. The primary disadvantage of quantile regression is that there is no closed form solution to Equation 8. Unlike standard OLS regression, where the parameters can be determined using arithmetic or simple linear algebra, quantile regression requires linear programming or some other optimization algorithm for parameter estimation. With modern computers this is hardly an insurmountable problem. Simple quantile regressions can easily be solved using Excel's Solver. Even still, the lack of a closed form solution makes understanding the parameters of a quantile regression more complicated. All of the tools that we have developed for OLS regression — goodness of fit, t-tests for parameter significance, etc. — need to be reevaluated for quantile regression. Because of this and because it is applicable to a much smaller class of problems, quantile regression is taught very rarely in statistics classes, even at the graduate level. This may sound like a catch-22, but is nonetheless a very real obstacle to the widespread adoption of this technique.

Backtesting

An obvious concern when using VaR is choosing the appropriate confidence interval. As mentioned, 95% has become a very popular choice in risk management. In some settings there may be a natural choice for the confidence level, but most of the time the exact choice is arbitrary.

A common mistake for newcomers is to choose a confidence level that is too high. Naturally, a higher confidence level sounds more conservative. A risk manager who measures one-day VaR at the 95% confidence level will, on average, experience an exceedance event every 20 days. A risk manager who measures VaR at the 99.9% confidence level expects to see an exceedance only once every 1,000 days. Is an event that happens once every 20 days really something that we need to worry about? It is tempting to believe that the risk manager using the 99.9% confidence level is concerned with more serious, riskier outcomes, and is therefore doing a better job.

The problem is that, as we go further and further out into the tail of the distribution, we become less and less certain of the shape of the distribution. In most cases, the assumed distribution of returns for our portfolio will be based on historical data. If we have 1,000 data points, then there are 50 data points to back up our 95% confidence level, but only one to back up our 99.9% confidence level. As with

any distribution parameter, the variance of our estimate of the parameter decreases with the sample size. One data point is hardly a good sample size on which to base a parameter estimate.

A related problem has to do with backtesting. Good risk managers should regularly backtest their models. Backtesting entails checking the predicted outcome of a model against actual data. Any model parameter can be backtested.

In the case of VaR, backtesting is easy. Each period can be viewed as a Bernoulli trial. In the case of one-day 95% VaR, there is a 5% chance of an exceedance event each day, and a 95% chance that there is no exceedance. Because exceedance events are expected to be independent, over the course of n days the distribution of exceedances follows a binomial distribution:

$$P[K = k] = \binom{n}{k} p^k (1 - p)^{n-k}$$

<div align="center">Equation 9</div>

In this case, n is the number of periods that we are using to backtest, k is the number of exceedances, and $(1 - p)$ is our confidence level.

Sample Problem

Question
As a risk manager, you are tasked with calculating a daily 95% VaR statistic for a large commodities portfolio. Over the past 60 days, there have been two exceedances. How many exceedances should you have expected? What was the probability of exactly two exceedances during this time? Two or less?

Answer
Remember, by convention, for a 95% VaR the probability of an exceedance is 5%, not 95%. Over 60 days, we would expect to see 3 exceedance events, 3 = 5% × 60.

The probability of exactly two exceedances is 22.59%:

$$P[K = 2] = \binom{60}{2} 0.05^2 (1 - 0.05)^{60-2} = 0.2259$$

To get the probability of two or less exceedances, we simply perform the same calculation of all values less than or equal to two: zero, one and two. It is important not to forget zero:

$$P[K \le 2] = \sum_{k=0}^{2} \binom{60}{k} 0.05^k (1 - 0.05)^{60-k} = 0.0461 + 0.1455 + 0.2259 = 0.4147$$

The probability of two or less exceedances was 41.47%. Even though we had less exceedances than expected, this outcome was not unlikely given the short time frame. In this case, the fact that we had less exceedances than expected is probably not a cause for concern.

As stated at the beginning of the chapter, another assumption of the VaR model is that the probability of an exceedance event conditional on all available information is equal to the unconditional probability of an exceedance. For example suppose we looked back over several years of data and observed that 30% of exceedances occurred on Tuesday and only 10% occurred on Thursdays. In theory, an exceedance should be equally likely on any given day of the week. In the long run, we would expect to see the exceedances equally allocated to each of the five business days, or 20% to each. Of course the 30%/10% split between Tuesday and Thursday could just be a fluke. However, if the split turned out to be statistically significant, we would need to make an adjustment to our model.

A common problem with VaR models is serial correlation in exceedances. The probability of an exceedance today, given that an exceedance occurred yesterday, should be no higher or lower than on any other day. A simple way to test for this serial correlation is to count the number of exceedances, and then count the number of times there is an exceedance on the day after an exceedance. The distribution of the day-after exceedances should also follow a binomial distribution. If we are forecasting a 1-day 95% VaR, and there have been 40 exceedances, then we would expect that 2 of those exceedances occurred after another of the 40 exceedances, 2 = 40 × 5%.

Another common problem with VaR models is the tendency for the probability of exceedances to vary with the level of risk. This may seem counterintuitive, but exceedances should be no more likely to occur when risk is high than when risk is low. If we are measuring 95% VaR then there is always a 5% chance of an exceedance. To test that there is no correlation between exceedances and the level of risk we could divide our sample into high and low risk days. Just as before, we could test the significance of the number of exceedances in each subset using the binomial distribution.

End of Chapter Questions

1. Prove the formula for the first Cornish-Fisher moment, that is, given:

$$dV \approx \tilde{\Delta}R + \frac{1}{2}\tilde{\Gamma}R^2 + \theta dt$$

prove that:

$$\mu_{dV} = E[dV] = \frac{1}{2}\tilde{\Gamma}\sigma^2 + \theta dt$$

Remember that R is normally distributed with a mean of zero.

2. You are the risk manager for a currency trading desk. The desk had a VaR exceedance today. What is the most likely day for the next VaR exceedance?

3. You are asked to calculate the 1-day 95% VaR for a portfolio using the historical method with a window of 256 days. The table below contains the 20 worst backcast returns for the portfolio along with the time at which the returns, t = 0, 1, 2, … , 255. t=0 is the most recent return, and t=255 is the return form 255 days ago.

	t	R
1	42	-35%
2	83	-29%
3	10	-26%
4	23	-25%
5	3	-24%
6	58	-21%
7	188	-20%
8	103	-19%
9	131	-18%
10	12	-16%
11	116	-16%
12	245	-16%
13	150	-15%
14	56	-14%
15	61	-14%
16	31	-13%
17	69	-13%
18	95	-13%
19	161	-13%
20	35	-12%

4. Using the same data as in the previous question, calculate the 1-day 95% VaR using the hybrid method with a window of 256 days and a decay factor of 0.99.

References

Miller, Michael B. 2012. *Mathematics and Statistics for Financial Risk Management*. New York: John Wiley & Sons.

Answers to End of Chapter Questions

1. We take the approximation for *dV* as a given, as an equality, and take the expectations of both sides:

$$E[dV] = \widetilde{\Delta}E[R] + \frac{1}{2}\widetilde{\Gamma}E[R^2] + \theta dt$$

Because *R* has a mean of zero, E[*R*] = 0 and the variance is equal to E[*R²*]:

$$\sigma^2 = E[R^2] + E[R]^2 = E[R^2] + 0^2 = E[R^2]$$

Substituting back into our previous equation, we have:

$$E[dV] = \widetilde{\Delta}0 + \frac{1}{2}\widetilde{\Gamma}\sigma^2 + \theta dt = \frac{1}{2}\widetilde{\Gamma}\sigma^2 + \theta dt$$

2. The most likely day for the next exceedance is tomorrow. This is something of a trick question. It is tempting to guess that the next exceedance is equally likely to occur on any future day. In fact, if the probability of an exceedance on any given day is α, then the probability of an exceedance tomorrow is α, but the probability that the next exceedance is the day after tomorrow is (1− α)α. This is because in order for the *next* exceedance to be the day after tomorrow, two things have to happen: there is no exceedance tomorrow, and there is an exceedance the day after tomorrow. The probability of these two events are (1− α) and α, respectively, so the probability that both happen is (1− α)α. Because (1− α) is less than one, (1− α)α<α. For example is we are calculating the 95% VaR, then α = 5%. The probability that the next exceedance is tomorrow is 5%, and the probability that the next exceedance is the day after tomorrow is 5% × 95% = 4.75%. The probability continues to decline further out. The probability that the next exceedance happens in *n* day is, (1− α)$^{(n-1)}$α, which is likewise always less than α.

3. Five percent of 256 is 12.8, so the 95% VaR is between the 12th and 13th worst returns. Because we do not know the distribution between the 12th and 13th worst returns, we use the 12th. The 1-day 95% VaR is −16%, or a loss of 16%.

4. In order to calculate the hybrid VaR we need to calculate weights for the returns. For each return, the corresponding weight is 0.99^t. We can convert this to a percentage weight by dividing by the total weight, 92.37 = (1 − 0.99^{256})/(1 − 0.99). Starting with the worst return, we then sum these percentage weights to get the cumulative percentage weight until we reach 5%. This time, 5% occurs between the 7th and 8th worst returns. The 7th worst return, −20%, is our 1-day 95% VaR. Many of the worst returns have occurred recently, so it is not surprising that the hybrid VaR is worse than the historical VaR.

	t	R	wt	%wt	cum. %
1	42	-35%	0.6557	0.71%	0.71%
2	83	-29%	0.4342	0.47%	1.18%
3	10	-26%	0.9044	0.98%	2.16%
4	23	-25%	0.7936	0.86%	3.02%
5	3	-24%	0.9703	1.05%	4.07%
6	58	-21%	0.5583	0.60%	4.67%
7	188	-20%	0.1512	0.16%	4.84%
8	103	-19%	0.3552	0.38%	5.22%
9	131	-18%	0.2680	0.29%	
10	12	-16%	0.8864	0.96%	
11	116	-16%	0.3117	0.34%	
12	245	-16%	0.0852	0.09%	
13	150	-15%	0.2215	0.24%	
14	56	-14%	0.5696	0.62%	
15	61	-14%	0.5417	0.59%	
16	31	-13%	0.7323	0.79%	
17	69	-13%	0.4998	0.54%	
18	95	-13%	0.3849	0.42%	
19	161	-13%	0.1983	0.21%	
20	35	-12%	0.7034	0.76%	