RiskMetrics Group

Research Monthly

VaR Is From Mars, Capital Is From Venus

April 2009

Christopher C. Finger
chris.finger@riskmetrics.com

In February, we commented on a Basel Committee proposal to implement a capital charge (the Incremental Risk Charge, or IRC) to cover default and migration losses in the trading book. Last month, the Financial Services Authority (FSA) of the UK released the Turner Review, an analysis of the origins of our financial crisis and recommendations for the future of financial services regulation, along with a discussion paper stating a more official FSA policy stance. Unlike the detailed proposal for the IRC, the FSA documents represent a broader review of regulation and capital generally.

Neither FSA document is guarded in its assessment of financial supervisors in the runup to the crisis. The discussion paper summarizes:

> . . . prudential regulation was too weak at a micro-prudential level, and almost completely lacking at a macro-prudential level. (Paragraph 4.2)

Central to the calls for macro-prudential regulation is the reaffirmation of the supervisors' role in ensuring the stability of the financial system. The papers propose that regulators take an active role in identifying systemic risks—easing of credit standards, rising systemic leverage, falling risk premia—through both macroeconomic analysis and knowledge of the actions of specific institutions. This is an ambitious step, broadening the mandate for regulation generally, but also requiring cooperation between the different agencies that today perform macroeconomic and institution specific analyses.

For the micro-prudential level—the supervision of individual institutions—the strongest call is for a "fundamental review" of the existing minimum capital rules. The tone of the papers make clear that the FSA holds strong views on where this review should take us. As for Value-at-Risk (VaR) models, the FSA states that the burden of proof lies with the industry to show that VaR is appropriate for capital rules.

But such a challenge sets up the review as a (potentially one-sided) discussion focused on the merits (or demerits) of VaR. This is far less productive than starting with a review of what trading book capital should be, with the notion that if a short-term measure of risk is appropriate, then VaR should prove its place. Our aim in this note, then, is to address some of the wrong reasons the FSA has put forth to cast VaR aside, offer some good reasons of our own, and comment on where short-term risk measures could contribute in a future trading book capital framework.

What went wrong?

Both FSA documents begin with the same chapter entitled "What went wrong?" The chapter reviews

the macroeconomic environment and the growth of the securitized credit model, and then moves to the now obligatory section on "Misplaced reliance on sophisticated maths". Central to this section is an exhibit (Box 1A in the Turner Review) summarizing in three main points the FSA views on the deficiencies of VaR-based estimates of risk.

One of the three VaR deficiencies is the "failure to capture fat-tail risks", which is further elaborated:

> Short-term observation periods plus assumption of normal distribution can lead to large underestimation of probability of extreme loss events.

While the symptom here (extreme loss events were underestimated) is uncontroversial, both aspects of the diagnosis are misguided.

To be clear, VaR models are intended to forecast (in a statistical sense) the possible loss on today's portfolio, based on today's market, over a short (one day to one month, for instance) risk horizon. Almost always, these models rely on historical price moves and some statistical inference to perform forecasting. In regulatory language, VaR by definition is a point-in-time measure.

One implication is that VaR models are verifiable: we can track both our forecasts of possible loss and the actual loss experience over time, and validate whether these two have the statistical relationship that they should. Banks perform this backtesting as part of their regulatory and disclosure requirements. Among the standard metrics is the number of VaR exceptions, that is, the number of days on which actual trading losses exceeded the VaR forecast.

Campbell (2009) recently surveyed bank VaR disclosure for 2008. Even in such an eventful year, there was a wide range of model performance. One of the better performers, Bank of America, reported two VaR exceptions at 99% confidence, almost precisely what one would expect. This was an improvement on the bank model's performance in 2007, which the bank attributed to their move to more frequent data updates in an effort to react more rapidly to higher volatility. In fact, this move to more reactive volatility measures was one of the best practices cited by the Senior Supervisors Group (SSG) in March 2008. This example also shows that good backtesting results do not guarantee that a bank had a peaceful year.

At the other end of the spectrum was UBS, which recorded 50 exceptions at 99% confidence. In other words, UBS observed what they thought was a one-in-a-hundred event on average once per week, a spectacular failure in risk forecasting. One of the bank's comments on this poor performance was to assert that their results "highlight the limitations of VaR". Such a broad dismissal of risk forecasting based is especially feeble in light of other banks' success.

A closer look at the UBS disclosure is illuminating. In its annual report, the bank discloses that it utilizes five years of historical returns, with equal weighting, to produce its VaR forecasts. It warns that this method "does not respond quickly to periods of heightened volatility". Indeed.

Thus, consistent with the SSG's recommendations, but at odds with the FSA evaluation, it was *long* observation periods that led to underestimation of risks. In the short term, risk changes, and risk models must react. The use of short[1] observation periods for VaR

[1] Or effectively short, in the case that past data is weighted unequally

forecasting is necessary to make good forecasts, and not just a convenient choice that lets us save money on data storage.

False relics

This brings us to the second part of the diagnosis—the use of the infamous normal distribution—which has been fodder for a thousand popular press vilifications of risk models. While it is true that the normal distribution is an oversimplification of empirical loss experience, the continued burning in effigy of Karl Friedrich Gauss has distracted the dialogue from numerous other points. And while we can ignore the counterproductive dialogue in the tabloid press or chat forums, the regulatory dialogue is more important. Unfortunately, to be heard above the din, a bit of yelling is in order.

First, any discussion of probability distribution (the description of our ignorance) must come only after using the available information to forecast what we can, in this case volatility. If asked to propose a distribution of the heights of a group of school children, it would be absurd to start deriving tail exponents before asking how old the children were.

Second, criticisms of the normal distribution are not new, and there exist numerous VaR model implementations that use alternatives. Not all of these are improvements, however. Those that use non-normal distributions but do not forecast changes in volatility perform categorically worse than those that stick with the beleaguered normal distribution but react appropriately to changing market conditions.

Third, the search for the right fat-tailed distribution at best gives us a more accurate view of the static part of our risks. As such, it is in some ways equivalent to shrugging our shoulders and saying "Shit happens". Maybe we recognize that market moves of ten times our volatility forecast are significantly more likely than we thought, not one-in-a-billion events but one-in-a-thousand events. But this teaches us nothing about the dynamics of the market we are trading, and gives us no warning signals of when those events are becoming even more plausible. The plea for the right fat-tailed distribution is the medieval response to the Plague, not to understand hygiene, germs or contagion, but to blame misfortune on the unknowable and immeasurable. This reaction is dangerous, as it presupposes not only that we don't know, but that we can't know. Our only hope is to trust in someone endowed with divine knowledge of the unknowably improbable.

Bigger distractions

The most dangerous distraction then is from an Enlightenment response to the Plague—a real analysis of what could have made underestimations of risk better, not just by making them larger, but by making them more timely. There are two crucial areas which have been all too often ignored. To their credit, the FSA does focus on one of these.

The second of the VaR deficiencies is "Failure to capture systemic risk", in which the FSA cites the assumption that "each institution is an individual agent whose actions do not themselves affect the market", and asserts that "interconnected market events can produce self-reinforcing cycles which models do not capture". This is potentially a much more productive

discussion than that of fat-tailed distributions, as it leads us away from a view purely based on historical returns data, and holds the promise of more than just larger risk estimates overall, but more timely signals of when risk has increased.

In fact, the language here lets us resolve one of the damaging rhetorical corners that the risk community has painted itself into. It is common to state that VaR models work in "normal markets", but without ever defining what "normal markets" are. The lack of a good definition turns what could be a useful guidepost into a useless circular statement: models work in normal markets, and normal markets are defined as those within which models work.

A better definition of a normal market is one in which the assumption mentioned above holds: institutions' actions do not themselves affect the market, and all market participants are purely victims of a set of exogenous price processes. In this regime, it is reasonable to expect that the historical data is sufficient as a basis for forecasting, and that VaR models will provide timely and accurate indicators of risk.[2]

To illustrate what we would call a non-normal market, consider the dislocation in tranched credit derivatives in 2005. The first two tranches of the North American credit derivatives index (CDX) typically exhibit a very tight relationship, with correlations in the 80-90% range. There is a fundamental reason for this, as the two tranches represent protection on the same underlying portfolio, albeit against different levels of loss. In the first two weeks of May 2005, however, the tranches moved significantly against each other, without the price of the underlying index moving much, causing heavy losses for investors with (seemingly) hedged positions across the two tranches.

This was a case where no model based on historical data would have foreseen the losses to come. Though a savvy risk manager would have known that the relationship between the two tranches was not perfect, and that such a dislocation was possible in theory, no historical precedent existed for the magnitude of the losses. This is not, however, a case where we shrug our shoulders and mumble about fat tails.

In the aftermath, it became clear that there had been a large buildup in positions that were short protection on the first tranche and long protection on the second. In such a position, a trader would reap a net quarterly premium and be hedged against moves in the underlying portfolio, or so it seemed. But with many investors in the same trade, the inevitable happened. Sparked by an event external to the tranche market (likely the earlier downgrades of Ford and GM), an initial set of investors closed their positions, pushing the prices of the two tranches apart and causing mark-to-market losses for those still holding the position; this sparked more position closing, which led to greater losses, and so on. In the end, this was a classic example of a crowded trade.[3]

But with no historical precedent, was there anything risk managers could do? Possibly. Market makers in these derivatives, themselves advocates for the "hedged" trade, could see the market flows leading to the crowded trade, prompting perhaps an investi-

[2] To be clear, we use the word normal here in its generic sense, meaning typical or regular. Market returns could well be non-normally distributed; all we are claiming is that useful statistical inference from historical data is possible.

[3] See Finger (2005) for more detailed analysis.

gation of their own positioning and susceptibility to a rush-to-the-exits scenario. Outside of flow desks, in 2005, there may have been nothing but market rumors to provide such insight. Today, however, the situation has improved.

The Depository Trust Clearing Corporation (DTCC) began in the fall of 2008 to provide weekly snapshots of the market exposures in credit default swaps, credit indices and tranches. For some of these products, the reports are granular enough to note changes in exposure on individual contracts. For tranches, this is not yet the case, but it is not far fetched to believe the members of the DTCC will agree to make this level of detail available in the future. Notably, this is true for an over-the-counter market which does not (yet) operate with a central counterparty.

In short, the FSA is correct to criticize the current generation of VaR models for their inability to uncover systemic risk or contagion effects. The data and tools to uncover such effects are closer than we might think, and should be the subject of significant attention in any regulatory review.

A new backtest

One year ago, we wrote in praise of the SSG for recognizing the description of positions in a risk model as just as crucial as the volatility model or distributional assumption. Neither FSA document makes enough mention of this point. This is the second casualty of the fat-tailed distraction.

The most overlooked source of bad VaR forecasts is the failure to adequately describe trading positions. This can take the form of a missing risk factor—assuming the basis between two similar instruments or the spread on a risky bond is constant—or a poor proxy choice—utilizing corporate bond yields to describe the risk of a securitization with comparable rating. As the SSG pointed out in early 2008, and as continued to be the case, it was modeling the wrong (or no) risk factors that was the root of the worst understatements of risk, not the choice of the wrong statistical distribution. The lack of coverage of credit risk in the trading book—what the IRC proposes to address—is a version of this same problem.

There are statistical arguments to tell us how many exceptions we should see in theory, assuming VaR models work as advertised. And there is a history of disclosure on simple, standard backtesting measures. We have a sense for how many exceptions the industry experiences, and know for instance that at 99% confidence, two VaR exceptions in a year is expected, eight is fair but slightly concerning and fifty is outrageous.

We propose that the description of instruments for risk purposes be backtested in their own right, independently of the statistical models used to forecast changes, and that the risk community establish a set of simple benchmarks for these tests. One candidate for such a benchmark is to compare over time the actual market price changes on an instrument to the price changes that would appear in the risk model, with whatever assumptions (proxy factors, constant spreads, linear price relationships) that might be entailed. Simple correlations of these changes, averaged across asset classes, would provide a first indicator of the quality of instrument representation.

Of course, this sort of backtesting is only possible if

we have actual market prices to compare. The absence of such prices, and the inability to perform the proposed backtesting, could be used as criteria to exclude instruments from a model-based capital framework, or from regulatory trading book consideration.

This is essentially the argument used (lack of prices, unclear pricing models and risk factors) for excluding securitizations from trading book consideration, both in the FSA's proposed capital review and the Basel Committee's IRC proposal. We are in agreement with this specific decision at the present time, but recommend that regulators and industry formalize what is expected for securitizations, or any instrument, to earn trading book treatment in the future.

Moving to procylicality

The last of the VaR deficiencies is that the procyclicality of VaR contributed to excessive risk taking in the period prior to the crisis. A procyclical capital framework is one that reinforces business cycles, requiring less capital when times appear good—encouraging greater risk taking—and more capital as the economy contracts—constraining banks' ability to lend and working against economic growth. Alongside this point, there is a demonstration that VaR based on short histories will produce such procyclical capital requirements. The solution is to use longer historical periods, exactly as UBS did. Once again, the symptom is correct, but not the diagnosis.

The conflict between the needs of capital and the output of VaR has existed since VaR first became part of the capital regime. The regulators have responded to this by placing restrictions on VaR models, in an effort to embed into VaR the properties that are desirable of regulatory capital. The call for longer observation periods in order to eliminate procyclicality is a continuation of this mindset. So is the recent proposal of the Basel Committee to apply risk forecasts from a turbulent period to the positions of today, in order to calculate a so-called stressed VaR.

The restrictions on VaR have not worked in the past—trading book capital under the current regime is still flawed—and further restrictions are unlikely to make it work in the future. At the same time, the UBS example demonstrates that VaR models that hew closely to regulatory desires perform poorly as risk forecasters. The effort to make a desirable capital rule out of a good risk forecasting model has resulted in something that is neither.

Looking to the horizon

The bottom line is that short-horizon risk forecasts should be procyclical, and efforts to dampen this produce worse forecasts. But two weeks is too short a horizon over which to set prudential capital. The horizons over which procyclicality matters are measured in quarters or years, not days or weeks. For an institution or a system to build up risks over this type of horizon, it is not enough for a portfolio at a specific point in time to go bad (something VaR might warn against); rather, the systemic risks are a result of institutions' reactions to an evolving market.

So the fundamental question for the fundamental review is how to define minimum capital for trading books such that the regime is risk sensitive, is countercyclical (or at least, not procyclical) and protects

not just a fixed portfolio, but rather an institution as a going concern. While VaR on its own satisfies the first of these requirements, it is incompatible with the second, and by construction inappropriate for the third. There is a strong argument, then, for moving away from a capital requirement defined as a multiple of VaR, not because VaR fails to perform as advertised, but because it succeeds.

If we accept VaR for what it is, there may be a role for the measure in a new capital regime, even if not the central role it plays today. And if not, the regulators should still emphasize that institutions should measure, manage and disclose the short-term risks in their trading portfolios. In either case, regulators should allow banks the freedom to build models that best forecast short-term risk. In return, banks should not only demonstrate adequate statistical performance of their risk forecasts, but also an acceptable link between any position's representation in the risk model and its actual market prices.

Finally, some short-term events can in fact produce systemic risks: crowded trades, contagion, liquidity shocks, the transition to non-normal markets. Signals for these should become part of the supervisory arsenal. But far from giving up and bucketing all of these as unknowable fat tails, we should be seeking indicators of when market imbalances have made these dynamics more likely. Regulators, banks and the rest of the financial community should share in this task.

Further reading

- Campbell (2009). The Risk of Value-at-Risk. *Risk*, **22**(4): 42–46.
- Financial Services Authority (2009). A Regulatory Response to the Global Banking Crisis. Discussion Paper 09/2. March.
- Financial Services Authority (2009). The Turner Review: A Regulatory Response to the Global Banking Crisis. March.
- Finger (2005). Eating Our Own Cooking, RiskMetrics Research Monthly, June.
- Finger (2008). The SSG and My Two Brains, RiskMetrics Research Monthly, March.