



Aaron Brown

# Forced by the Sternest Circumstances

*The American Constitution, one of the few modern political documents drawn up by men who were forced by the sternest circumstances to think out what they really had to face instead of chopping logic in a university classroom*  
George Bernard Shaw, *Getting Married*

In every field there is some discrepancy between what practitioners do and what textbooks teach. This is proper; only vocational education attempts to instill specific practical skills. Textbooks should provide a higher-level framework that will not only prepare students for their chosen field, but also give them the tools to change it for the better. On the other hand, when the discrepancy gets too big, one or both sides should move to narrow it.

I have in mind a simple, crucial quantitative fact that is second nature to any practicing financial risk manager, which I have never seen in a risk management textbook. It has to do with Value at Risk (VaR). There is a 1 percent chance your trading positions will lose more than VaR over one trading day, if you do not change positions and markets are normal (other confidence levels and horizons can be used as well; 5 percent and ten days are common).

To be useful, a VaR estimate must be very close to the true probability, taking into account all available information. More concretely, no one should be able to make money betting for or against VaR breaks (days when you lose more than VaR) at 99 to 1 odds. In fact, any quant managing market risk in the late 1980s or early 1990s ran a book on some type of VaR-like risk measure. You can imagine what traders would think of a guy who wanted to tell them how to run their billion-dollar book, but wouldn't bet \$1,000 of his own money that he was right. Only after a few months of not being able to beat a game offered at zero spread did traders grudgingly concede that the VaR might not be wildly overstated. And their attempts to predict VaR breaks better than the risk manager gave him useful tips.

In the executive suite, betting was frowned



upon, but people insisted on rigorous statistical analysis before they would entertain the idea that VaR was not ridiculously understated. One of the triumphs of VaR is that it made sense on both the trading floor ("it's like a point spread") and the corner office ("it's like an actuarial prediction").

## Passing grades

The three most important tests for a VaR methodology are:

1. The actual fraction of VaR break days is 1 percent, within statistical tolerance;
2. The VaR breaks are randomly distributed in time;
3. The VaR breaks are independent of the level of VaR.

One simple way you might think to use is to look at how the current portfolio would have done over past days, and pick the 1 percent

loss point. This is a distribution-free approach. Regardless of the distribution of price changes, as long as it is constant and outcomes on different days are independent, the probability of an observation being among the five worst in the last 500 days is 1 percent; therefore, the chance of it being worse than the fifth worst of the last 499 days (which is the equivalent event) is 1 percent. This should guarantee that all three tests above are met. This method shows up in all risk management texts as "historical simulation VaR."

So, let's take the simplest possible application. You have a portfolio that consists only of \$1 million invested in the S&P 500 Index. You use a three-year window for your VaR estimate. Since the beginning of 1930, you would have made 19,922 VaR predictions, so you should have had 199 breaks. In fact, you had 314 breaks – more than 50 percent too many. The chance of this being random bad luck is essentially zero. Three

years is the most common window used for historical simulation VaR, and you don't improve things significantly by choosing a shorter or longer window. Historical simulation VaR is a handy number to know, mostly to compare it to the real VaR, but it is not a VaR.

Should we try for two out of three? If the breaks were randomly distributed in time, 1 percent of them (3.14 breaks) would occur the day after another VaR break. In fact, 23 of the breaks came the day after a break. Ten percent, or 31.4, should occur within ten days of another VaR break. The actual number is 136. But if it's been more than 25 days since the last VaR break, your chance of a break drops far below these levels.

The third test concerns the level of VaR. The average VaR is \$26,500; the average VaR on break days is \$24,100. This is not a huge failure, like the first two tests, but the deviation is too large to be random chance. You get your breaks disproportionately when you say things are safe. The flip side of that is bad too – when you say VaR is high, you are usually exaggerating the danger.

You might try to improve things by using a lower confidence interval. If you set VaR to the 0.56 percentile instead of the 1.00 percentile, you get exactly 199 breaks and pass the first test. One problem with this is that the appropriate percentile is not constant. If you set it using recent past data, you will not get the correct number of future breaks.

More importantly, this does not help the second and third tests. Now you should have two VaR breaks the day after a break, and 20 within ten days of a break. You actually get eight and 80, respectively. Your average VaR estimate is \$31,200, but your average VaR on break days is \$27,300.

This is not a function of the S&P 500 or the time period chosen. Historical simulation VaR does not work on financial time series, and cannot be made to work. Volatility clusters. Your observations are not all drawn from the same distribution. You will lose money betting if you rely on it. Your VaR will be too low in good times, when you should be reminding people that good times always end, and too high in bad times, when you should be pushing people to take sensible risks. You will induce procyclical behavior

in your institutions, helping to inflate bubbles and exacerbate crashes. No one has ever passed a nonrubber-stamp backtest with historical simulation. Since this is an easy observation to make, and important for a risk manager to know, you would think it would work its way into someone's textbook.

### When in doubt, parameterize!

The only other major flavor of VaR I see in textbooks is parametric VaR. The most common estimation parameter is the original Riskmetrics 94 percent decay exponentially weighted moving average. So, you take yesterday's variance estimate, multiply it by 0.94, add 0.06 times today's move squared, and set VaR at 2.33 (the one percentile point of a Gaussian distribution is  $-2.33$ )

## Many people's first reaction is that this is cheating. It's easy to get the right number of VaR breaks this way

times the square root of that number.

That will get you a whopping 1,298 VaR breaks – six and a half times the expected 199. Instead of 13 of them coming one day after another break, you get 154. Instead of 130 coming within ten days of a VaR break, you get 711. Your average VaR is \$15,000 and your average VaR on break days is \$13,800.

Okay, you know financial return series have fat tails. If you take 7.63 standard deviations instead of 2.33, you get exactly 199 breaks. As with historical simulation, the problem is that this number is not constant, so if you estimate it from the past, it will not be reliable in the future.

Unlike historical simulation VaR, parametric VaR improves on test 2 when you fix test 1. You expect two VaR breaks the day after a break, and you get eight. But once you get four days beyond a VaR break (during which you get 19 breaks instead of the expected eight), the frequency of breaks drops to about the right level. Unfortunately, test 3 gets worse. Your average VaR is \$27,200 and your average VaR on break days is \$22,300.

Of course, you can try higher or lower decay factors, or use fixed intervals or some method other than exponentially weighted moving average to estimate standard deviation, but it won't get you past these tests. Simple parametric VaRs do not work.

### A better way

If there were no good alternative to these methods, you could understand them being taught without dwelling on their defects. But there is a perfectly simple and obvious way to estimate VaR that works quite well. In one form or another, often deeply disguised, it is at the heart of all successful VaR systems.

You start with any simple VaR estimate. For this example, I'll use 2.33 standard deviations

estimated over the previous three years. If you get a VaR break, you double the estimate. If you don't get a VaR break, you take 0.94 times yesterday's estimate plus 0.06 times your updated simple estimate.

Many people's first reaction is that this is cheating. It's easy to get the right number of VaR breaks this way. For example, you could raise VaR to infinity after a break, then bring it down to negative infinity exactly 100 days later. But if you did this, you would fail test 2 and have the worst possible performance on test 3.

Another objection is that you have three parameters: the estimation interval for the standard deviation, the amount to increase after a VaR break, and the decay factor. I don't count the 2.33 standard deviations for VaR, since this comes from the Gaussian distribution. Parametric VaR has only one (the estimation interval), two if we fudge the percentile. Parametric VaR is the same. This would be a problem if the parameters were crucial to the performance, but they're not. This works well for a wide variety of reasonable parameter choices,

and, more importantly, good parameter sets do not seem to change over time. The volatility of financial time series changes, but the characteristics of volatility bursts seem to be reasonably stable. Also, you'll note I didn't overfit here, I just used the same parameters from the previous two methods – three-year estimation interval and 0.94 decay factor.

How does this simple procedure stack up? It gives 188 VaR breaks – less than one standard deviation away from 199 – so perfectly consistent with a 1 percent probability of break. It has one, two, or three VaR breaks for each of the first ten days after another VaR break, except zero, nine days away, and 15 in total over the first ten days. This is perfectly consistent with breaks distributed randomly in time. The average VaR is \$25,100, and the average VaR on break days is \$25,700 – again, consistent with random distribution.

Two other aspects of test 3 are notable. The first is that the average VaR is the smallest of the three methods, when the first two are calibrated to give exactly 199 breaks, and it has had fewer breaks. That makes it more aggressive in allowing risk, with fewer misses. The second is that the VaR on break days is higher than the average VaR (although within the expected variation). This is the better side on which to err. If you overestimate risk before a big loss and underestimate it when there is no loss, you might even have better results than a perfect VaR would give. Finally, it shows that we're not cheating. If we lowered VaR until we got a break, then raised it to get the right number of nonbreak days, we would have much lower VaR on break days than nonbreak days (we'd also have too few breaks immediately after a break).

### **A VaR by any other name...**

Every VaR that passes a nonrubber-stamp backtest has a mechanism for sharp VaR increases after large market moves, with a following decay. I like simple methods, so I often use exactly what I described, choosing some parameters that seem to work pretty well. Other people like GARCH, or even more complicated models. Another group uses a lot of math to solve for an optimal Bayesian regime shift estimate (when asked what I call my simple method, I answer "Bayesian VaR" because it represents my

subjective posterior belief, as evidenced by the fact that I'll take either side of a real money bet on it). A different sort of risk manager brings in the spikes by using market data, like implied volatilities. Then there are people who use my simple version but are ashamed of it – they call it something like "fat tail scaling algorithm" or "volatility override."

In my experience, successful quantitative risk managers are agnostic pragmatists and use any combination of the above that seems to work, and satisfies the people they work with. If the CEO likes GARCH or implied volatilities, I can work with those. If the head trader likes a simple rule like "double after a break," I won't object. It's even possible without too much work to accommodate both at the same time.

The point is that every risk manager responsible for producing a daily VaR and held to a non-rubber-stamp backtest uses this method. It has been rediscovered independently many, many times. All you have to do is try to predict for a while and you'll find yourself doing it. Yet it is not in any textbook I have ever seen.

To put this in context, it's important to remember that estimating the distribution of underlying market parameters is just the first and easiest step to developing a VaR system. Next, you have to estimate all market factors at once. Historical simulation fails here, because the past is too sparse. Parametric methods fail because you can't possibly estimate all the covariances, and covariance does not capture the full dependence effect. Every day will be a surprise if you try either of these methods.

What does work pretty well is to use any simple prediction method – say, a four-factor linear regression model – but inflate the volatility of any factor that is more than 2.33 standard deviations from prediction. Let the inflation decay for market factors that did not have VaR breaks. It's possible to have a successful VaR without anything more sophisticated than this, although you can do better. It's not possible to have a successful VaR without some sort of volatility inflation.

Next, you have to gather all the positions. This is a tremendous challenge in most institutions, and you never get things perfect. You need to build in good data quality checks to identify

missing or erroneous data and substitute good default values. Another fact that I don't find in textbooks is that a significant component of VaR is position uncertainties and other data errors. To publish a VaR daily that can pass a backtest without subsequent adjustment (which is contrary to the entire idea of risk management; if you publish a number for people to use in decision-making, after-the-fact adjustments are irrelevant), you have to add cushions for data and system problems. In a real VaR system, more time and effort is spent finding clever ways to generate accurate results from inaccurate inputs than in financial modeling or statistical reasoning.

Finally, you have to price all the positions. For something like an equity, it's easy – the market factor is the price. But for derivatives, pricing can be an enormous challenge. Front-office models are often the wrong tool. A VaR simulation may price all the firm positions under thousands of scenarios; pricing models that involve their own simulations or other computationally intensive procedures are impractical to include. Also, front-office models sometimes contain many parameters that can only be calibrated from market data; no one has any idea how to predict their future evolution. Finally, front-office models have to be very precise because they're used for real trades and real hedges, but it's acceptable if they break every once in a while (the front office can just throw it out and build a new one; every desk quant sooner or later finds herself building a model while people are trading on the outputs). A risk manager needs a more robust model and can sacrifice a lot of precision.

When you contemplate this daunting task, you need confidence that your basic underlyings, the marginal distributions of individual market factors, have the right statistical properties. This is one of your few reliable reference points in a chaotic sea. Historical simulation and simple parametric models are not suited to this demanding environment. Textbooks are where they belong, but they should come with a warning label "for toy problems and indicative reports only." And perhaps someday, someone who has delivered successful VaRs for many years will write a textbook and tell us what he really did, instead of what he thinks he should have done.