Deloitte.

In Defence of VaR

Risk measurement and Backtesting in times of Crisis

1 June, 2020

Executive Summary

Since the Financial Crisis, Value at Risk (VaR) has been on the receiving end of a lot of criticisms. It has been blamed for, amongst other things, not measuring risk accurately, allowing banks to get away with holding insufficient capital, creating over reliance on a single model, and creating pro-cyclical positive feedback in financial markets leading to 'VaR shocks'.

In this article we do not seek to defend VaR against all these claims. But we do seek to defend the idea of modelling risk, and of attempting to model risk accurately. We point out that no single model can meet mutually contradictory criteria, and we demonstrate that certain approaches to VaR modelling would at least provide accurate (as measured by Backtesting) near-term risk estimates. We note that new Market Risk regulation, the Fundamental Review of the Trading Book (FRTB) could provide an opportunity for banks and regulators to treat capital VaR and risk management VaR sufficiently differently as to end up with measures that work for both rather than for neither.

Contents

1 Market Risk VaR and Regulatory Capital

Market Risk VaR was developed at JP Morgan in the wake of the 1987 crash, as a way to answer then chairman Dennis Weatherstone's question "how much might we lose on our trading portfolio by tomorrow's close?" VaR attempts to identify a quantile of loss for a given timeframe. For example, the 95th percentile would mean that I would only expect to lose more than VaR on one in twenty days.

This was a seemingly transparent and sensible approach to risk management and the use of VaR spread across the industry. When the Basel accords began their focus on traded risk in the 1990s, a version of VaR became the basis of the Market Risk Capital Charge.

1.1 History repeating itself

Typically VaR for complex portfolios is calibrated to a history of risk factor returns (this is explicit in Historical Simulation, used by the majority of large banks, and implicit in the calibration of the Monte Carlo approach used by a diminishing minority). So VaR often substitutes the question "how much did I lose on the the percent worst day in the scenarios covered in my history?" for the question "how much might I lose on the 5th worst day in the future".

This means that if the scenarios I use to calibrate my VaR did not contain a period of market stress, then it is unlikely to work for a period of market stress. In particular, if a sudden shock leads to an uptick in market volatility, and VaR is calibrated to a lengthier period of recent history, then my measurement of risk will fail to keep up with the actual risks posed by real market conditions.

It is actually not too difficult to create a more responsive VaR measure that will keep up with market volatility, a point we will come on to later. But for now we note that many vanilla VaR measures (including measures used by banks and mandated by regulators) are unresponsive to upticks in stress.

1.2 The current regulatory framework

The financial crisis occurred in 2008, after an unusually long (in historical terms) period of calm markets. Losses exceeded VaR (a "Backtesting exception") far more than the number of times one would expect if the model were accurately measuring the risk of losses at the given quantile. And banks turned out to be insufficiently capitalised across the board, leading to bail outs and all the political and economic turmoil that followed.

As a consequence of the post crisis Basel reforms, Market Risk capital is now composed of (amongst other things) a VaR (a VaR measure based on recent history) a Stress VaR (SVaR, a VaR measure based on the worst year of scenarios available in the bank's history, usually 2008) and a multiplier. The VaR and SVaR are both based on 10 days of trading losses and a 99th percentile. They are then scaled up by a multiplier based on the number of backtesting exceptions the VaR model has experienced. The multiplier starts at three but if the VaR model performs worse than expected, with four or more exceptions in the past year, the multiplier starts to increase.

The current Market Risk capital requirements are considered a stop-gap, to be replaced by FRTB at some point in the 2020s (dependent on jurisdiction, but most jurisdictions have not yet set dates for FRTB as a capital requirement).

1.3 Market Risk Capital in the Current Crisis

VaR has performed no better this time round than it did in 2008. According to Risk.net, US banks experienced 42 exceptions in Q1 2020, while the US units of a Canadian and a European bank each exceeded their VaR by more than 1000%. In Europe things didn't look much better, with one typical bank experiencing six exceptions in Q1 (for a 99th percentile VaR one would expect less than 1 exception per quarter).

By rule these exceptions should be leading to an increase in the multiplier, which could lead to a significant increase in Market Risk capital requirements (the multiplier begins to increase after a bank has experienced 4 or more exceptions in a one year window). Moreover, VaR models do eventually incorporate the COVID-related market swings into their histories and therefore register a very sharp (albeit too late) increase in VaR. This 'pro-cyclicality' of market risk capital requirements has caused concern amongst regulators, many of whom appear to be opting not to apply the multiplier increase to banks that experience exceptions.

Meanwhile, so far at least, the overall market risk capital levels do not appear to have been insufficient. Banks do appear to have weathered the storm better than they did in 2008 (though it was not market risk losses that directly led to the bankruptcies then either). The inclusion of a Stress VaR measure and a multiplier does seem to have created a reasonable capital buffer.

1.4 Risk Management vs Capital

So regulators want a capital measure that is sufficient for banks to withstand a major crisis, while risk managers presumably want a metric that accurately measures risk.

It should be clear that no single number can satisfy both. The trading risk associated with a 10 day horizon genuinely did increase significantly between January and March 2020. It is unlikely that any regulator wanted to see capital go up quite that much in such a short timeframe (hence the inclusion of SVaR in capital).

Yet somehow VaR manages to be criticised from both angles. It is attacked for being pro-cyclical but also for being inaccurate.

It is our argument that no single model works for both purposes. The same underlying 'VaR engine' could work, but a VaR that is reactive to market volatility is, by construction, highly pro-cyclical, while a VaR that produces a sufficiently high through the cycle capital measure is, by construction, non-responsive to changes in market conditions.

The problem is that the focus on regulatory capital since 2008 has meant that banks and regulators alike have tended to focus on methodological changes that produce 'the right' capital number, not that sharpen the accuracy of their risk models. Many banks use a plain vanilla historical simulation VaR both for regulatory capital (where there is a 10 day holding period and a 99th percentile and a stressed and an unstressed measure) and for internal risk management (usually the recent VaR only, a one day holding period and a 95th and 99th percentile). It is this 'risk management' VaR (at the 99th percentile) that is used for the regulatory backtesting that sets the 'multiplier'.

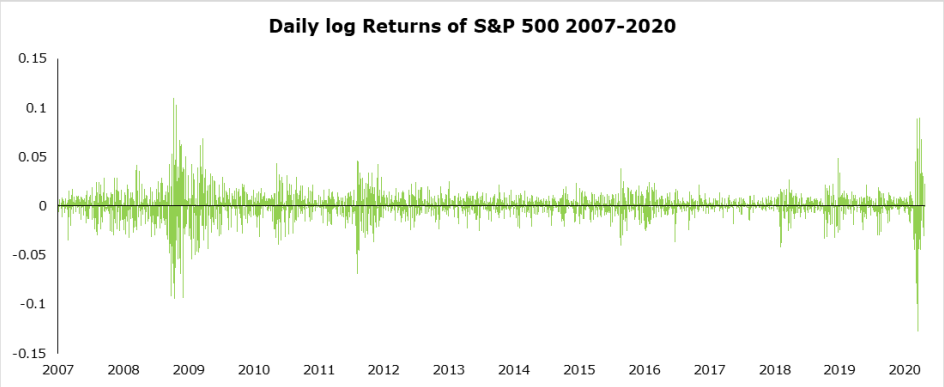
2 Can VaR accurately measure risk?

We now ask the question "could one produce a VaR model that could accurately measure near term market risk?" We show that this is possible. To be clear we do not use a fancy new model. We simply apply the recommendations of the RiskMetrics Technical Documents published in the early days of VaR, and building on the pioneering implementations at JP Morgan.

2.1 Volatility Clustering

It is a 'stylised fact' about markets that while the *direction* of yesterday's move tells us nothing about the direction of move we should expect today, the *size* tells us about the size we should expect today. In other words, markets exhibit volatility clustering: there are periods of calm, low volatility, and periods of high volatility in which there are a succession of large moves. Clusters of high and low volatility are clearly visible, for example, in Figure 1,Figure 1: which shows returns for the S&P500 from 2008 to present.

Figure 1: Daily log returns of S&P 500 (2007-2020). Clusters of higher volatility (larger bars, both in the positive and negative direction) are clearly visible.



'Fat tails' are often posited as the reason that VaR models can be unsuitable for risk measurement, but in fact, it is volatility clustering that tends to cause real problems for VaR models. The issue is not simply that the tails are fat, but that, much like London buses, you can spend ages waiting for a move in the tail and then see a whole number of big moves come along all at once.

Any VaR model that aims to accurately measure near term market risk, would do well to establish whether it believes 'current volatility' is high or low, rather than to simply use a uniform time period form history and hope things are 'about average'.

A simple way to achieve this, suggested by RiskMetrics in the 1990s, is to track an Exponentially Weighted Moving Average (EWMA) volatility, to 'normalise' historic moves by the observed EWMA volatility at the time, and to 'scale up' VaR by the present EWMA volatility.

Thus, when initial market tremors begin to appear the VaR model starts to realise that markets are becoming more volatile and begins to predict risk measures incorporating the possibility of larger quakes, well before a 'vanilla' historic simulation model would start to show increased risk.

EWMA comes with a 'decay factor' that suggests how strongly to weigh more recent observations (the smaller the 'lambda', the faster the decay, the more the most recent observations outweigh previous dates). Risk Metrics recommend a 'Lambda' of 0.94, which is what we have used for our analysis.

2.2 Our analysis shows that EWMA VaR performs far more strongly than vanilla VaR

Previous experience shows that EWMA VaR would have suffered far fewer backtesting exceptions during the last financial crisis, 2008-9. Our analysis updates this to show that the same was true of the COVID-19 related market turbulence. EWMA VaR's superiority is evident over a range of asset classes. It holds whether one is comparing 99% VaR, 95% VaR or Expected Shortfall measures (although ES does perform somewhat better than vanilla VaR). Even when we add a 'lag' of five working days to the vol scaling (perhaps in line with banks' market data update frequency, EWMA vol still outperforms.

The graphs below are typical for EWMA VaR. The long time series of P&L has clusters of volatility. When these appear, the VaR line 'hugs' the P&L volatility and rises almost seamlessly with it. The vanilla VaR, on the other hand, takes time to 'catch on' and rises far more jerkily.

Figure 2: Comparison of Risk Measures - Vanilla VaR vs EWMA VaR. A backtesting exception occurs where a daily loss exceeds a risk measure. The EWMA VaR line 'hugs' the P&L volatility and rises almost seamlessly with it. The vanilla VaR, on the other hand, takes time to 'catch on' and rises far more jerkily.

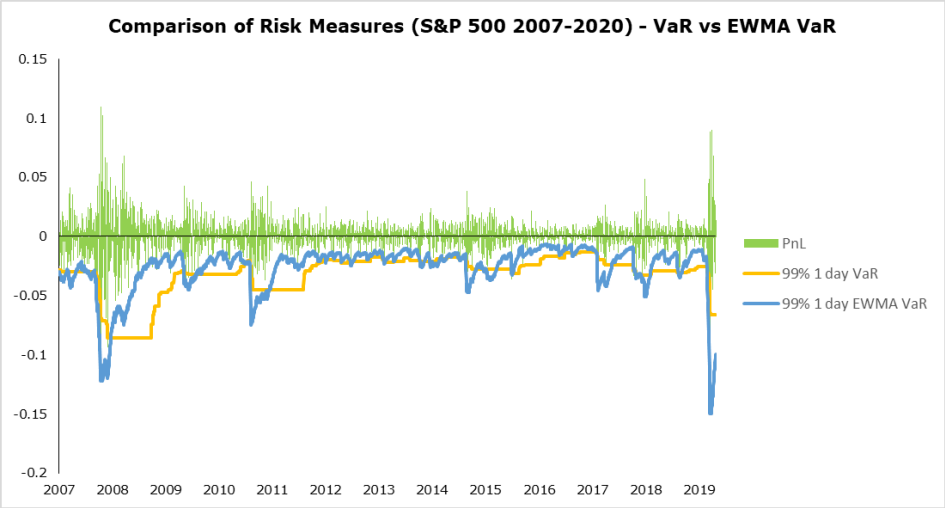


Figure 3: Comparison of Risk Measures – ES vs EWMA ES. A backtesting exception occurs where a daily loss exceeds a risk measure. While ES is more reactive than VaR, EWMA is still better at capturing the volatility clustering

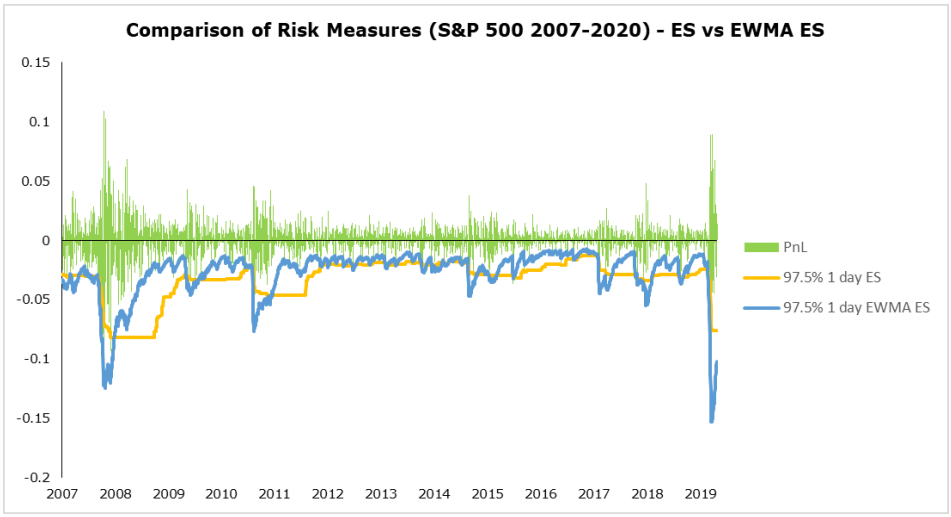
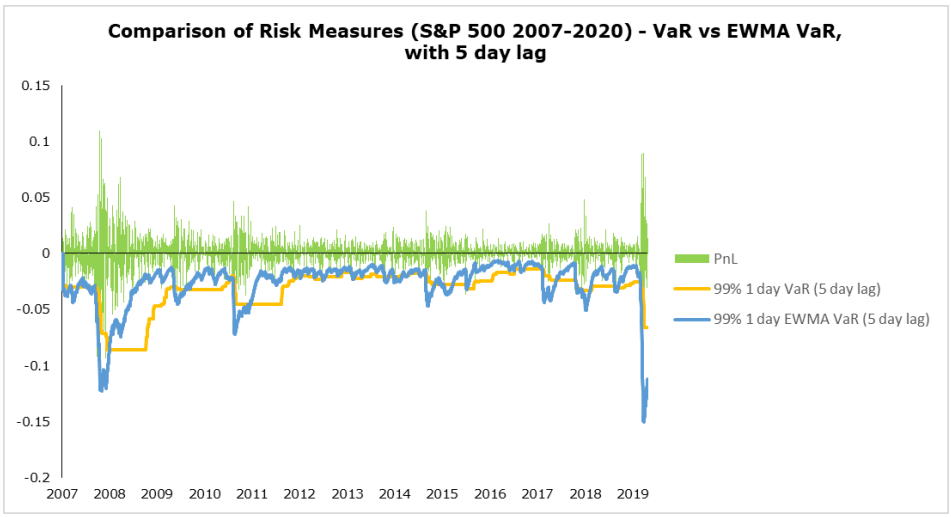


Figure 4: Comparison of Risk Measures - Vanilla VaR vs EWMA VaR, with 5 day lag in time series updates. A backtesting exception occurs where a daily loss exceeds a risk measure. The EWMA VaR is now less responsive than with a zero lag, but still better at capturing volatility clustering than its vanilla counterpart.



In the graphs shown we have used a vol scaling that scales down as quickly as it scales up, leading to slightly more exceptions in the calmer years. This is easily resolved by flooring EWMA VaR at vanilla VaR levels (bounding the scaling factor by 1). Table 1 presents some detailed comparisons.

Table 1: Backtesting Exceptions under each Risk Measure during "stress" (2008 and 2020) and "quiet" years (2018 and 2019) for S&P 500. Note that EWMA VaR could be floored at VaR, meaning it would never suffer more exceptions.

Number of Backtesting Exceptions	2008	2018	2019	2020
Vanilla VaR vs EWMA VaR				
99% 1 day VaR	13	7	1	8
99% 1 day EWMA VaR	3	3	4	3
ES vs EWMA ES				
97.5% 1 day ES	13	7	1	8
97.5% 1 day EWMA ES	3	3	2	2
Vanilla VaR vs EWMA VaR, with 5 day lag				
99% 1 day VaR (5 day lag)	18	10	2	11
99% 1 day EWMA VaR (5 day lag)	10	7	5	8

We show that EWMA VaR performs better in Backtesting during 2008 and Q1 2020. Without flooring it does perform slightly worse in the quiet years.

We have performed similar analysis across a variety of asset classes[1]. For Delta-1 portfolios in US Treasuries, EUR:USD and Oil (WTI ETF) between 2008 and Q1 2020, the findings are consistent, EWMA VaR outperforms VaR and ES equivalents during stress. Although some time series would fail backtesting in 2020 even with EWMA, EWMA still offers the best performance compared to both VaR and ES with equal weighting.

Table 2: Backtesting Exceptions under each Risk Measure during "stress" (2008 and 2020) and "quiet" years (2018 and 2019) for a range of other portfolios. Note that EWMA VaR could be floored at VaR.

Portfolio	Number of Backtesting Exceptions	2008	2009	2018	2019	2020
US Treasury Yields 5Y (2008 - 2020)	99% 1 day VaR	11	1	1	11	9
	99% 1 day EWMA VaR	5	3	5	4	2
	97.5% 1 day ES	10	1	1	7	8
	97.5% 1 day EWMA ES	5	1	3	2	1
	99% 1 day VaR (5 day lag)	16	1	1	15	10
	99% 1 day EWMA VaR (5 day lag)	10	4	7	8	6
EURUSD FX Rate (2008 - 2020)	99% 1 day VaR	10	0	4	1	5
	99% 1 day EWMA VaR	5	4	4	3	2
	97.5% 1 day ES	6	0	1	1	3
	97.5% 1 day EWMA ES	5	3	2	3	2
	99% 1 day VaR (5 day lag)	11	0	4	1	5
	99% 1 day EWMA VaR (5 day lag)	8	5	8	4	2
Wisdom Tree WTI Crude Oil (2009 - 2020)	99% 1 day VaR		2	8	2	6
	99% 1 day EWMA VaR		3	5	3	3
	97.5% 1 day ES		2	9	2	4
	97.5% 1 day EWMA ES		3	5	3	3
	99% 1 day VaR (5 day lag)		1	7	2	10
	99% 1 day EWMA VaR (5 day lag)		4	11	4	6

[1] We only present a subset of results here. Our flexible toolkit allows us to run a variety of configurations for any given portfolio. If you would be interested in seeing a particular analysis, or an analysis on your own portfolio, please get in touch.

2.3 So VaR can measure risk?

A 99% 1d EWMA VaR on a Delta-1 S&P 500 portfolio would have experienced only 3 exceptions the first four months of 2020. Somewhat more than one might expect of an accurate model but not enough to attract a multiplier increase and a great improvement over some of the numbers seen in industry.

So if we are looking for a number that does what Dennis Weatherstone asked, that gives us an estimate of how much our trading portfolio could lose tomorrow, we could do a lot worse than use EWMA VaR. If senior management and head traders want to know how their portfolio measures up against the current nature of the market, they could do a lot worse than look at EWMA VaR.

Of course, there are things VaR could never do. It cannot capture a complete 'Black Swan' event the like of which has never been seen before (if we believe such events exist). It cannot predict that markets will become more volatile if the uptick in volatility has not already begun (no VaR model could have been developed that would have risen with COVID in January, before markets took note). It cannot predict in which areas market losses will be experienced (although it can measure which areas the bank is more exposed to). It would not be very helpful at obtaining a measure of market risk months into the future.

Banks have rightly focused on stress testing, scenario modelling, reverse stress testing and other techniques to try to measure risk in these additional ways. But none of this is to say that that daily report of VaR risk should not be produced, or to say that efforts should not be made to make it as accurate as possible.

3 Regulation and Risk Management

The focus on regulatory capital since 2008 has meant that most risk management and measurement projects at major banks have been driven by regulatory requirements, and relatively little additional effort has been made beyond evidencing compliance with these – much stricter than pre crisis – rules. As we have seen, there are good arguments against using a EWMA VaR for regulatory capital: it could be far too volatile for banks trying to maintain stable capital ratios. A Stress VaR measure is a simpler way of ensuring sufficient capitalisation.

While some banks do use vol scaled VaR, or similar approaches, either internally or as part of regulatory capital calculations, many use a 'plain vanilla' measure for both. In our view this leads to an ostrich-like approach to risk measurement. We all know we could produce better risk reports using EWMA, but because regulations imply a vanilla VaR is more appropriate, we use this worse risk measure in our reports, and pretend we couldn't do any better. And then complain about "VaR" in general terms.

It is our argument that if we know we can measure risk with reasonable near term accuracy, and we could do so with relative ease, there is little excuse for not producing this report for senior management. If we can measure risk reasonably well then we should. The analysis above and plenty of other analyses show that we can. So we should.

Given the continued pressure to deliver regulatory change, the questions we ask are: will we also try to produce better risk measures? Perhaps the recent market volatility will provide an incentive to better measure risk? Perhaps the new FRTB regulation will provide the space?

3.1 FRTB and Capital Calculations

FRTB fundamentally overhauls Market Risk capital calculations. Banks can choose to use the Standardised Approach or Internal Models. Here we focus on Internal Models. These must use an Expected Shortfall measure calibrated to a stressed period, accounting for only limited asset class diversification and adjusting for liquidity horizons. Despite many industry concerns around implementation challenges, many of these changes are, in our view, fundamentally sensible. Some could feed into near term risk measurement, but overall FRTB is designed to be a Through the Cycle capital requirement, reducing pro cyclicality and creating a more stable capital number, as well as a more robust one.

We provide a brief overview of the key changes[2]:

- Expected Shortfall: Many of the headlines around FRTB focused on the move from VaR to ES. We argue this actually delivers relatively little change. Expected Shortfall asks, 'if losses exceed VaR, then on average what loss can I expect'. It is a 'coherent' risk measure (satisfying a set of academic axioms) whereas VaR is not sub-additive. It should capture 'tail risk' better than VaR (since it explicitly measures the tail). In reality though, regulators do not expect banks to model an

[2] MAR 33, Minimum capital requirements for market risk, BIS, Jan 2019.

'extreme value', we just use the average of the worst 6.25 scenarios in a year rather than the 2.5th worse. While our analysis shows that ES is more responsive than VaR in a crisis (since it picks up a very bad scenario immediately) the FRTB measure is anyway stress calibrated so this provides no real advantage. And we should point out that FRTB IMA is no more sub-additive than VaR since stressed calibration breaks this property.

- Stress Calibration: this makes FRTB a 'through the cycle' measure. It removes the VaR / SVaR double count and reduces the pro cyclicality of current Market Risk capital. It makes sense from a capitalisation point of view but clearly increases the divide between capital and near term risk measurement.
- Limited Asset Class diversification: This limits the degree to which previously observed cross asset class co-movements are recognised in the modelled number. A vanilla 'hist sim' may pick up on 'spurious' or 'one-off' co-movements that occurred in one crisis that may not occur in another (for example Treasuries and Equities both fell during COVID-19, not something that happened in 2008). This innovation could be ported across to VaR measurement if so desired.
- Liquidity Horizons: This scales up the risks associated with 'less liquid' risk factors. For example, high yield credit is given a horizon of 60 days compared to a base of 10 days. Different liquidity horizons could also be ported across to a point in time risk measure if so desired.

3.2 FRTB and Backtesting

Backtesting and P&L Attribution are an integral part of FRTB. Each desk must pass certain test thresholds in order to obtain (and retain) approval to use internal models. However it is not simple to backtest a through the cycle capital number. Therefore it is point in time (one year lookback, none of the FRTB adjustments defined above) VaR that is backtested. This VaR must be generated using the same risk management model as the FRTB Expected Shortfall and therefore would be expected to highlight any issues resulting from insufficient risk factor coverage or poor approximation of pricing models.

Implicit in this backtesting requirement is an assumption that this backtesting VaR may be used for risk management purposes as well as backtesting, but this is not made explicit. It is made clear though that measures used for risk management are again expected to come from the same core model as backtesting VaR and FRTB ES. There is also a statement that the scenarios from the past year must be "equally weighted"[3] which could be interpreted as saying that EWMA cannot be used for Backtesting. It does seem somewhat odd to prescribe so precisely the model and then tell banks they must test the model and suffer penalties for its poor performance.

However, we have spoken to some in industry who believe that since EWMA is a weighting applied to aggregate portfolio P&Ls, rather than scenarios, it could still meet the criteria for Backtesting VaR (the historic returns of risk factors are given equal weighting, while a scaling factor is applied at the end of the process to reflect the moving average P&L volatility of the portfolio). While we have not had this confirmed to us by regulators, some have expressed a view that if a bank could demonstrate that an internally used EWMA VaR did not suffer backtesting exceptions when the regulatory measure did, this could at least help them make a case for leniency in terms of consequence (desk failure or multiplier increase) – since it would be clear the exception was not a result of poor P&L representation or risk factor coverage.

3.3 FRTB and Risk Management

FRTB is primarily a framework for market risk capital, but also incorporates regulatory expectations of enhanced risk management, for example around limit frameworks,

[3] MAR 32.18, Minimum capital requirements for market risk, BIS, Jan 2019.

stress testing and intraday risk. There is no requirement for banks to develop more accurate VaR for risk management purposes, but the framework nonetheless presents an opportunity for this.

By explicitly using a Through the Cycle measure for capital, yet still requiring banks to have robust risk management frameworks, FRTB arguably creates an opening for banks to develop more accurate VaR measures. The fear that capital will become too volatile is removed. And the benefit to Backtesting could be made to matter.

One potential objection to using a more predictive internal VaR would be that there would then be pressure to use it to generate a capital add-on, for example in Pillar 2. However the difference between a one day VaR measure and the explicitly through the cycle stressed calibrated approach to capital should be sufficient to allay these fears; it is highly unlikely that a one day moving average volatility risk management VaR would exceed FRTB through the cycle pillar 1 capital.

4 Conclusion

VaR has come under a lot of criticism in recent years, some of which is justified. But much of this criticism is around the use, or misuse, of VaR for regulatory capital purposes. The simple ask of providing a good measure of near term market risk for MI / risk management purposes remains a good one. And VaR is not too bad a response to this ask.

While 'vanilla' historical simulated VaR failed Backtesting miserably both in the financial crisis and again in the recent COVID-19 related turbulence, it is eminently possible to design simple VaR measures that would not have failed. EWMA Vol Scaling is in our opinion one of the simplest and best. Our analysis shows that it would experience far fewer backtesting exceptions than vanilla VaR.

There is little justification for knowing that a better risk measure could easily be produced while wantonly continuing to output a measure that is patently a poor measure of risk. The fact that regulatory capital may prefer the poorer measure should not be enough to justify using it for internal risk management.

While banks have spent the past 12 years playing regulatory catch-up, there is now a chance to try to get market risk measurement right with FRTB implementation. The backtesting failures around COVID-19 should be a wake-up call: we should be using better VaR models to measure risk.

Contacts

Daniel Mayer
Director, Market and Traded Risk
Deloitte LLP
dmayer@deloitte.co.uk

Stephen Weston
Partner, Artificial Intelligence Lead
Deloitte LLP
sweston@deloitte.co.uk

Julian Leake
Partner, Head of Financial Services
Risk Advisory
Deloitte LLP
jileake@deloitte.co.uk

Zeshan Choudhry
Partner, Head of Financial Services Quants
Deloitte LLP
zchoudhry@deloitte.co.uk

Diane Pham
Senior Manager, Market and Traded Risk
Deloitte LLP
dpham@deloitte.co.uk

Farhan Kadri
Senior Manager, Market and Traded Risk
Deloitte LLP
fkadri@deloitte.co.uk

Deloitte.

This publication has been written in general terms and we recommend that you obtain professional advice before acting or refraining from action on any of the contents of this publication. Deloitte LLP accepts no liability for any loss occasioned to any person acting or refraining from action as a result of any material in this publication.

Deloitte LLP is a limited liability partnership registered in England and Wales with registered number OC303675 and its registered office at 1 New Street Square, London, EC4A 3HQ, United Kingdom.

Deloitte LLP is the United Kingdom affiliate of Deloitte NSE LLP, a member firm of Deloitte Touche Tohmatsu Limited, a UK private company limited by guarantee ("DTTL"). DTTL and each of its member firms are legally separate and independent entities. DTTL and Deloitte NSE LLP do not provide services to clients. Please see www.deloitte.com/about to learn more about our global network of member firms.

© 2020 Deloitte LLP. All rights reserved.