Basel Committee on Banking Supervision Consultative Document

Fundamental Review of the Trading Book

Dated May 2012

International Swaps and Derivatives Association, Inc.

the Global Financial Markets Association

And

The Institute of International Finance, Inc.

Further Response Covering

Calibration of Alpha and Beta

March 2013

Final Version

Contents

1. Introduction

This paper continues the proposal made in the Industry's February 2013[1] follow up paper to the Basel Committee on Banking Supervision's (BCBS) Fundamental Review of the Trading Book (FRTB). The February paper examines how total risk should be split for purposes of deriving the regulatory capital requirement and introduces two parameters α, which governs diversification benefit, and β, which controls the penalty for poor model performance. Here we explore an approach to calibrating α and β which is a general test on whether the scenarios used to generate daily Value at Risk (VaR), or Expected Shortfall (ES)[2], accurately describe the distribution of the Profit and Loss (P/L) being observed. We believe this approach would be a strong candidate for replacing the existing 'traffic lights' approach to back testing model based capital given the shortfalls of this measure previously noted by BCBS and summarised below. However, given the special problems with back testing ES, as noted by many leading authors, we believe this proposal addresses those concerns by testing the adequacy of the P/L distribution used to construct risk measures such as VaR and ES, rather than by testing the risk measures specifically.

We first briefly recall the proposal made in our previous paper. We then summarise the issues with back testing and recall the Probability Integral Transform (PIT) as a way of measuring model performance. PIT has been well documented in the academic literature. Then we propose a couple of methods for calibrating α and β based on PIT. One is a through the cycle method and the other is more of a point in time measure which will be more sensitive to recent model deterioration. In so doing, we avoid some theoretical issues about back testing expected shortfall by instead attempting to answer the question: "what is the probability that our model scenarios describe the underlying probability distribution of our observed P/L?" This probability can be directly interpreted as a measure of reliability of our scenarios for constructing a regulatory capital requirement, without dependence on the choice of risk metric itself. We believe it is this measure which should be the test of model risk, not the choice of risk metric constructed from the empirical distribution. That is, we test the adequacy of the scenarios used to construct VaR or ES, not the metrics themselves. We also note in passing that because we are testing scenarios directly as a description of the P/L distribution function, rather than test for specific weaknesses of the model, the same test metric can be used to calibrate α and β. This should be a welcome reduction in burden for firms and regulators.

[1]Further Response Covering Diversification and Model Approval, February 2013

[2]The precise form of this model – e.g. whether Value-at-Risk or Expected Shortfall, based on parametric simulation or historical data, etc., is of course another important topic for FRTB, but out of scope for this paper, hence we shall refer to 'internal model based capital' generically.

2. The proposed single framework

Our previous paper noted that there are different 'natural' dimensions of risk for different applications. No single cut of risk can address all relevant applications. We suggested, however, that the different approaches can be combined into a single capital framework, which by aligning the application with the natural cuts, enables regulatory capital to be aligned with risk taking. This is operationally efficient, and avoids unwanted outcomes or inappropriate incentives which could arise from an inappropriately specified approach.

Taking each component in turn:

Model-based capital for desks which *the firm* defines and selects for inclusion in internal models should be capitalized as the weighted sum of a fully diversified internal models number and the sum of individual components by risk factor, i.e.

$$Capital(\text{Internal Models}) = \propto M\big(\text{In-scope Desks, All Products, All Factors}\big) \quad (1)$$

$$+ (1-\propto) \sum_{\substack{i\in Risk \\ Factor \\ Categories}} M(\text{In-scope Desks, All Products, } i\,)$$

where α is a parameter between 0 and 1 determined based on the quality of the firm's model of diversification[3] (with higher α corresponding to a better model) and *M(x, y, i)* denotes the internal models capital charge for risk factor *i* on the desks *x* and products *y*. We view this as being strongly preferable to having supervisory specified correlations (as proposed in Annex 6 of the FRTB), for the reasons outlined in our February 2013 paper. Notice that equation (1) computes internal model-based capital on *all* products on in-scope desks, whether they are approved or not, and hence aligns to the actual P&L incurred – a key benefit, since we retain the link between model-based capital and actual risk taking.

Any trades for which the firm has not received regulatory approval on that product, or which are on a desk nominated as 'out-of-scope' by the firm, should additionally receive a full standard rules capital charge, i.e.

$$Capital(\text{Standard Rules}) = S\big(\text{Out-of-scope Desks, All Products}\big) \quad (2)$$

$$+ S(\text{In-scope Desks, Unapproved Products}),$$

where *S(x, y)* denotes the standard rules capital charge on desks *x* and products *y*.

Clearly, there is a double count here on unapproved products on in-scope desks, which are counted towards capital in both (1) and (2), but we see this as tolerable to retain the link between model-based capital and economic risk.

Lastly, we have the penalty function for desks which are nominated as in-scope by the firm, but found to perform poorly in backtesting. As argued above, the natural cut of risk in this case is by

[3]As opposed to the performance of the model on specific desks, see equation (3) below.

desk, and only desks which are in the scope of the internal model would be subject to testing. Here we propose a factor $\beta(j)$ be assigned to each in-scope desk, based on the desk-level performance[4] of the internal model, as assessed by backtesting and similar methods, with the penalty capital set as:

$$Capital(\text{Backtesting Penalty}) = \sum_{\substack{j\in \text{ In-scope} \\ \text{Desks}}} penalty \times \beta(j) \tag{3}$$

The penalty should smoothly transition from $\beta(j)$ = 0 in the case of ideal (or over-conservative) model performance, to $\beta(j)$ = 1 in the case of very poor (i.e. underestimation of risk) model performance. This enables areas where the model performs poorly to smoothly transition to a punitive, charge, without the 'cliff effect' of immediate model approval withdrawal seen in the existing framework.

Now, combining (1), (2) and (3), we have total capital for Market Risk as

$$\begin{aligned} Capital =& \propto M\big(\text{In-scope Desks, All Products, All Factors}\big) \\ &+ (1-\propto) \sum_{\substack{i\in Risk \\ Factor \\ Categories}} M(\text{In-scope Desks, All Products, } i) \\ &+ \sum_{\substack{j\in \text{ In-scope} \\ \text{Desks}}} penalty(j) \times \beta(j) \\ &+ S\big(\text{Out-of-scope Desks, All Products}\big) + S(\text{In-scope Desks, Unapproved Products}) \end{aligned} \tag{4}$$

Before addressing how to calibrate α and β we would like to provide a brief summary of the issues around the traffic lights system for back testing VaR, together with further issues around using the same framework for back testing ES.

[4]Out of scope desks could be thought of as 'automatically failing', and hence have beta(i) =1, and incur standard rules as a penalty.

3. Issues with back testing using the exceptions approach

It is known that the regulatory back testing approach suffers from "Type II Error[5]". The threshold for the test is set such that the probability of not rejecting the model when it is incorrect is high. From a practical perspective we note the following problems with the current back testing procedure when applied to VaR:

- It does not penalise firms based on the size of exceptions; all breaches count.
- It does not penalise firms which frequently come close to exceptions but just avoid a breach.
- Firms with multiple small breaches are penalised relatively harshly.

Furthermore if we apply back testing to ES there are additional theoretical problems. On any given day ES can relate to a particular VaR percentile but that percentile can vary from day to day as the tail of the P/L distribution changes through time and as a result of changes in the portfolio composition. Therefore ES breaches by construction are not independent Identically Distributed (IID) variables even if exceptions of VaR at a particular percentile are. Hence the null hypothesis required for the Kupiec Test is not valid. This means we cannot, from a statistical perspective, agree an acceptable count of exceptions. These observations are not new, see for example Dowd[6]. More recently Paul Embrechts has pointed out that ES is not an elicit risk measure and it is invalid to back test a risk measure that is not elicit[7].

4. Proposed replacement for the traffic lights approach to back testing

Purely on theoretical grounds then there seems to be a good case for reviewing the way we test the adequacy of models for regulatory capital. However it would be useful if in doing this we could come up with a more robust testing framework that also delivers a calibration of α and β, the two regulatory parameters that we propose replace the current regulatory multiplier which is floored at 3. Specifically we require α and β each to lie between 0 and 1 (inclusive) in order to deliver regulatory requirements to control diversification benefits taken by firms, and also to apply a capital penalty for poor model performance.

Given the problems identified above with testing particular risk measures, VaR and ES, we propose the use of PIT advocated by Crnkovic and Drachman (1996)[8] and Diebold et al (1998)[9]. An excellent

---

[5] See, for example, "Supervisory framework for the use of "back testing" in conjunction with the internal models approach to market risk capital requirements – Basel(1996)
and Pena, Rivera and Ruiz-Mata, 'Quality control of risk measures: backtesting VAR models', Journal of Risk (Volume 9/Number 2, Winter 2006/7)

[6] "Backtesting market risk models in a standard normality framework", Journal of Risk (Volume 9/Number 2, Winter 2006/7)

[7] Risk Minds conference, Amsterdam, December 2012.

[8] Crnkovic, C., and Drachman, J. (1996). Quality control. *Risk* **9** (September), 139–43

[9] Diebold, F. X., Gunther, T. A., and Tay, A. S. (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review* **39**, 863–83.

discussion of the approach is given in Dowd. We also note that Wolf Muller, Risk Analyst-BIS Market Risk, discussed PIT and the traffic lights system in an excellent presentation on backtesting VaR models in June 2008.

The PIT approach essentially allows us to assign a cumulative probability to each daily P/L observed, based on our VaR scenarios for that day. Over time the set of probabilities will have the Uniform U(0,1) distribution if our null hypothesis that the VaR scenarios describe the probability distribution of the observed P/L. We can then apply tests to empirically observed probabilities to see if they are indeed consistent with the Uniform distribution. For calibrating α and β we propose a variant of PIT but before doing that we will explain the PIT in a little more detail. We also note in passing that Dowd describes a number of extensions to PIT, a truncated distribution test -for example for testing just the lower tail – losses, several methods for testing – a multi-step P/L – for example 10-day P/L, tests where there is a dependence structure in the transformed P/L's – auto-correlation, and multivariate approaches including testing copula models.

The PIT transform is simply to calculate for each day $p_i = F(pl_i)$

Where F denotes the cumulative distribution function of pl_i, and pl_i is the P/L observed for day i. Now F is based on the scenarios generated to compute VaR on day i so that formally p_i is a conditional cumulative probability based on the scenarios employed for the model being adequate:

$$p_i = prob\left(x < pl_i / Model\right)$$

In the case of a historical simulation p_i can be obtained simply by observing the rank of the scenario closest to the pl_i observed. That is, for a historical simulation based on 1000 scenarios we would use:

$$p_i = \frac{rank(pl_i)}{1000}$$

Where *rank* (pl_i) is the rank of the scenario closest to pl_i

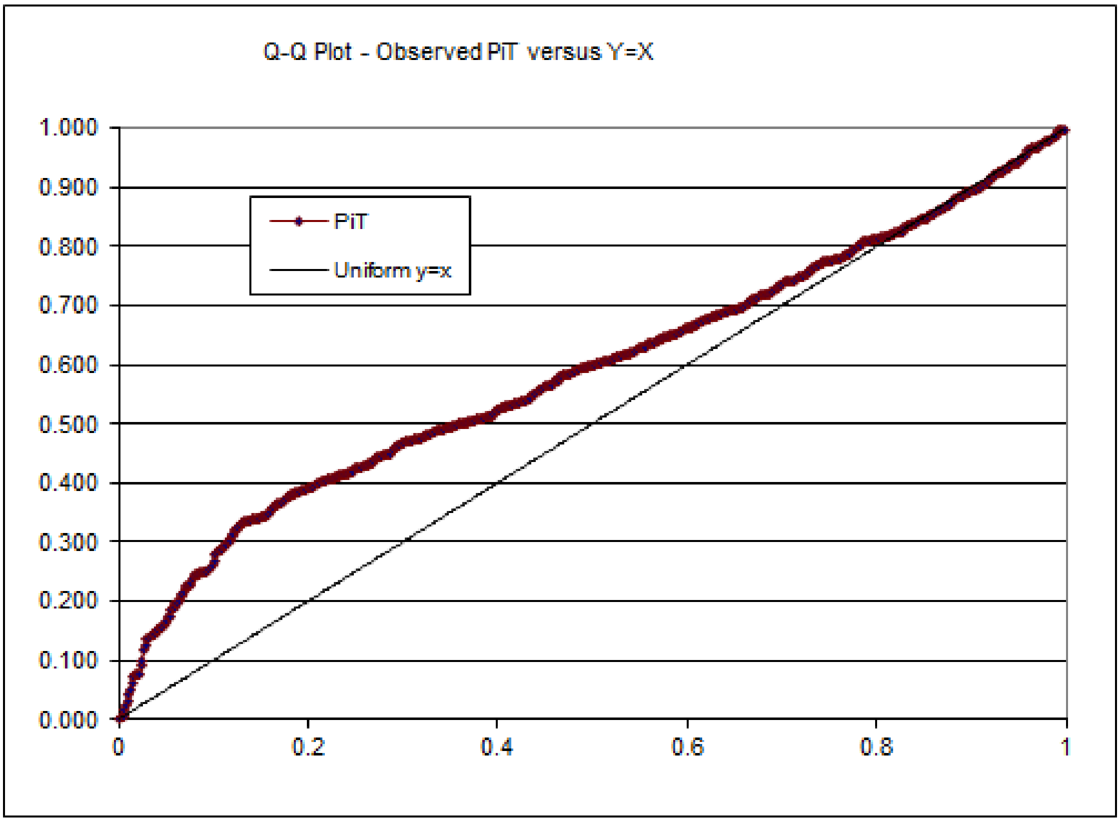
Note we are measuring rank here so that the most negative P/L scenario is 1 and the most positive is 1000. Hence, if we observe a P/L_i that falls closest to, say, the 250th scenario we would have p_i = 0.25. A loss greater than any scenario would be assigned a probability of 1/1000 and a profit above our highest scenario would be given $p_i = 1$. In the case of scenarios generated by Monte-Carlo simulation, given enough simulations it should always be possible to generate a loss scenario more negative than any loss observed, and a profit scenario greater than any profit observed. That is the end-point probabilities of 0 and 1 may never be observed. However, we do not believe any of the tests will be materially affected by the tests we discuss.

A number of tests can then be constructed on the p_i values by noting that, under the null hypothesis, the model scenarios adequately describe the distribution of the observed pl_i then the p_i values are uniformly distributed U(0,1). Dowd discusses a proposal by Berkowitz, to convert the p_i to standard normal variables by using the transformation:

$$q_i = N^{-1}(p_i) \qquad (5)$$

Where N denotes the standard normal distribution function

We can then test the q variables for normality under the null hypothesis the model scenarios are correct. There are also a number of tests we can apply for the uniform distribution, Kolmogorov-Smirnov, Kuipier test (not to be confused with Kupiec) and Anderson-Darling etc. However these tests are either quite difficult to apply or have weaknesses as Dowd notes. In fact simply plotting the ordered p_i variables against the line y=x provides a useful visual test for uniformity. Further, the deviations of the PIT variables from y=x provides insights into types of model failure, for example skew and fat-tail under/overestimation.



5. Calibrating alpha and beta

PIT is a useful tool for analysing model risk but for the FRTB we wish to extract from the data a measure of model risk that can be expressed in a single parameter. Note our philosophy here is to calibrate the parameters based on the performance of the model the firm proposes regardless of the alternative. For example, whether the regulator is concerned with diversification benefit being taken by banks across risk factors, across desks or across products we propose to calculate just one parameter α based purely on the model that the bank proposes to regulators. In a sense we calculate α under the null hypothesis that the bank has properly captured the benefit in its proposed model. Similarly, for the desk level add-on that regulators require we calculate β under the null hypothesis that the desk level model the firm proposes is appropriate. Conceptually the calibration of α and β is the same and this allows the firm to use a single framework for both, and regulators to be able to view the two parameters in the same way.

Regulators and firms may wish to consider many alternative test statistics that can be constructed from PIT that will provide a measure of model risk. We refer the reader to Dowd and other references for further discussion on PIT, including extending to multiple periods and allowing for time dependency.

Here we suggest two ideas which we believe meet regulatory needs and would provide a simple and intuitive calibration for α and β.

The measures presented lead to values of α and β continuous in the range 0 to 1 and are defined for some observation period. As with the current model validation regime, it is expected that results would be reported at regular intervals e.g. quarterly, and that the capital model i.e. the values of α and β that a firm is required to use, be updated only periodically and not daily. Further, for practical reasons (transparency, audit, consistency across jurisdictions), we suggest a granular lookup table be used according to 'tolerance bands' based on the statistical properties of finite sample estimates.

5.1 Through the Cycle Calibration (TTC)

Since the Cumulative Distribution Function (CDF) is bound between 0 and 1 a model adequacy parameter similarly bound between 0 and 1 could be constructed by simply as measuring the maximum absolute deviation of the empirical CDF from the y=x line. In the chart above the maximum deviation is 0.19.

An alternative approach would be to consider a 'goodness-of-fit' function quantifying the area between the empirical and uniform CDF. Since we are primarily concerned with under-estimation of losses, we propose to limit the range to (0, 0.5) and seek to preserve the sign of the difference between the two CDFs. We also propose that more weight should be given to differences in the extreme tail, and this can easily be accommodated for example by using a power function of (2z-1).

One possible example is given below – we note that this is closely based on an approach introduced by Zumbach[10]:

$$d_p = 2 \cdot (p+1) \cdot \int_0^{0.5} \left(CDF_{emp}(z) - z \right) \cdot |2z-1|^p \cdot dz \quad (6)$$

With p=8 almost 90% of the weight is given to CDF differences in the range (0,0.1), i.e. losses beyond the 90th percentile.

One of the key benefits of this approach is that it provides a structured way to 'blend' the results from testing multiple quantiles. Under the null hypothesis of a uniform distribution, it is straightforward[11] to numerically determine (e.g. by Monte Carlo simulation) the distribution of the test statistic d_p. This can then be used to inform the resulting scalar via the use of a suitably granular lookup table.

5.2 Point in Time Calibration (PiT)

Given the performance of models can vary through time we suggest also monitoring current model performance as a way of detecting any recent deterioration of model performance. In a sense this can be achieved by updating the TTC using a moving window of observations but the PiT approach can easily be adjusted to allow greater or less sensitivity depending on perceived systemic risk. Under PIT when a P/L is observed we record p_i, as 0 for large losses and p_i rises towards one for large positive profits. A p value close to 1 implies a high probability of observing a profit less than the P/L scenarios used for VaR, or ES. Lower observations of p imply a lower probability of observing that P/L if the VaR scenarios truly reflect the underlying P/L distribution. That is p_i is the probability of observing a P/L less that observed conditional on the scenarios being a valid description of the probability distribution of the observed P/L. We will denote this as:

$$p_i = prob(PL / Scenarios)$$

What we really require to know, in order to calibrate α and β is the probability that the scenarios are valid given the observed P/L. If the scenarios are valid then we have a valid model[12].

Using Bayes' theorem we have:

$$prob(Scenarios / PL) = \frac{prob(scenarios)}{prob(PL)} prob(PL / Scenarios) \quad (7)$$

---

[10] See for example 'Backtesting risk methodologies from one day to one year' – Gilles Zumbach, Journal of Risk, Volume 9/Number 2

[11] A working paper by David Phillips at RBS discusses calibration in more detail.

[12] A valid model does not necessarily imply an appropriate choice of risk metric. However as argued above, the risk measure is the choice of the regulator, and the firm for internal purposes. A poorly performing risk measure does not imply a poor model.

Where, in the usual way, we note that prob (scenarios) is the unconditional probability of the scenarios being correct and prob (P/L) is the unconditional probability of observing a profit less than PL, independently of the scenarios. We can therefore write this as:

$$prob(Scenarios/PL) \propto prob(PL/Scenarios)$$

Or

$$prob(Scenarios/PL) = K * prob(PL/Scenarios)$$

Hence we can use the constructed p_i as a measure of model performance and K is a normalisation factor.

We could, on a daily basis simply set

$$\alpha = Kp_i \qquad (8)$$

Consistent with the discussion on TTC we wish only to test that the model under predicts losses. That is we are interested in the left tail of the distribution. To achieve this in the PiT approach we select only those p_i values below 0.5. This is consistent also with the approach outlined in Dowd and again requires scaling the pi values by 2 since we are considering only the interval (0,0.5).

So finally we select only p_i values less than 0.5 and then set:

$$lp_i = \frac{p_i}{0.5} \qquad (9)$$

However lp_i will be very volatile from day to day since, even if the model is perfect and the lp_i are uniformly distributed as expected, we would still expect lp_i to vary over that full range. Hence we need to smooth the time-series of *l*p values that emerge or simply take an average over some appropriate window. The method of averaging, or smoothing, the p time series is something that regulators will want to prescribe if this approach is adopted. However our recommendation would be to use an exponential smoothing factor of around 0.99. This provides a path for p which is reasonably stable but remains sensitive to model deterioration.

A more formal foundation for exponential smoothing can be derived using a State Space[13] model where exponential smoothing is a simplified form of a Kalman Filter. The State Space model provides a framework for modelling an observed time series in terms of one or more unobserved components but requires assumptions to be made about the distribution of the unobserved components. In this particular application to model risk we model the observable lp_i as a function of the unobservable probability of the model being valid. In particular we assume that the observed lp_i values are a linear function of an underlying probability of the model being correct, but where that probability itself follows some transition rule. A typical formulation would be:

[13] For a detailed discussion on State Space models see: *Forecasting, Structural Time Series Models and the Kalman Filter*, Harvey, Andrew C, Cambridge University Press. Or for a brief introduction, “Understanding the Kalman Filter”, Meinhold and Singpurwalla.

$$p_i = \theta_i + \varepsilon_i$$
$$\theta_i = \theta_{i-1} + \beta_i + \nu_i$$
$$\beta_i = \beta_{i-1} + \varsigma_i$$

Here θ and β are system variables that describe the evolution of the true unobserved probability of the model being appropriate and ε, γ and ς are unobserved random variables. This kind of model can be estimated using a Kalman Filter. A Kalman Filter is an optimal rule for updating the system variables as now observations of the observed variable (*lp*$_i$) become available. When β is set to zero for all time periods so that θ just follows a random walk, and where the variance of the unobserved random terms is assumed to be constant through time, then the Kalman Filter reduces to exponential smoothing. The choice of smoothing parameter corresponds to the ratio of the variance of ε to the variance of γ and is in fact consistent with the calibration of tau in BCBS working paper 22 on securitisations.

Finally, on choice of K, we note again that a set of scenarios that perfectly describe the probability distribution of observed P/L would deliver uniformly distributed p_i values centred on 0.5. For such an ideal model we would want to assign a value of 1 to α in order to give full weight to the model. This suggests K should be set to 2 but we note though that regulators could set K to less than 2 as a regulatory override if, for example, they became concerned that the model was failing to capture some new risk that is unlikely to be reflected in historical data, or they were otherwise concerned about model risk notwithstanding the historical performance. As noted above they could also require firms to reduce the exponential smoothing parameter so that capital becomes more sensitive to model deterioration. A K factor of 2 could theoretically generate α value greater than one however and so we suggest the following rule for calibrating α.

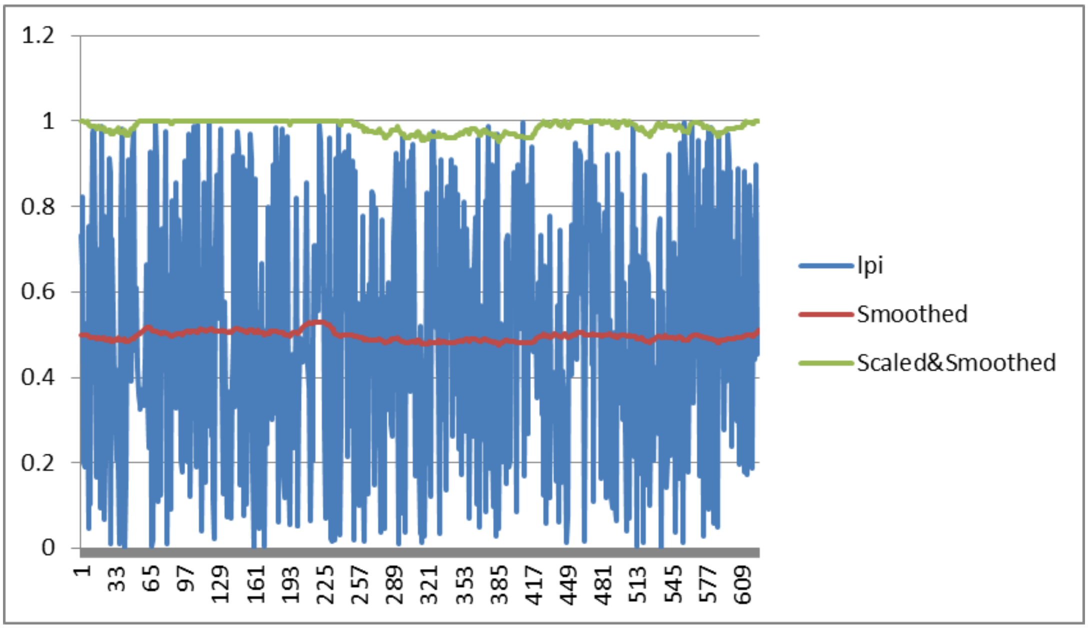
$$\alpha_i = Min(1, 2 * \theta_i)$$
$$\theta_i = 0.99 * \theta_{i-1} + 0.01 * p_i$$

Precisely the same method can be used for calibrating β at desk level.

We provide some sample charts to demonstrate the performance of the calibration.

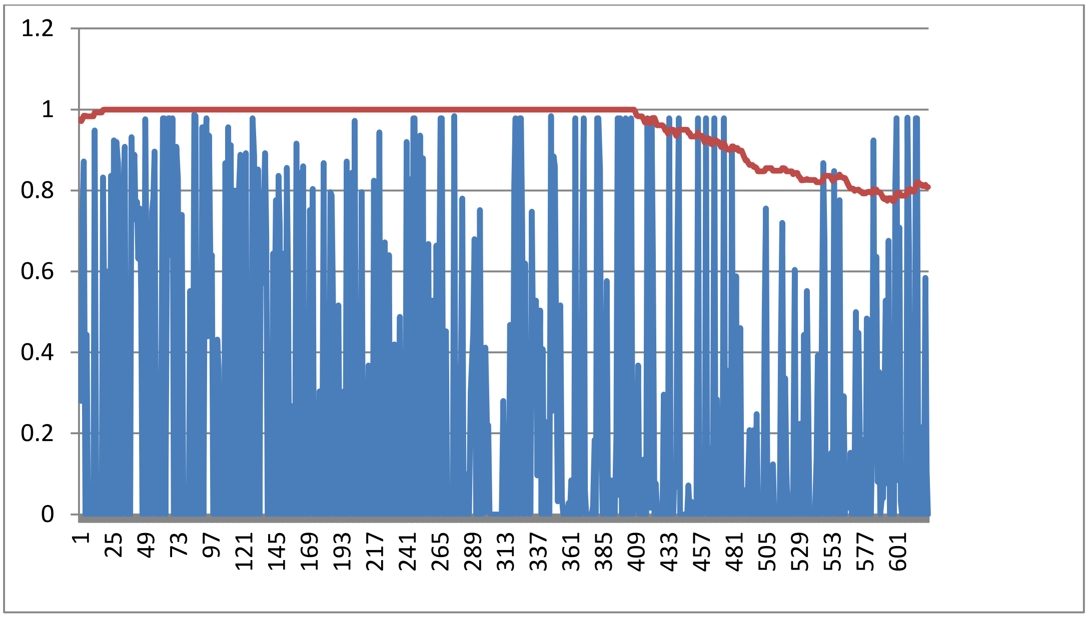
First for uniformly generated *l*p$_i$ to demonstrate that the distribution is centred on 0.5 if K is set to 1 (smoothed) together with K=2 (scaled and smoothed).

***l*p$_i$ distribution with K set to 1 and 2**

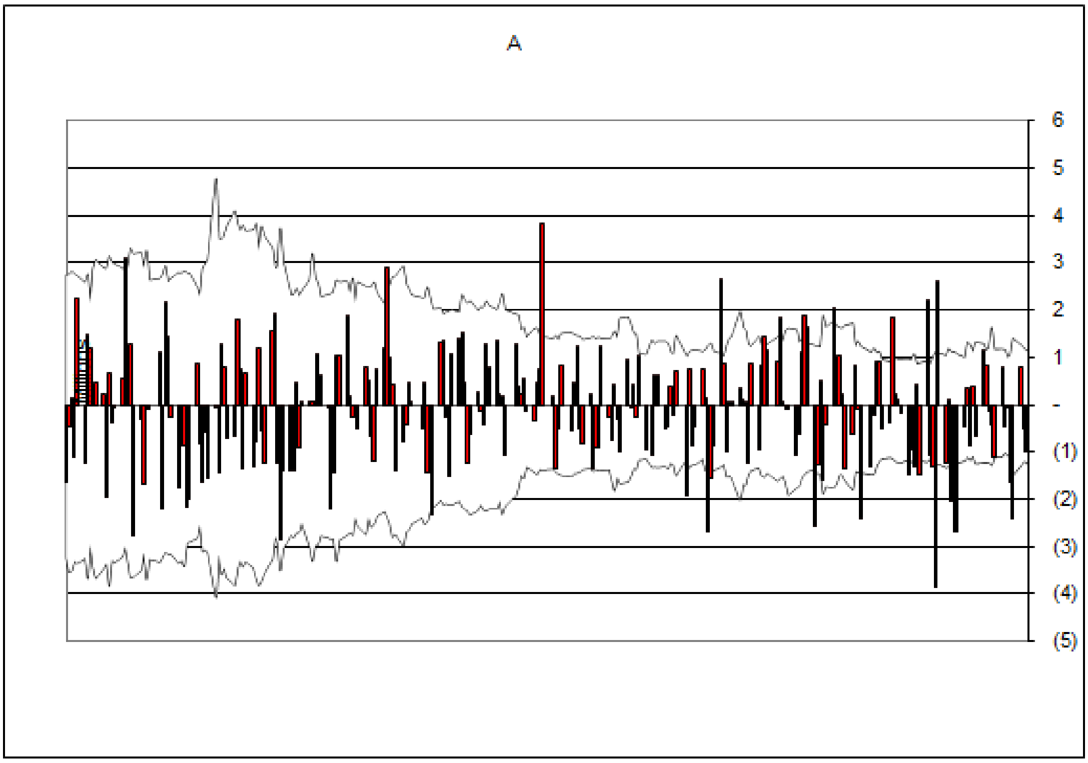


Then for some real desk level portfolios, using K=2

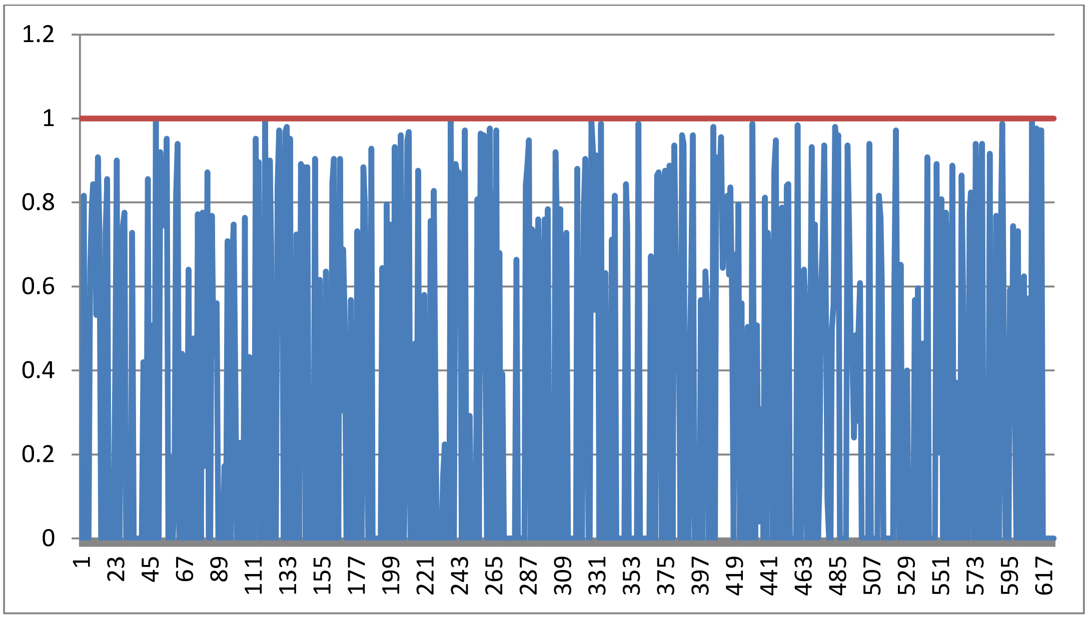
lp_i distribution with K set to 2



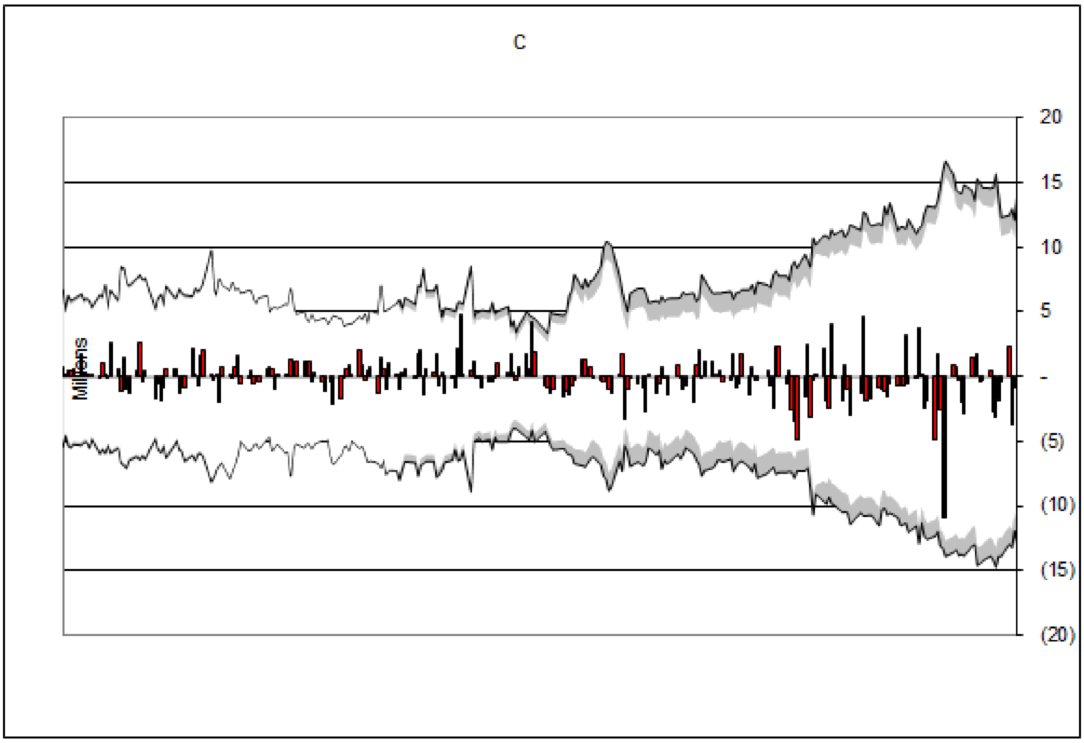
The corresponding VaR back testing chart is as follows:



lp_i distribution with K also set to 2, but for a different portfolio



The corresponding VaR back testing chart is as follows:



Just as with the TTC calibration it is desirable to use capital penalty tolerance bands based on the statistical uncertainty in an estimation of α from a finite sample. By way of illustration, bands based on confidence limits are shown in Table 5.1.

The tolerance bands depend on the size of the sample, and on an initiating value, or "seed", for θ. A sample size of 250 was used (and others for comparison) corresponding to the current backtesting observation period and representing just under four smoothing half-lives (with l=0.99, the half-life, or "memory", is about three months). A value of 0.5 for the θ seed (α=1) is advised because this leads to faster convergence of tolerance bands with respect to sample size. This choice of seed value can be thought of as an "innocent until proven guilty" approach, as a failing model will drift away from 0.5 while the successful model remains near its start point.

Table 5.1 – Lower Bound (LB) on α for a selection of confidence levels (CLs)

θ_0=0.5	N=100	N=250	N=500	N=1000
mean	0.98	0.98	0.98	0.98
median	1.00	1.00	1.00	1.00
LB CL 95%	0.94	0.93	0.93	0.93
LB CL 99%	0.91	0.91	0.91	0.91
LB CL 99.5%	0.90	0.90	0.89	0.89
LB CL 99.9%	0.88	0.87	0.87	0.87
LB CL 99.99%	0.86	0.85	0.85	0.85

Application of these results is envisaged to have the following structure (for a 250 day observation period);

- Capital model α = 1 while the observed value is greater than 0.93
- Capital model α = 0.90 while the observed value lies between 0.90 and 0.93
- Capital model α = 0.85 while the observed value lies between 0.80 and 0.85
- ... etc

Details of band definition and capital model α values are left to central regulators.

One great benefit of PIT over the traffic lights account is that model risk is identified in benign periods simply because VaR contracts. It is not necessary to observe exceptions for increased model risk. Note though that the risk measure we have proposed is asymmetric. It penalises the firm for losses that are extreme compared to the P/L scenarios (i.e. negative exceptions) but does not penalise positive exceptions. This is in line with the current traffic lights approach and seems sensible in that it would be unreasonable for a firm to take a capital add-on for excessive profit. Also risk management and risk modelling tend to focus on loss mitigation, not profit. A measure that treats model risk symmetrically would risk creating bad incentives. However we should note that large unpredicted profits are also a sign of poor model performance. It is possible to adapt the PIT approach so that it favours P/L in the centre of the P/L scenarios used for VaR or ES and penalises tail events at either end. We can pursue this if required and in any case a firm may choose to do this for internal use.

Appendix: A Short Note on a State Space Model for Alpha and Beta

A State Space form of a time series model is:

$$y_t = m(\theta_t, t) + \varepsilon_t$$
$$\theta_t = T(\theta_{t-1}) + \gamma_t$$
$$Var(\varepsilon_t) = \sigma_\varepsilon^2 \quad etc$$

Where y is an observation of a time series and m is an unobserved component – e.g. a trend component.

A frequently used simple form of the model is:

$$y_t = \mu_t + \varepsilon_t$$
$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t$$
$$\beta_t = \beta_{t-1} + \varsigma_t$$

$$Var(\varepsilon_t) = \sigma_\varepsilon^2 \quad etc.$$

In State Space form this is written:

$$y_t = Z\alpha_t + \varepsilon_t$$
$$\alpha_t = T\alpha_{t-1} + G$$
$$where$$
$$y_t = \begin{bmatrix}1 & 0\end{bmatrix}\alpha_t + \varepsilon_t$$
$$\alpha_t = \begin{bmatrix}\mu_t \\ \beta_t\end{bmatrix} = \begin{bmatrix}1 & 1 \\ 0 & 1\end{bmatrix}\begin{bmatrix}\mu_{t-1} \\ \beta_{t-1}\end{bmatrix} + \begin{bmatrix}\sigma_\eta & 0 \\ 0 & \sigma_\xi\end{bmatrix}$$

Where

$$Z = \begin{bmatrix}1 & 0\end{bmatrix}, \quad T = \begin{bmatrix}1 & 1 \\ 0 & 1\end{bmatrix} \quad and \quad G = \begin{bmatrix}\sigma_\eta & 0 \\ 0 & \sigma_\xi\end{bmatrix}$$

The Kalman Filter (KF) is an updating rule for estimating the time varying parameters contained in α, for example μ and β. The rule is a recursive one that works through time as follows:

Assume at the start of time t we know α_{t-1}. Then our best predictor for α_t before y_t is observed is:
$\alpha_{t/t-1} = T\,\alpha_{t-1}$

Then, given $\alpha_{t/t-1}$, when we observe y_t the Kalman Filter provides the following updating rule:

$$a_t = a_{t/t-1} + P_{t/t-1}Z_t'F_t^{-1}(y_t - Z_t a_{t/t-1})$$
$$and$$
$$P_t = P_{t/t-1} - P_{t/t-1}Z_t'F_t'Z_tP_{t/t-1}$$
$$where$$
$$F_t = Z_tP_{t/t-1}Z_t' + H_t$$

H is the time varying variance of ε and P is the time varying variance-covariance matrix of the parameters in α; that is the covariance matrix of μ and β.

The Kalman Filter can be shown to be an optimal predictor for α_t when the stochastic error terms are normally distributed. For a much fuller discussion of the Kalman Filter and State Space Models see Harvey.

In the simple case where β is assumed to be zero, so that θ is stationary, Z and T becomes simply one and can be dropped.

So the KF becomes simply:

$$\alpha_t = \alpha_{t-1} + \frac{P_{t/t-1}}{F_t}\left(y_t - \alpha_{t-1}\right)$$

Since F_t is a scalar given by:

$$F_t = P_{t/t-1} + H_t$$

And P_t is a scalar given by:

$$P_t = P_{t/t-1} - P^2_{t/t-1} * F_t$$

This can be rearranged as follows:

If H is not time varying then F and P are also time invariant so that finally we get:

$$\alpha_t = \alpha_{t/t-1} + \frac{P}{F}\left(y_t - \alpha_t\right)$$
$$or \quad \alpha_t = \alpha_{t-1} + \lambda(y_t - \alpha_{t-1})$$
$$and \quad finally \quad \alpha_t = (1-\lambda)\alpha_{t-1} + \lambda y_t$$

The Kalman Filter reduces to an exponential smoothing formula when the parameters are stationary and we assume constant variance of the stochastic terms.

Application to calibrating alpha and beta

Denote p_i =P(PL/M) to be the probability of observing PL on day I given the model scenarios are correct. Using Bayes' theorem we have:

$$q_i = Min(1,2 * P(M / PL)) = Min(1,2 * \frac{P(M)}{P(PL)} P(PL / M))$$

Our observation equation is:

$$q_i = \theta_i + \varepsilon_i$$

So, y_i is an observed value of p_i where p_i is the underlying 'true' probability of observing PL_i conditional on the scenarios used to calculate VaR or Expected Shortfall.

If we assume the transition equation for p_i is simply a random walk:

$$\theta_i = \theta_{i-1} + \nu_i$$

Then – as above – if we assume the variance of γ and ε are constant through time then the Kalman Filter updating equation becomes simply:

$$\hat{\theta}_i = \lambda \hat{\theta}_{i-1} + (1-\lambda) y_i$$

Where here (1- λ) is the ratio of the variance of p_i to the variance of y_i.

Harvey, Andrew C. *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press.