Novosyolov, Arcady and Satchkov, Daniel (2010)
Crash Testing: Don't be a dummy, Investment & Pensions Europe, Feb 2010.

Instead of chasing after an infinite number of possible events, a risk manager must consider a limited number of impacts when stress or crash testing, argue Arcady Novosyolov and Daniel Satchkov. The challenge is to pay more attention to more plausible impacts without making specific timing predictions

In January the Basel banking committee described the shortfalls of many risk models coming into the latest crisis: "Given a long period of stability, backward-looking historical information indicated benign conditions." For the last three years our research has focused on making models resistant to this very problem. Here we present our solutions in the realm of stress testing.

Stress testing is like car crash testing. Car designers are not concerned with the precise scenario that might lead to a given accident (did the car hit a pole or a tree?) but instead focus on a set of impacts, such as frontal, side, and rollover, that every car must be tested against. Similarly, we limit portfolio tests to a number of impacts that are observable in the financial markets, selected from categories such as: market-wide (for example, S&P 500); sector (especially sectors that do not comprise a large part of portfolio holdings and therefore affect the portfolio in an uncertain manner); or commodity or economic variable.

We will compare two methods for calculating actual global portfolio losses conditional on shock impacts, lasting one month, using six actual extreme events between 1998 and 2008. We disregard 'naïve' methods that consist of shocking one factor while leaving others untouched, thereby assuming zero factor correlation, and move directly on to the 'time-weighted' (TW) method described by Paul Kupiec in 1998, and our own modification, which we call 'event-weighted' (EW). TW uses correlation and volatility estimates obtained by using time decay, that is, it assumes that covariance structure during the extreme event is not different from such structure in today's environment. EW modifies the weights assigned to each observation for calculation of covariance estimates to better mimic the extreme conditions being tested.

Both methods are based on conditional multivariate normal distribution, in which the exponential decay factor assigns higher values to those observations deemed to have the most valuable information content. Under TW, typical risk models assign higher weights to more recent observations

– a valid assumption when the objective is to predict a risk profile of a portfolio in the present environment, but one which does not take into account the probability that today's correlations will not remain relevant when a major impact occurs. Under EW, we simply change the importance order of the observations based on their similarity to the event we are trying to model.

We must remember that testing of any prediction is a comparison of a prediction made before the fact with the actual outcome observed subsequent to established prediction. Because our approach is explicitly designed to go beyond the pre-defined factors in a given covariance matrix and allows flexibility in shock definition, we potentially have an infinite number of ways to specify the impacts. To narrow the number of tests, we must again use the fact that various extreme shocks manifest themselves through a limited number of observed indices and metrics. For example, the credit crunch started primarily with shocks to financials and then spread. The S&P 500 Diversified Financials index declined 14.4% between 15 July and 15 August 2007. If we define such a shock to the financials sector on 14 July and use the factor model data available prior to that date, we can make a prediction of a monthly return for a portfolio and then compare it with the way that the portfolio actually performed over the subsequent month. Repeating this procedure over many portfolios and over available history of actual extreme impacts would give us a good idea of the system's predictive power.

We tested EW on 176 randomly-chosen global institutional portfolios, against six historical shocks: LTCM (by applying a -14.6% shock to the S&P 500 on 31 July 1998); the end of the internet bubble (a -22.9% shock to the NASDAQ on 31 October 2000); the 9/11 attacks (a -8.2% shock to the S&P 500 on 31 August 2001); the first major impact of the credit crunch (a -14.4% shock to the S&P 500 Diversified Financials index on 14 July 2007); the continuing manifestations of the credit crunch (a -19.5% shock to S&P 500 Diversified Financials index on 31 May 2008); and the culmination of the credit crunch (a -17.3% shock to S&P Diversified Financials index on 16 August 2008).

The clearest result across most of these scenarios is that EW produces more conservative estimates of conditional average portfolio losses than TW – that is, its loss predictions are more extreme. This is as expected, since by assigning higher weights to similar extreme events from the past, EW's correlation and variance estimates are mostly higher than TW's. EW does considerably better at estimating conditional losses in the LTCM scenario and, even more importantly, it produces far fewer significant underestimations, which constitute one of the key dangers in a stress testing process. The 9/11 scenario produces more underestimations, but EW still performs better than TW – and we see a similar pattern for the culmination of the credit crunch in fall 2008. TW performs almost as well as EW during the earlier, less extreme credit-crunch shocks.

Where TW outperforms the predictive power of EW is in the dot-com crash scenario. There we see that EW is far too conservative for most of the portfolios. Again, given the effect of EW's assumptions on correlation and variance estimates, this is not surprising: this was a unique period characterised by both extreme sector impact (NASDAQ shocks) and non-rising (and even reduced) correlations. The TW method should be used when correlations are not expected to move; otherwise the EW method presents a more powerful alternative.

It may be objected that in order to give higher weights to events similar to those one wishes to test, one has first to decide which types of event are most likely to be repeated in the near future – in other words, to be able to predict the nature and timing of extreme events. But it is possible in many cases to refine the process and discriminate between the impacts based on their plausibility.