**Quantitative Risk Management Tools for Practitioners**

In this chapter, we give an overview of the techniques many banks and investment firms use to satisfy regulatory capital requirements and internally manage their overall market risk. Our focus here is not how front desks hedge their positions, but rather methods aimed at protecting the firm from major losses from market moves or credit events over a given time horizon.

# Value at Risk - Introduction

We will start by defining the most basic measure of the risk of large losses, namely Value at Risk, or VaR. VaR, is roughly speaking, a measure of how much money a bank or other financial firm can lose on its positions in a fixed period, such as 1 day, 10 days, or 1 year in a “worst case” (e.g. worst 1 percent) scenario.

**Basic Assumptions:** The environment we work in is always defined by , a triple consisting of, respectively, (1) a *probability space,* or set of scenarios, (2) a -*field* or set of measurable sets, and finally (3) a *probability measure* on that -field*.*

**Definition 1.1.** Let V(t) be the value of your portfolio on a date t in the future. Let P be the *real-world* (NOT risk-neutral) measure, and let > 0 be a fixed *time horizon.* Let denote today. Then

For VaR as specified by Basel II, p = 0.99.

This measure of risk of loss has some drawbacks:

1. VaR tells you where the beginning of the tail is, and does not measure the overall tail impact on risk (Expected Shortfall, which will be covered later, is a stronger attempt to do that)
2. VaR does not consider liquidity – a 1-day VaR only considers changes in *mid market* price, and does not consider inability to sell at that price in extreme conditions. The concept of *liquidity horizon,* the minimum time it takes to unwind the position and get something close to the market value, tries to address this issue. For this reason we also will discuss 10-day VaR, which is part of Basel II capital requirements and accounts for the fact that it might take 10 days, rather than 1 day, to unload a position.
3. Because of illiquidity and because of potential model risk on the future realizations of portfolios, it is inaccurate, strictly speaking, to think of VaR as saying that the bank will not lose “this much” tomorrow with a 99 percent probability. It will be much worse if you try to actually unload the position, and your models may be off. An example of this will appear in the model risk chapter.
4. VaR does not always behave well when aggregating portfolios, meaning that the sum of VaRs for two portfolios is sometimes less than the VaR for the aggregate portfolio. It is quite possible for two portfolios to each have a VaR of less than $500,000, but the aggregate portfolio to have a VaR of greater than $1,000,000. In other words, the diversification principle can fail in the case of VaR.

**Example 1.2.** Suppose we have a single unhedged stock position that follows a real-world lognormal process

Then for a given value of p, we can compute VaR by solving

This translates in this case to

VaR.

In this notation, we are using the cumulative normal probability distribution

In particular, if the 1-day 99% VaR works out to $3.58, and the 10-day 99% VaR is about $10.77.

**Example 1.3. VaR based on factor model (variance-covariance estimate).** Suppose that there are a total of risk factors for a position with value V and their shifts are jointly normal with

Let the first order sensitivities to the risk factors be

Then, up to first order, the one-period loss is

and L has a normal distribution with mean and standard deviation , where

It then follows, using an analysis similar to that of Example 1.2, that

Note that this approximation is exact if the position value V is linear in all the risk factors, as in the case of a portfolio of stocks. However, a portfolio of derivatives has a non-linear dependence on all the risk factors, and its VaR will be discussed later in the upcoming analysis.

Overall, there are three basic calculation methods for VaR and each of them is discussed in greater length separately in the subsequent discussion:

1. **Analytical formulas or approximations** – rarely possible unless the underlying pricing models are extremely simple.
2. **Historical simulation** – last N 1 or 10-day interval market changes are applied to current conditions, and we take the loss corresponding to the 99th percentile. The Feds allow N as low as 251 (one year), with the second worst loss chosen as the VaR. Some banks use 2 or 3 years’ worth of data, with the VaR being the fifth or seventh worst loss in those cases.
3. **Monte Carlo model** – Create a parametric model for the next period’s moves based on current pricing models and sufficient historical data, simulate N times, take the least of worst 1% of losses.

We have just given examples of (1). We will now describe (2).

# Value at Risk – Historical Simulation and Loss Calculation

We assume that the firm has trading positions with values and that each position has risk factors chosen primarily from market observable inputs, rather than from underlying calibrated parameters (see Section 2.3 below for more discussion of this topic). Furthermore, each risk factor has historical time series and 1-day shifts (absolute) or (relative).

## 2.1 Full Revaluation

Now for each time t, and each position indexed by j, define

Then the total P&L across all positions is

Finally, sort all N (losses) one for each business day over the N-day historical period ending yesterday, from high to low, and take the one most closely corresponding to the 99th percentile. If using N=251, as federal regulators allow, take the second worst loss. Note that this is not really the 99th percentile of losses, really closer to the 99.2 percentile, but the regulators do not allow interpolating between the 2nd and 3rd worst P&L.

## 2.2 Delta-Gamma Approximation

The *delta-gamma approximation* to the daily historical P&L for a position is

This is just the 2nd order Taylor series approximation to (2), in contrast to the first order approximation of Example 1.3.

The purpose of this approximation is to potentially save time; the number of position values we need to compute for (2) is M(N+1), which can be quite large. In practice, when computing (4) it is common to leave out the cross gamma terms, which are usually (but not always) relatively small, and also the time derivative (theta) term at the beginning, to arrive at

The most common way to calculate these “Greeks” is to use central finite differences, with a *bump size* h, as in

For certain risk factors, such as stock price, it is common to use *relative* bump sizes, which means substituting for Observe that the total number of prices you need to calculate is now

This is usually a much smaller number than

The bump size can affect the accuracy of the delta-gamma approximation, and we can argue that we can achieve the smallest mean-squared error versus full revaluation by setting

To see this, consider the case of a single risk factor, and the absolute case. In that case we want

to be as close as possible, on average, to

By expanding the Taylor Series on both sides, we see that we want

From the last 2 terms we see that we really want

on average, which implies the result stated above. The argument is very similar for two or more risk factors, given that we do not include the cross gammas in the approximation.

## Market Data Inputs versus Calibrated Parameters

It is worth mentioning that the *risk drivers* for VaR, or the values are typically market inputs, e.g., spot, at-the-money volatility, some sort of volatility skew measure, and interest rates, rather than underlying model parameters. Full revaluation requires recalibration of the underlying parameters for every historical shift. Delta-Gamma approximation requires inverting a Jacobian matrix. More precisely, suppose that

Then

In this notation, an matrix such that the (i,k) entry is

Second derivatives are messier but can be worked out.

To give an example which illustrates this concept, consider a model for credit default swaps out to two years which depend on a one-year credit spread and a two year credit spread Suppose for simplicity that discount rates are 0, and the recovery rate is a constant R, and that coupons are paid with accrued interest up to the time of default. Then we can evaluate any credit spread out to 2 years if we know the underlying instantaneous piecewise-constant hazard rate for times and the forward instantaneous hazard rate for times

By assuming that a credit spread with tenors one or two years with par coupons respectively, have a net present value of 0, we can work out the following simple relationships between and the hazard rates

Then the Jacobian matrix, as discussed above, will have the following entries, with the first two being on the first row and the last two being on the last row:

Then, rather than bumping the credit spreads and backing out the hazard curve each time, we can instead bump the hazard rates and use the Jacobian formula shown above.

## Grid Approximation for Risk Factors

Another commonly used method of accounting for a risk factor is to replace the delta-gamma approximation for that risk factor with a one-dimensional grid. Clearly higher-dimensional grids are possible as well, but the number of NPVs to calculate rapidly grows with the dimension. It is common practice among the banks to combine a grid approximation for some risk factors with delta-gamma approximations for other risk factors.

If we wish to create a grid for the first risk factor, we create a symmetric array of the form with and compute NPVs of the form

for an *absolute* grid and

for a *relative* grid. In either case we might refer to these NPVs as

More generally, if we were using the kth risk factor, we would refer to these NPVs as

Relative grids are more common because the major risk factors for which banks employ a grid are typically spot price or volatility, which are always positive. For a relative grid in the risk factor , suppose that

Then we set

This is the P&L contribution from the first risk factor. Though 1-dimensional grid approximations still leave out the cross term risk, they are often more accurate than delta-gamma approximations for larger shifts since the latter is a parabola in the risk factor shifts. This grows much more rapidly than most pricing functions, in terms of spot, volatility or interest rate shifts. For example, a vanilla call option approaches linearity as spot increases, and approaches 0 as spot decreases. If we were computing VaR for an at-the-money call with strike equal to spot, the parabolic function given by the delta-gamma approximation would become quite large as spot approached 0, rather than going to 0, and as spot increased in the other direction, the parabolic function would grow much faster than a function approaching a straight line.

## One Day versus Ten Days

Federal regulators require that banks calculate VaR with both a 1-day and 10-day time horizon. For 10-day risk factor shifts, we let

or

If is the total number of historical returns we use, the returns begin on consecutive days, and overlap, meaning that we require days of data. If daily returns are i.i.d. (independent and identically distributed), then

In the case of an organization with limited computing power, the Feds may sometimes accept this approximation for 10-day VaR under certain conditions. See Section 5 of [1].

## Stressed VaR vs General VaR

In the case of stressed VaR, the calculation is exactly the same as for general VaR, except that the historical shifts are derived from a “stressful” period in the past, where the risk factors were likely to be much more volatile. In general, if we want N k-day intervals, we choose a set of business days in the past of the form

Then for a position with value we would calculate the loss

The argument 0 as usual stands for today or the most recent close date, but the difference here is that the last close date does not coincide with the last day of the history, namely as it does for general VaR. Typically for regulatory purposes and it is standard practice to choose the historical period so as to maximize the resulting total VaR for the bank’s entire trading book subject to regulation. One example of a very stressful period is the one-year period from April 2008 to April 2009. Once again, the Feds allow as little as a year’s worth of starting times, that is N = 251. For details on this, see [1], Sections 4 and 5.

## 2.7 Value at Risk – Backtesting versus Actual P&L

The purpose of *backtesting* is to determine whether a particular VaR model can be considered a good or reasonable measure of risk. More precisely, a VaR model is considered to be a good measure of risk if the actual loss in a particular 1-day or 10-day period does not exceed the VaR for the beginning of that period, more often than with probability Generally, if we are using a VaR calculation with we should not see more than around 2 to 3 *breaches,* or *exceptions,* per year. To formalize this concept, we note that, using the real-world probability measure we have

If you experience m exceptions over a period of m business days and this probability is less than 5% it is considered unlikely that the probability of an exception is 1-p. If m is too high, the VaR model is not adequately capturing the risk, and if m is too low (e.g. you never see an exception year after year), then you are *too conservative.* Generally speaking, since this results in an overestimate of regulatory capital, the federal regulators are never concerned if a bank is too conservative.

Here is an example in which and



Note that the Feds will start to have serious doubts about the VaR model if there are six or more exceptions in a year, and in fact, there are increasing capital requirements for four or more.

# Value at Risk – Monte Carlo Simulation

The last approach for computing the Value at Risk is to through a Monte Carlo simulation engine. Briefly speaking, Monte Carlo simulation works by creating a parametric model for the historical distribution, and simulating it a number of times which is typically much larger than the number of days covered by the history. This removes the primary limitation of historical simulation, that is, if you are only going back a year, you only have about 250 business days to work with. By contrast, Monte Carlo simulation enables us to use 10,000, 50,000 or even a million scenarios, if there is enough hardware available. Therefore, the limitation of Monte Carlo is in the quality of the model and computation time, not in the number of scenarios.

## Monte Carlo Basics

We assume that there are risk factors affecting your position, and we denote these by We assume that the 1-period changes in these risk factors have a cumulative probability distribution

As usual, we assume that the probability measure P is the *real-world ­*measure. We simulate this distribution times, and denote a particular scenario out of N by Now compute

Sort the losses in ascending order, and choose the loss at the correct percentile, i.e. if p=0.99 and N=10000, then choose the 9900th loss.

We assume for this exposition that the actual historical risk factors shifts, or *time series,* denoted by are independent and identically distributed (i.i.d.). The model we create is known as an *unconditional* model. A *conditional* model is based on a time series in which we assume that the distribution of can depend on the values of for There is a great deal of literature on conditional models, (see [1]) but these are beyond the scope of this brief overview.

## Monte Carlo VaR for Beginners

The simplest Monte Carlo algorithm involves assuming that the distribution of the risk factor shifts is joint normal. Specifically, we express a multivariate normal distribution in vector form as

This distribution is joint normal with means and covariance matrix and we write this as

To simulate this distribution, we first make use of the

*Cholesky decomposition*: Suppose that **X** is d-dimensional with . Suppose, further that has full rank d. Then you can write the covariance matrix as

with **A** being *lower triangular (d x d)* with positive diagonal entries.

One then follows the following four steps for a 99% VaR with 10,000 scenarios:

1. Compute the historical means over the last year, or several years, denoted by and their historical standard variances, denoted by .

2. Compute the historical covariances rounding out the rest of covariance matrix, denoted by

3. Use Cholesky decomposition to simulate as multivariate Normal 10000 times and each time compute by delta/gamma approximation, grids or full revaluation, if you have enough machines.

4. Choose the 100th worst of these losses for VaR and take the average of the 100 worst for ES, for p = 0.99.

Well, of course, the joint distribution of changes of n risk factors for a typical asset class would not be multivariate normal – they usually have *fat tails –* in other words the risk of an extreme move is far greater than what would be implied by a normal distribution. In fact, this “beginner” method would never pass muster with the regulators. We will now look at methods of creating more realistic models for Monte Carlo VaR.

## Some Important Basic Statistics

*Theoretical skewness, kurtosis*. Let be the standard deviation of a probability distribution of a random variable X, and let be the mean. Then skewness and kurtosis are respectively,

If X is a normal random variable, and

*Sample mean and covariance matrix:*  Suppose we have n samples of the d-dimensional random variable **X**, namely

*Then the sample mean and the covariance matrix are:*

It is known that using 1/(n-1) instead of 1/n makes the covariance estimator *unbiased*.

*Sample skewness and kurtosis:* Let be n samples of the scalar random variable X, and let be the sample mean. Now let

Why do skew and kurtosis matter? Kurtosis is a measure of tail risk. If that means that the tails are “fat”, and computing VaR using a normal distribution with the same mean and variance will likely understate the risk. Stock and FX price returns are typically “leptokurtic” (fat-tailed). Skew is a measure of asymmetry about the mean. Stock price returns are negatively skewed, with big losses more likely than big gains. Our next goal is to go beyond “beginner” and create joint distributions with the right covariances and whose marginal distributions are skewed and fat-tailed.

## Some Important Tests

As mentioned in section 3.2, the Monte Carlo simulation for “beginner” is to employ a joint normal distribution. A natural question is: *Are the marginal distributions normal?* The *Jarque-Bera* test indicates whether a 1-dimensional distribution is likely to be normal. Using the sample kurtosis and skew above, we define

As the sample size n gets larger, if the distribution is normal, T will approach a chi-squared distribution with 2 degrees of freedom, which turns out to be exponential:

for large values of n.

*How well does your model fit the data?* Let be n samples of an unknown d-dimensional distribution, and suppose we want to fit a distribution with a density of the form

is a set of parameters. Since we assume that the samples are independent, we want to maximize, over all possible choices of the parameters, the log of the likelihood of seeing these samples, namely

We call this the *maximum log likelihood* calculation.

## A Simple Type of Fat-Tailed Distribution

Let

Next we require that W be independent of the Z’s. Then the random vector

is said to be a *normal mixture* model. The simplest example of this sort of model would be the one-dimensional case

Observe that, for this one-dimensional case,

In the special case where we get

It follows immediately that the marginal distributions of a normal mixture *always* have fat tails. We can then consider a 2-dimensional example, expressed as

with Then note also that This means that we can use normal mixtures to add kurtosis without changing the correlation matrix. Also note that because of independence of and and more generally

Note, however, that there is only one kurtosis level.

To simulate such a fat-tailed distribution of **X**, it is enough to generate instances of the d-dimensional normal distribution and also the single scalar random variable In the case where is absolutely continuous with a density function and cumulative distribution function , a simple (even though it is not always the most efficient) way to simulate is to choose a uniform random variable and set

To compute the density of this distribution is straightforward. Note that if is the domain of , then

## Special Case: the t-Distribution

To illustrate the aforementioned idea, we focus on only one simple example of a fat-tailed distribution that we can try to fit to the historical data for d risk factor shifts. Consider the following distribution for , known as the *inverse Gamma distribution, or :*

Then the d-dimensional normal mixture random vector with has what is known as the *multivariate t distribution* with degrees of freedom. From the formula in Section 3.5, it is possible to derive the density in closed form (see [1], Section 3.2) as

This is a popular fat-tailed distribution for fitting certain common types of risk factor sets, such as portfolios of stocks.

To choose a suitable t-distribution for our data, one simple and intuitive approach is to match the mean and covariance matrix and choose the parameter which yields the maximum log likelihood. First, we express the multivariate t distribution as

Here, for notational simplicity, we are using the notation to refer to risk factor shifts which we would ordinarily refer to as To match the covariance matrix, let be the sample mean of , let be the sample covariance matrix of and let be the covariance of **.** Then we want to have

The density function now becomes

which implies that the log likelihood we want to maximize is now

There are any number of simple numerical algorithms which can maximize this expression, and the result will be a distribution with the same means and covariances as your sample, but with the optimally chosen kurtosis for best fit to the data.

## Fitting a Distribution with the E-M Algorithm

Another way to fit a t distribution, or any other distribution of this type, is with an iterative procedure known as the *E-M(expectation-maximization) algorithm.* In this case we are not insisting that we match the sample covariance matrix exactly first, but as a result we might obtain a better overall fit based on maximum likelihood.

The way to think of the E-M algorithm is that it “seesaws” between estimating the multidimensional parameters and the parameters of the distribution of W. The basic tasks are:

1. Express the joint density of (**X**,W) as the product of the density of W and the density of **X**|W.
2. Estimate the parameters based on the latest estimates of your W parameters and the known values of the samples
3. Then do maximum log likelihood to get the parameters of W using the density function *h* but you don’t have the so instead you use expectations of certain functions of the which in turn were derived from the latest and the distribution of given those parameters.
4. Keep doing (A) and (B) until you achieve convergence.

We will now present a more precise, numbered, set of steps for the t distribution. First note that in the case of the t distribution, has only one parameter and by Bayes’ Theorem we can express

so that

Thus the conditional distribution of W given **X** is Inverse Gamma with parameters that is, with

In addition, the log likelihood of the overall density breaks down as

Armed with this important information, we can now give an exact recipe for the algorithm.

Step 1: Set

the sample means and covariances, and let be some “reasonable” first guess for For notational conciseness, let Let be the iteration counter, and set that equal to 1.

Step 2: Calculate the following:

Step 3: Let

Step 4: Define an intermediate set of parameters . Then let

Step 5: In the equation

substitute for and for Now maximize function over all possible values of , to obtain .

Now let , replace k by k+1, go back to Step 2, and repeat Steps 2-5 until you achieve convergence.

This algorithm can be generalized to any multivariate distribution which takes the form of a normal mixture. The only difference might be that if the density has more than one parameter, then Step 5 will be more complex, and the conditional distribution may be more challenging to work out, but other than that the algorithm would be the same.

## Expected Shortfall

Within a few short years, it will no longer be acceptable to regulators for banks to use general and stressed VaR as key risk management tools; they will be required to replace it with a variant known as *Expected Shortfall,* or ES for short.

To understand what ES is, think of VaR as a high percentile of the possible losses; ES, on the other hand, is the *average* of the tail beyond a certain percentile. More rigorously, define

To define in terms of a percentile, we may write, for some probability ,

If a bank uses for VaR, it is likely to use a somewhat lower probability, such as for ES, and in fact upcoming federal regulations will specify this level (See Section D3 of [2]). Though this risk measure is harder to do backtesting for than VaR, it has one key advantage over VaR, which is that it satisfies *subadditivity,* meaning that if you have two portfolios and , then

As mentioned above, VaR is not guaranteed to satisfy this property, and it’s easy to come up with an example. Suppose that on October 1, 2015, you are doing 1-year historical VaR, using the second worst loss, and has the two worst losses of $110 and $150 corresponding to historical shifts that occurred on January 14 and May 1. On the other hand, had its two worst losses on the exact same two historical shift days, but those losses were $130 and $100, respectively. Then the aggregate portfolio must have its worst two losses on those exact days of $240 and $250, giving us a VaR for of $240. But note that

Since ES is the average of the tail, it is also considered a better measure of tail risk since it contains information about the extreme outliers that you might miss with VaR.

# Stress Testing

A somewhat simpler, but equally important aspect of risk management is the concept of *stress testing.* A stress scenario consists of applying a single extreme set of shocks to the current values of bank’s risk factors, and computing the change in net present value that results. Stress scenarios take 2 forms, *business as usual* and CCAR, or Comprehensive Capital Analysis and Review, which is the regulators’ annual review of all the major banks’ risk management practices.

A business as usual (BAU) stress scenario is a choice of two dates in the past, For a position indexed by we compute

On the other hand, in the case of a CCAR stress scenario, the regulators decide on a set of fixed shift amounts, so that

An example of a CCAR stress scenario might be one in which the regulators instruct the bank to increase all equity volatilities by 30 percent on a relative basis, and decrease all stock prices by 20 percent on a relative basis. In both BAU and CCAR stress scenarios, the bank may need to adjust the modified market data so that there is no resulting arbitrage, and the positions can price successfully. In that case the realized risk factor shifts may be different from the original prescribed shifts.

The most difficult aspect of stress testing is defining what scenarios to use. In the case of BAU, this means choosing the date intervals The concept of what is a “good” stress scenario is an extremely ill-defined problem and the subject of much current research. Some examples of stress scenarios which a bank might use are:

*Financial crisis, 4th quarter 2008.*

*September 10, 2001 to a couple of weeks later (9/11 terrorist attack)*

*Subprime crisis, from around February 2007 to around August 2007*

*U.S. credit downgrade, August 2011.*

**References**

1. *Federal Register, vol 77, No. 169, Rules and Regulations.* Office of the Comptroller of the Currency, August 2012.
2. *Instructions:**Impact Study on the Proposed Frameworks for Market Risk and CVA Risk.* Basel Committee on Banking Supervision, July 2015.