Model Validation Analytics in DROP

v4.02 24 January 2019

Probability Integral Transform

Introduction

1. Definition of Probability Integral Transform: In statistics, the probability integral transform or transformation relates to the results that data values that are modeled as being random variables from any continuous distribution can be converted to random variables having a standard uniform distribution (Dodge (2003), Wikipedia (2018)).
2. Exact vs. Approximate PIT Map: This holds exactly provided that the distribution being used is the true distribution of the random variables; if the distribution is the one fitted to the data, this will hold approximately in large samples.
3. Alternate Target PIT Distributions: The result is sometimes modified or extended so that the result of the transformation is a standard distribution other than the uniform distribution, such as the exponential distribution.

Applications

1. PIT for Statistical Hypothesis Testing: One use for the probability integral transform in statistical data analysis is to provide the basis for testing whether a set of observations can be reasonably modeled as arising from a specified distribution.

2. Using PIT to generate $U[0, 1]$: Specifically, the probability integral transform is applied to construct an equivalent set of values, and a test is then made to check whether a uniform distribution is appropriate for the constructed data set. Examples of this are the P-P plots and the Kolmogorov-Smirnov tests.
3. Copula Tests for Multivariate Data: A second use of the transformation is in the theory related to copulas which are a means of defining and working with distributions for statistically dependent multivariate data.
4. Applying PIT to Marginal Distributions: Here the problem of defining or manipulating a joint probability distribution for a set of random variables is simplified or reduced in apparent complexity by applying the probability integral transform to each of the components and then working with a joint distribution for which the marginal variables have uniform distributions.
5. PIT for Inverse Transform Sampling: A third use is based on applying then inverse of the probability integral transform to convert random variables from a uniform distribution to have the selected distribution; this is known as inverse transform sampling.

Statement and Proof

1. Statement: Suppose a random variable X has a continuous distribution for which the cumulative distribution function (CDF) is F_X. Then the random variable Y is defined as

$$Y = F_X(X)$$

It has a uniform distribution.

2. Proof: Given any random continuous variable X, define

$$Y = F_X(X)$$

Then

$$\begin{aligned} F_Y(y) &= \mathbb{P}[Y \le y] \\ &= \mathbb{P}[F_X(X) \le y] \\ &= \mathbb{P}\left[X \le F_X{}^{-1}(y)\right] \\ &= F_X\left(F_X{}^{-1}(y)\right) \\ &= y \end{aligned}$$

If F_Y is the CDF of a uniform $[0, 1]$ random variable, Y will have a uniform distribution on the interval $[0, 1]$.

Examples

1. <u>Underlying Univariate Distribution – Standard Normal</u>: As an illustrative example, let X be a random variable with the Standard Normal distribution $\mathcal{N}(0, 1)$. Then its CDF is

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt = \frac{1}{2}\left[1 + \text{erf}\left(\frac{x}{\sqrt{2}}\right)\right] \quad x \in \mathbb{R}$$

where $\text{erf}(\cdot)$ is the error function. Then the new random variable Y defined as

$$Y = \Phi(x)$$

is uniformly distributed.

2. <u>Underlying Univariate Distribution - Standard Exponential</u>: If X has an exponential distribution with unit mean, then its CDF is

$$F(x) = 1 - e^{-x}$$

and the immediate result of the probability integral transform is that

$$Y = 1 - e^{-X}$$

has a uniform distribution. The symmetry of the uniform distribution can then be used to show that

$$Y' = e^{-X}$$

also has a uniform distribution.

References

- Dodge, Y. (2003): *The Oxford Dictionary of Statistical Terms* **Oxford University Press**
- Wikipedia (2018): Probability Integral Transform

Basel III Framework for Backtesting Exposure Models

Abstract

1. Standard Practices for IMM Back-testing: A central component of the Basel III (B3) standards is the *Sound Practices for Backtesting* (Basel Committee for Banking Supervision (2010)), i.e., a summary of the strict regulatory guidances on how to validate and back-test Internal Methods Models (IMM) for credit exposure.
2. Statistical Credit Exposure Backtesting Framework: In their work, Anfuso, Karyampas, and Nawroth (2017) define a comprehensive framework to backtest credit exposure models, highlighting the proposed features against the regulatory requirements.
3. Risk Factor Dynamical Evolution Backtesting: Their framework contains four main pillars. First is the *risk factor backtesting*, i.e., assessment of the forecasting ability of the stochastic differential equations (SDEs) used to describe the dynamics of the single factor.
4. Risk Factor Correlation Estimator Backtesting: Next is the *correlations backtesting*, i.e., the assessment of the statistical estimators used to describe the cross-asset evolution.
5. Representative Firm Portfolio Backtesting: Third is the *portfolio backtesting*, i.e., the assessment of the complete exposure model – SDEs + correlations + pricing – for portfolios that are representative of the firm's exposure.
6. Computation of the Capital Buffer: Last is the *computation of the capital buffer*, i.e., the extra amount of capital that the firm should hold if the model framework is not adequate – using the outcomes of the pillars above.
7. Distributional Tests for Collateralized/Uncollateralized: Anfuso, Karyampas, and Nawroth (2017) show with concrete examples in the cases of collateralized and uncollateralized models how to perform distributional tests with respect to different risk metrics.

8. Discriminatory Power Analysis across Forecasting Horizons: They produce discriminatory power analysis for all the tests introduced, providing exact methods to aggregate backtesting results across forecasting horizons.
9. Capital Remedies for Model Deficiencies: Most importantly, the third and the fourth pillars define a sound quantitative approach for computing capital remedies for potential model deficiencies.

Introduction

1. Validating and Backtesting IMM: The central pillar of the Basel III (B3) document is the *Sound Practices for Backtesting* (Basel Committee for Banking Supervision (2010)), i.e., a summary of strict regulatory guidelines on how to validate and backtest Internal Methods Models (IMM) for credit exposure.
2. European Basel III - The CRD4 Requirements: Similarly, a series of requirements for CRD4 – the European equivalent of B3 – indicate and define backtesting and validation as core component for good governance of IMM firms.
3. Importance of IMM for CVA: From a dealer perspective, the new regulatory changes introduced with B3 – e.g. CVA capital charge – have stressed even more the importance of IMM in making the capital costs of the businesses sustainable.
4. Importance of Backtesting for IMM: At the same time, the increasing complexity of the capital framework requires a thorough approach to the validation and the monitoring of the model performance.
5. Support from Regulators and Stake-holders: A sound backtesting methodology is therefore the key tool to both prove to the regulators the soundness of the models and to assure the stakeholders that the capital position of the firm is in sound modeling grounds.
6. Holistic Qualitative/Quantitative Model Assessment: The assessment of a model is a holistic process that has both qualitative and quantitative elements. While the former may have a

decisive weight for the choice of a given model many possible, the latter are the ones to be considered for backtesting.

7. Validation of the NULL Hypothesis: In particular, the performance of a model should be judged in terms of its forecasting ability. In statistical jargon, backtesting should address the question: *Can we reject the NULL hypothesis – i.e., the model – based on the available historical data*?
8. Statistical Testing of the Forecasting Ability: In the definition employed in this chapter, backtesting is therefore a set of statistical tests that measures the forecasting ability of the model using the data history available as comparison.
9. Assessment Metric - Aggregation of p-values: The final metric is based on the aggregation of the given p-values of the single tests, rejecting the model if an *a priori* determined threshold is breached.
10. Framework Features vs. Regulatory Guidelines: This chapter presents a complete framework to backtest credit exposure models. The next section gives a brief overview of the relevant metrics for the counterparty credit risk and summarizes the features of the framework against the new regulatory guidelines.
11. Complete Backtesting Cycle Details: In subsequent sections the methodology is presented in detail for the full backtesting cycle, i.e., the risk factor evolution models, the correlation models, and the portfolio exposure metrics. Later sections show how to compute the capital buffers based on the back-testing results. Finally, conclusions are drawn.

Basic Concepts and the Need for Backtesting

1. Credit Counterparty Exposure - Definition: Credit counterparty exposure is defined as the amount a dealer A could potentially loos in the case that its counterparty B defaults.

2. Estimating the CP Exposure Distribution: The exposure – from A's perspective – is computed from the forecasted distribution of prices of the financial contracts that constitute the portfolio of counterparty B at any future date.
3. Components required for the Exposure Estimation: The main building blocks required for the computation are the following: first, the scenario simulations for the underlying risk factors – generated with what is referred to here as the Risk Factor Evolution (RFE) models – and second, the pricing at each scenario to generate the $MTM(t)$ distribution at any future date t.
4. Expression for the Expected Positive Exposure: The relevant exposure metric from the regulatory perspective is the Expected (Positive) Exposure at time t which is defined as:

$$EE(t) = \mathbb{E}[MTM^{+}(t)]$$

where

$$MTM^{+}(t) = \max(MTM(t), 0)$$

5. RWA for CVA and Capital: The same exposure profile $EE(t)$ enters in the computation of the Risk Weighted Assets (RWA) of a given counterparty both for the CVA and for the default capital charges.
6. Modeling the Risk Factor Evolution: To compute the $MTM(t)$ distribution – and most importantly the $EE(t)$ profile – at any future time t one needs to forecast the evolution of the risk factor values. Those risk factors can also be dependent on each other – *correlation* assumption between the risk factors.
7. Loss Impact of RFE Mis-specification: The more accurately the RFE model is specified, the more realistic is the exposure calculation. If the RFE is mis-specified the exposure figure can be wrongly stated and the losses may occur with higher probability than expected in the case of counterparty defaults.
8. Risk Neutral Measure for the CVA: As observed in Kenyon and Stamm (2012), there is a potential for divergence in the choice for calibration for the RFE models. On the one hand,

the $EE(t)$ profile used for CVA calculation – a *price* – should be based on market calibration.

9. Historical Measure for Capital Charges: On the other hand, default charges require a forecast in the real-world measure and therefore a historical calibration would be most suitable.
10. Complication from the Dual Measures: The backtesting methodology described in the following is agnostic to the choice of the model calibration. Nevertheless, by construction, the regulatory requirements for backtesting models are addressed generally by historical models. In view of Anfuso, Karyampas, and Nawroth (2017), this apparent dichotomy is one of the key quantitative challenges for the industry after Basel III.
11. Regulatory Valuation Validation Framework: Following the guidances from regulators, Anfuso, Karyampas, and Nawroth (2017) define a framework that has four main pillars.
12. Risk Factor Dynamics Backtesting: The first is Risk Factor backtesting, i.e., the assessment of the forecasting ability of the Stochastic Differential Equations (SDEs) used to describe the dynamics of the single risk factors. It can be seen that the calibration of the SDE – market implied or historical – has a crucial influence on this assessment.
13. Risk Factor Correlation Backtesting: Next is the correlations backtesting, i.e., the assessment of the estimators used to model the cross-asset evolution.
14. The Representative Portfolio Backtesting: Third is the portfolio backtesting, i.e., the assessment of the complete exposure model – RFEs + Correlations + Pricing – for portfolios that are representative of the firm's exposure.
15. Model Reserve Capital Buffer Calculation: The computation of the capital buffer, i.e., the extra amount of capital that the form should hold if the model framework is not adequate – see the outcomes of the three pillars above – is the final pillar.
16. Diagnostic vs. Deficiency Remedy Pillars: The first three pillars are diagnostic whereas the fourth comes as a remedy for potential deficiencies in the exposure models.
17. Regulatory Guidance vs. Framework Response: The section below summarizes the relevant regulatory guidances for backtesting and how the framework in this chapter addresses them. The specifics of the proposed solutions are described in more detail in the sections below.

Regulatory Guidances

1. Guidance #1: The performance of the market risk factor should be validated using backtesting. The validation must be able to identify poor performances in individual risk factors.
    a. Methodology => The forecasting capability of the RFE models and their calibrations is back tested at multiple backtesting horizons, making use of different distributional tests.
    b. Compliance => Full.
2. Guidance #2: Validation of the EPE models and all the relevant models that input into the calculation of the EPE must be made using forecasts initialized on a number of historical dates.
    a. Methodology => The sampling of the backtesting is on a bi-weekly frequency, spanning all the available data history. Since the sampling is very close, no statistical bias due to the selection of the sampling frequency is introduced.
    b. Compliance => Full.
3. Guidance #3: Historical backtesting on representative counterparty portfolios and market risk factors must be part of the validation process. At regular intervals, as dictated by its supervisor, the dealer must conduct backtesting on a number of representative counterparty portfolios and the market risk factor models. The representative portfolios must be chosen based on the sensitivity to the material risk factors and correlations to which a dealer is exposed.
    a. Methodology => The portfolio backtesting is performed with suitable metrics that do not penalize the conservative estimates of the EPE. Representative counterparties can be chosen by a given dealer based on their RWA contributions.
    b. Compliance => Fully compliant from a methodology perspective. The dealer should additionally ensure that the selected counterparties are representative.

4. Guidance #4: Backtesting of the EPE and all the relevant models that input into the calculation of the EPE must be based on the recent performance.
    a. Methodology => The framework is agnostic to the data history that is selected, since that is an input. It can therefore be applied to assess recent and longer-term performances – though the statistical significance of the results will not be equivalent given the different amount of historical realizations that are being backtested.
    b. Compliance => Full.
5. Guidance #5: The frequency with which the parameters of an EPE model are updated needs to be assumed as part of the ongoing validation process.
    a. Methodology => The calibration is fully accounted for the RF, the correlations, and the portfolio backtesting.
    b. Compliance => Full.
6. Guidance #6: Dealers need to unambiguously define what constitutes acceptable and unacceptable performance for their EPE models and the models that input into the calculation of the EPE and have a written policy in place that describes how unacceptable performance will be remediated.
    a. Methodology => The framework gives a quantitative probabilistic interpretation of the performance that allows unambiguous acceptance or rejection of a model.
    b. Compliance => Fully compliant from a model perspective. The dealer should additionally ensure that the acceptance threshold is sufficiently conservative.
7. Guidance #7: IMM firms need to conduct hypothetical portfolio backtesting that is designed to test risk factor model assumptions, e.g., the relationships between the tenors of the same risk factors, and the modeled relationships between the risk factors.
    a. Methodology => The chapter does backtest the model correlation assumptions with a coherent extension of the methodology applied to the single risk factors.
    b. Compliance => Full.
8. Guidance #8: Firms must backtest their EPE models and all relevant models that input into the calculations of the EPE out to long horizons of at least one year.
    a. Methodology => Multiple horizons - shorter and longer than one year - are backtested at every level of granularity, i.e., RFs, correlations, and portfolio exposures.

b. Compliance => Full.

9. Guidance #9: Firms must validate their EPE models and all relevant models that input into the calculation of the EPE out to the time horizons commensurate with the maturity of trades covered by the IMM waiver.
    a. Methodology => Multiple horizons commensurate with the maturity if the trades are backtested at every level of granularity, i.e., RF's, correlations, and portfolio exposures.
    b. Compliance => Full.
10. Guidance #10: Prior to the implementation of a new EPE model or a new model that inputs into the calculation of the EPE, a dealer must carry out a beck testing of the EPE model and all relevant models that input into the calculation of the EPE at a number of distinct time horizons using historical data for movements in the market risk factors for a range of historical periods covering a wide range of market conditions.
    a. Methodology => The framework described, because of its granularity and modularity, can be applied for periodic regulatory back testing as well as initial validation of a given model.
    b. Compliance => Full.

RF Backtesting: The Backtesting Construction for Collateralized and Uncollateralized Models

1. Basic Components of the Framework: The RFE models are the most atomic components of the exposure framework.
2. Cross Horizon RF Backtesting: As for regulatory guidances 1, 8, and 9, their performance should be assessed for different forecasting horizons and the predicted distributions should be consistent with the realized history of the corresponding risk factors.

3. Probability Integral Transform (PIT) Definition: The statistical tool that is the basis for the backtesting is the Probability Integral Transform (PIT – Gunther, Diebold, and Tay (1998), Kenyon and Stamm (2012)) defined as

$$F(r_n) = \int_{-\infty}^{r_n} \phi(x) dx$$

where r_n is the realization of the given random variable and $\phi(\cdot)$ is its predicted distribution.
4. Application to the RF Distribution: It is clear that when one applies PIT to a set of i.i.d. variables r_n using the correct distribution of r_n the transformed set

$$y_n = F(r_n)$$

is uniformly distributed.
5. Statistical Metric for Mismatch/Departure: The distance between the transformed set y_n and $U[0, 1]$ distribution – in a statistical sense – can therefore be used as a goodness of the model $\phi(\cdot)$ to describe the random variable r_n.
6. RF Market Generation and Evolution: In practice, the input for RFE backtesting analysis is the time series of the given risk factor and the model – the SDE and its calibration – used to describe its evolution.
7. PIT Map onto $[0, 1]$ Space: The PIT for the predicted distribution can be used to map the realized values of a risk factor – or their variations, see below – to a set of values in the interval $[0, 1]$.
8. The Transformed Set Model Performance Assessment: The transformed set is then is used to assess the model performance by applying the standard distribution tests.
9. Collateralized vs. Uncollateralized Time Grids: For collateralized and uncollateralized models the construction differs because of the presence of the multiple time scales. In both cases a grid of sampling points t_k is used.

10. Criteria for Time Grid Selection: The sampling points define the origin of the backtesting experiment and they should be on a sufficiently fine grid and for a sufficiently long time series so as to ensure the following:
    a. An acceptable discriminatory power for the test – in relation to the size of the data history – and:
    b. An absence of significant statistical bias caused by the sampling sequence – if the sampling points are too distant, the backtesting results from an equivalent sequence with different arbitrary starting point may differ.
11. Horizon used for Backtesting: The backtesting is carried out for an arbitrary set of horizons $\{h_1, \cdots, h_n\}$ and for every h_i a single result is produced.
12. Horizon Matching Firms' Exposure Structure: The choice of the set $\{h_1, \cdots, h_n\}$ should be so as to reflect the portfolio structure of the firm.
13. Uncollateralized RF Model - Time Scale: In case of the uncollateralized RF models the horizon h_i is the only timescale. At

$$t = t_k$$

the forecast forward distribution for

$$t = t_k + h_i$$

is constructed based on he given RF model, and on the filtration $\mathcal{F}(t_k)$ at t_k.
14. PIT Based Transform to the $[0, 1]$ Scale: Using the PIT transform – where $\phi(\cdot)$ is given by the forecast conditional distribution – the realized value of RF at

$$t = t_k + h_i$$

is mapped to a value in the $[0, 1]$ scale.
15. Collateralized RF Model - MPoR Impact: For collateralized models the presence of the margin period of risk ($MPoR$) should be additionally accounted for. The MPoR is the time

required for the dealer to liquidate the collateral a given counterparty posted to finance its exposure.

16. Collateralized RF Model - Primary Focus: Therefore, the primary focus of a collateralized model is to describe the variation of the RF over the MPoR at any future horizon.
17. Distribution of the RF inside the MPoR: In the backtesting exercise $\phi(\cdot)$ is the forecast RF variation distribution in the interval $[t = t_k + h_i, t = t_k + h_i + MPoR]$ conditional on $\mathcal{F}(t_k)$ and the realized value is the historical variation of RF in the same interval.
18. PIT Transformations on RF Realizations: The result of the PIT transformation on the sampling sequence is a set of values $\mathcal{F}\left(r_{t_k}\right)$ in the interval $[0, 1]$.
19. Statistical Properties of the PIT Transformation: At this stage the standard statistical tests are applied to check for the different properties of the RF distribution.
20. Enhanced CDF Generalized Distribution Metric: In particular one can introduce a generalized distance metric as

$$d_w = \int_{\Gamma}^{\Gamma} [F_n(x) - F(x)]^2 w(x) dx$$

where

$$\Gamma = [0, 1]$$

is the domain of $F(x)$, $F_n(x)$ is the empirical cumulative distribution function CDF of the $F\left(r_{t_k}\right)$ values,

$$F(x) = x$$

is the CDF of the $U[0, 1]$ distribution, and $w(x)$ is a weight function that can be chosen so as to emphasize a given quantile domain. In Kenyon and Stamm (2012) it is suggested to link $w(x)$ to the portfolio structure.

21. Individual RF Realizations as Pillars: Conversely, in this chapter the risk factors are backtested independently and the choice of $w(x)$ is unrelated to the portfolio composition. The portfolio backtesting is a separate pillar as described above.
22. Testing across different Weight Distribution: The literature provides many different statistical tests to check the match between the two distributions.
23. Cramer von Mises vs. Anderson Darling: This chapter considers the well-known Cramer-von Mises – CVM with

$$w(x) = 1$$

and Anderson-Darling (AD)

$$w(x) = \frac{1}{F(x)[1 - F(x)]}$$

tests where the first focuses more on the center of the distribution whereas the latter focuses more on the tails.
24. Distance Metric as a Single Value: It can be seen that d_w is a single *distance* value obtained from the realized cost at $\mathcal{F}\left(r_{t_k}\right)$.
25. Monte Carlo based p-Value Estimate: The final outcome of the backtesting should be a p-Value. Therefore, the single realization d_w is assigned a p-Value based on the construction of the corresponding test statistic – distribution of the outcomes of d_w – using Monte Carlo.
26. Generation of the Test Statistic Distribution: The test statistic for a given history, sampling frequency, and horizon is produced by repeating on a large number of simulated paths the back-testing calculation/mapping described above.
27. Calibrated RF Model Path Generation: The paths are generated with the same model/calibration which is going to be used for calculating their corresponding d_w's.
28. Horizon Realized RF Value Distribution: Therefore, the test statistic is the distribution of the expected the correct model.

29. <u>p-value from the d_w Quantiles</u>: With the computed quantile of the realized d_w the p-value for the back-testing can be derived.
30. <u>Handling of the Overlapping Forecasting Horizons</u>: Notice that this derivation of the statistic allows for overlapping forecasting horizons since the auto-correlation among the $F\left(r_{t_k}\right)$ is correctly reflected in the construction.
31. <u>Backtesting Longer Dated Horizons</u>: This feature is particularly useful in backtesting longer horizons – see guidance #9, e.g., long-dated inflation/IR trades – for which the data history is comparatively short in most of the cases.
32. <u>Sample Test - CHFUSD Exchange Rate</u>: Anfuso, Karyampas, and Nawroth (2017) show uncollateralized and collateralized backtesting results for the CHFUSD exchange rate where they apply the CVM backtest to the last 15 years of history with bi-weekly sampling frequency for different forecasting horizons

$$h_i = \{1m, 3m\ 1y\}$$

33. <u>RFE Process Underlying the FX</u>: The RFE is a Geometric Random Brownian Motion (GBM) with drift

$$\mu = 0$$

and volatility calibrated using a rolling 1Y window.
34. <u>Backtesting Collateralized/Uncollateralized Exposures</u>: For the considered case, both the uncollateralized and the collateralized cases

$$MPoR = 2w$$

backtesting gives acceptable results at all horizons.
35. <u>Discriminatory Power PLUS Horizon Aggregation</u>: The next sections analyze two further aspects that were mentioned above – the discriminatory power of the statistical tests and the aggregation of the backtesting results across RF's and horizons.

Discriminatory Power RF Backtesting

1. Data Size Impact on Assessment: The assessment of the RF models depends crucially on the amount of data history.
2. Model Mis-specification Sensitivity to Size: In case of availability of large data sets, the backtesting can resolve very tiny model mis-specifications.
3. Less Data - Easier Backtesting: Conversely if the data is too few the model uncertainty will be larger – especially for longer horizons – and it will be comparatively easier to pass backtesting.
4. Quantitative Analysis of Data Sizes: To have a quantitative understanding of the above described effect, Anfuso, Karyampas, and Nawroth (2017) ran backtesting analysis on the CVM and the AD tests on synthetic data made from 1000 paths for 15 years generated by the same stochastic model – Geometric Brownian Motion with annualize drift and volatility of

$$\mu = 0$$

and

$$\sigma = 10\%$$

5. Impact of Drift/Volatility Mis-specifications: For every one of these paths, the p-value is determined for

$$h_i = \{1m, 3m, 1y\}$$

and for different mis-specifications of μ and σ.

6. Distance Metric as Sensitivity Metric: For a given h_i the average of the p-values across the paths is an intuitive measure of the sensitivity of the test for the given data history w.r.t. a mis-specified model calibration – for the correct model

$$\langle p \rangle = 0.5$$

7. Illustration of the above Analysis: Anfuso, Karyampas, and Nawroth (2017) illustrate the above analysis by highlighting the correctly specified model.
8. Backtesting Sensitivity to Shorter Horizons: As is evident from their tables, the backtesting results are more sensitive at the shorter horizons because of the larger number of independent observations taken into account:

$$h_i = 1m \rightarrow 180$$

vs.

$$h_i = 15y \rightarrow 15$$

independent observations.
9. AD vs. CVM Comparative Performance: It can also be noticed that AD slightly out-performs CVM in detecting model mis-specifications for the constructed example.
10. Discriminatory Power Analysis Regulatory Reporting: In the documentation that a dealer should provide to the regulators for the backtesting of the IMM, the discriminatory power analysis is a useful complementary information to assess the tolerance of the backtesting methodology.

The Aggregation of Backtesting Results

1. Lowest Granularity of Backtesting: The general indication from the regulators is to show the RF backtesting results at the most granular level, i.e., single RF's and horizons.
2. Asset Class and Horizon Aggregation: Nevertheless, the analysis should be complemented with aggregated results that assess more holistically the performance of the exposure models – e.g., by asset class and/or including several horizons.
3. *Boot-strappable* RF Probability Framework: The RF backtesting scheme presented above allows for further aggregations within the same probabilistic framework.
4. Aggregation over multiple RF Horizons: The scheme considered here is for a single RF over multiple horizons, i.e., the aim is to produce a single aggregated result where the importance of the different horizons is given an arbitrary weighting function $\theta(i)$ with

$$\sum_i \theta(i) = 1$$

   and

$$\theta(i) > 0 \; \forall \; i$$

5. Incorporating the Dealer Portfolio Exposure: This case is of relevance in the common situation where the same RF model is used for products/portfolios of very different maturities and the assessment of mode performance should encompass many different time scales.
6. Standard Deviation of the Test Statistic: It can be seen that for a given sampling frequency that the standard deviation of the test statistic distribution scales linearly with the horizon. This is a direct consequence of the auto-correlation among back-testing experiments at different sampling points that grows linearly with the length of the horizon.
7. Standard Deviation for GBM AD/CVM: Anfuso, Karyampas, and Nawroth (2017) illustrate this behavior for both AD and CVM tests where the standard deviation has been determined numerically for the case of GBM for different forecasting horizons.

8. Horizon Normalization of the Standard Deviation: As a consequence of the linearity, the test normalized test distances

$$\bar{d}_{w,i} = \frac{d_{w,i}}{h_i}$$

with the normalization given by the forecasting horizon h_i - are measured in equivalent units.

9. Realized Historical Standard Deviation Metric: One can therefore define a single distance

$$d_{w,AGG} = \sum_{i} \theta(i)\bar{d}_{w,i}$$

with the desired weightings across the horizons.

10. Realized Standard Deviation Metric: Therefore, the realized historical value for $d_{w,AGG}$ is derived straightforwardly from $\bar{d}_{w,i}$ and $\theta(i)$.
11. Path-wise Test Statistic Distribution: The corresponding distribution of the test statistic can be computed by applying path by path the given test at different horizons and aggregating $d_{w,AGG}$ using the equation above.
12. Equi-Weighted Horizon Discriminatory Analysis: Anfuso, Karyampas, and Nawroth (2017) show the discriminatory power analysis – shown in the previous section – for the case of equally weighted aggregation across horizons

$$\theta(i) = \frac{1}{N_h}$$

where N_h is the number of horizons considered.

Correlations Backtesting

1. Basel III CRD Correlation Backtesting Estimator: One of the novelties introduced by Basel III and CRD4 for backtesting is the explicit requirement to check the correlation estimator – see Guidance #7.
2. Capturing Bear Market Correlated Moves: This is an important ingredient of the exposure framework, especially in periods of extreme bear markets when correlations rise significantly (Nawroth, Anfuso, and Akesson (2014)).
3. Enhancing PIT for Correlations Testing: While the previous sections were about the performance of single RF models, this section defines a method to measure how well a correlated set of SDE's for multiple RF's describe the RF's co-movements. As will be seen the PIT methodology can be suitably generalized for this purpose.
4. Backtesting Pair-wise Correlations: For a set of N RF's the aim is to backtest $\frac{N(N-1)}{2}$ correlations among all the – upper or lower – off-diagonal pairs.
5. Correlation Matrix for N GBM's: To explain the methodology consider the simplest model framework, i.e., N GBM's with a given – e.g. historical – calibration for $\vec{\mu}$, $\vec{\sigma}$, and the correlation matrix $[\rho]$.
6. Synthetic Consolidated Random Factor Consolidation: Defining the process for RF_i as

$$RF_i(t) = RF_i(0)e^{\mu_i t+\sigma_i W(t)-\frac{1}{2}\sigma_i{}^2 t}$$

for every non-equivalent pair $\{i, j\}$ one can introduce the synthetic RF Z_{ij} as follows:

$$Z_{ij}(t) = RF_i(t)^{\frac{1}{\sigma_i}} RF_j(t)^{\frac{1}{\sigma_j}} e^{\frac{1}{2}(\sigma_i+\sigma_j)t-\frac{1}{2}(2+2\rho_{ij})-\left(\frac{\mu_i}{\sigma_i}+\frac{\mu_j}{\sigma_j}\right)t}$$

7. Volatility of the Synthetic Random Factor: By construction Z_{ij} is a drift-less GBM with volatility given by

$$\sigma_{ij} = \sqrt{2 + 2\rho_{ij}}$$

8. Computing the Historical Z_{ij} Realizations: The historical realizations of Z_{ij} can be obtained at every sampling point using the estimators for the marginal distributions of RF_i and RF_j - i.e., the volatilities and the drifts $\sigma_i(t_k)$, $\sigma_j(t_k)$, $\mu_i(t_k)$, and $\mu_j(t_k)$, and the correlations among RF_i and RF_j.
9. Z_{ij} Volatility Measures ρ_{ij}: Z_{ij} can be therefore backtested as was done for single RF's, but its volatility is a direct measure of the correlation to be verified.
10. Inadequate Marginal Distribution Estimators: It can be seen that if the estimators of the marginal distributions are inadequate Z_{ij} is likely to fail backtesting independently of ρ_{ij}.
11. Consequence of Inadequate Marginal Estimations: This feature is not a drawback but rather a desirable property given the regulatory purpose of the backtesting analysis.
12. Need for Valid Marginal RF's: The RFE models are perceived as the atomic components of the exposure framework while the correlations are the second layer.
13. Valid Z_{ij} Estimators and Invalid Marginals: Whenever the underlying RF models fail the information on the correlation performance is of little value from a regulatory perspective.
14. SnP500 vs CHFUSD Correlation Tests: In their illustrations Anfuso, Karyampas, and Nawroth (2017) present uncollateralized backtesting results for SnP500 index and CHFUSD exchange rate.
15. GBM Martingale vs 1Y Rolling Volatility: They consider a correlated GBM with

$$\mu = 0$$

    and the correlation ρ estimated using a 1Y rolling window.
16. Negative Correlation between SnP00 and USDCHF: The correlation is mostly negative, especially in correspondence of the lows of the SnP500.
17. Multi Horizon Correlations Backtesting: At the three-time horizons considered

$$h_i = \{1m, 3m, 1y\}$$

and for a data history of 15 years the correlation model passes backtesting at

$$CL = 99\%$$

- the two tickers pass RF backtesting independently for the same set of horizons. Results for the single factors are discussed earlier.

18. Reusing Single Factor Discriminatory Power Analysis: The mapping to the single RF problem allows the inheritance of all the results derived in that context – e.g., the discriminatory power analysis can be obtained as for single RF as is illustrated by Anfuso, Karyampas, and Nawroth (2017).
19. Collateralized/Uncollateralized Backtesting and Metric Aggregation: In particular the collateralized and the uncollateralized models can both be backtested based on the single RF factor scheme, and the correlation backtesting results can be aggregated at different horizons using the method discussed in the previous section.
20. Aggregation across a Single Correlator: Given the large number of entries in the correlation matrix for a given scenario, it is very convenient also to aggregate results across correlation elements – to a given forecasting horizon.
21. Block Level Correlation Metric Aggregation: A powerful visualization of the correlation testing is, e.g., aggregation by asset classes and the assignment of a given p-value for every block of the correlation matrix.
22. Block Level CVM/AD PIT: To obtain such a result, the aggregated distance for a given subset of the correlation matrix

$$\Omega = \{i \in \alpha, j \in \beta\}$$

can be obtained by applying the chosen distribution tests (CVM or AD) to the union of all the PIT's of the synthetic RF's

$$Z_{ij} \in \Omega$$

23. Model Comparison using the Departure Metric: The corresponding test statistic distribution can be derived with the model vs. model approach described in the previous sections, considering the correlated paths of the RF's in every scenario in Ω.
24. Cross RF Aggregation Technique: Anfuso, Karyampas, and Nawroth (2017) provide a detailed illustration of such an example of the aggregation methodology.

Portfolio Backtesting

1. Backtesting the Underlying RF's: The discussion so far has focused on the backtesting of the underlyings.
2. Primary Focus of the Regulators: However, the primary regulatory focus is on the performance on the overall regulatory framework, i.e., the ability of the dealer's IMM models to assess the RWA – i.e., the $EE(t)$ profile and hence the capital – accurately or conservatively – see Guidance #3.
3. RF vs. Regulatory Backtesting: It is obvious that the portfolio backtesting is conceptually different from the RF's and the correlations backtesting precisely owing to the above statement.
4. Tolerance for Overstating the Exposure: While in the case of individual RF's and correlations the model that is systematically understating or overstating a certain quantile domain is expected to fail, for portfolios and model feature that leads to systematically overstating the $EE(t)$ should not be penalized – at least from a regulatory perspective.
5. Asymmetry Adjusted PIT Capital Testing: The proposal of the third pillar of the framework – the portfolio backtesting – is again based on the PIT, with suitable modifications that account for the asymmetry discussed above.
6. Steps Involved in the Methodology: The methodology applies to both collateralized and uncollateralized books, and comprises of the following steps.

7. Scheme Based Counterparty Identification: The identification of a set of counterparties is based on, e.g., the B3 Default Capital RWA.
8. PIT for MM and r_t:The construction of the empirical uniform with the PIT using

$$F(r_n) = \int_{-\infty}^{r_n} \phi(x) dx$$

   where $\phi(\cdot)$ is given by the forecasted MTM distribution for uncollateralized counterparties – or by the ΔMTM distribution for the collateralized ones – and r_t by the realized values for uncollateralized counterparties – or by realized MTM variation for collateralized ones.
9. Backtesting the Horizon-Appropriate Portfolio: At every sampling point

$$t = t_k$$

   the composition of the portfolio should be the correct historical one. If no information on the historical trade composition is available one can simply backtest the current portfolio in a manner similar to what is discussed in the next section. The realized MTM value is obtained by re-pricing the same portfolio at

$$t = t_k + h$$

10. Statistically Conservative Test Portfolio #1: The statistical analysis of the $F\left(r_{t_k}\right)$ set based on a test with a notion of conservatism embedded is referred to as the conservative portfolio CPT.
11. Statistically Conservative Test Portfolio #2: The CPT is defined as

$$CPT = \int_{\Gamma}^{\Gamma} [\max(F(x) - F(x_n), 0)]^2 w(x) dF(x)$$

where Γ and $F(x)$ are defined as for

$$d_w = \int_{\Gamma}^{\Gamma} [F_n(x) - F(x)]^2 w(x) dF(x)$$

where $F_n(x)$ is the empirical CDF from the forecasted MTM distribution and $w(x)$ is a weight function.

12. Affirmation of the Conservative Behavior: For a given quantile the $\max(\cdot, 0)$ function ensures that no test distance is accrued when the empirical uniform is *more conservative* than the theoretical one.
13. Considering Strictly Positive Exposures: In the exposure language if

$$F_n(x) < F(x)$$

the contribution of the quantile x to EE is greater or equal to the value from the exact model – equality holds in the case of x being negative when the contribution is zero.

14. Sample vs. Distribution Departure Characteristic: As shown in Anfuso, Karyampas, and Nawroth (2017), in such cases as the mis-specification of the volatility of the MTM distribution, $F(x)$ and $F_n(x)$ can have multiple crossing points.
15. Low/High MTM Volatility Impact: This implies that the lower quantiles of the estimation may be conservatively estimated while the higher values are below the correct values for low MTM volatilities. The inverse is true for high MTM volatilities.
16. Mean of the MTM Distribution: Additionally, the EE sensitivity of the MTM volatility depends on the mean of the MTM distribution.
17. IMT/ATM/OTM MTM Characteristic: Again, as shown in Anfuso, Karyampas, and Nawroth (2017), in the two limiting cases of the MTM distribution – deep in and deep out of the money, the EE is almost independent of the MTM volatility – as opposed to the more intuitive linear dependence for an MTM distribution centered at 0.
18. Differential Ranking of the EE Quantiles: For the determination of $w(x)$ the different quantiles should be ranked – for e.g., - their importance to the EE must be quantified.

19. Metric for the Quantile Distribution Importance: Their relative contribution$\alpha(q)$ can be defined as

$$\alpha(q) = \underset{\Delta q \to 0}{Limit} \max\left(\frac{1}{\Delta q}\int_{\Phi^{-1}(q)}^{\Phi^{-1}(q+\Delta q)} x\phi(x)dx\,,0\right)\cdot\frac{1}{EE} = \frac{1}{EE}\max(\Phi^{-1}(q),0)$$

where $\phi(x)$ and $\Phi^{-1}(x)$ are the probability density function and the inverse of the cumulative density function of a reference MTM distribution.

20. Application of the Mean Value Theorem: The identity in the above equation can be derived by applying the mean value theorem in the limit

$$\Delta q \to 0$$

21. Normal MTM Distribution - Relative Contribution: It can be seen that in the case of a normal MTM distribution, $\alpha(q)$ has the following closed-form solution:

$$\alpha(q,\mu,\sigma) = \frac{\max\left({\Phi_{\mathcal{N}(\mu,\sigma)}}^{-1}(q),0\right)}{EE_{\mathcal{N}(\mu,\sigma)}} = \frac{\max\left(\mu+\sqrt{2\sigma^2}EF^{-1}(2q-2),0\right)}{\mu\Phi_{\mathcal{N}(0,1)}\left(\frac{\mu}{\sigma}\right)+\sigma\phi_{\mathcal{N}(0,1)}\left(\frac{\mu}{\sigma}\right)}$$

where $EF^{-1}(x)$ is the inverse of the error function, and μ and σ are the mean and the volatility of the Normal Distribution.

22. Same μ but varying σ: Anfuso, Karyampas, and Nawroth (2017) show $\alpha(q,\mu,\sigma)$ for a set of normal MTM distributions with equal volatility but different reasons.
23. Shape/Slope Dependence on μ: The shape and the slope of α vary dramatically with the level of MTM.
24. Par Contract α Volatility Dependence: It can be seen though that for

$$\mu = 0$$

the standard case of collateralized portfolio $\alpha(q, \mu, \sigma)$ is independent of the volatility, i.e.

$$\alpha(q, \mu = 0, \sigma) = \bar{\alpha}(q)$$

25. Elliptical Distribution μ Scale Invariance: This additional scale invariance property is quite robust and holds beyond the normality assumption, i.e., for all members of the elliptical family.
26. Inadequacy of a Single $w(x)$: For real portfolios the level of MTM varies across a wide spectrum of values and it is unlikely that a single choice of $w(x)$ can be optimal across all cases.
27. Special Case - Fully Collateralized Portfolios: In their analysis their focus has been on an MTM distribution centered at 0 – given also its special relevance for collateralized portfolios – and the case

$$w(x) = \bar{\alpha}(x)$$

is considered.
28. Application of the Appropriate $w(x)$: The dealer can fine-tune its own representation $w(x)$ by looking, for e.g., to the historically realized MTM's for the set of portfolios to be backtested.
29. Test Statistic appropriate for Portfolio Backtesting: A final remark is in order for the derivation of the test statistic for portfolio backtesting.
30. Multiple Times/MTM Sequence Simulation: The exact methodology – outlined for the RF's and correlations – would imply that every portfolio would have to compute its test statistic using two nested Monte Carlo simulations – one for the path of the underlyings, and one to generate the conditional MTM distribution at every sampling point and along every path.
31. Feasibility of the Dual Sequence Simulations: This approach can become computationally very expensive and requires a sophisticated parallel implementation on a cluster for large portfolios.

32. Options for avoiding Dual Simulations: If that is not feasible the following two alternative options are available.
33. Portfolios for Proxying Key Trades: For every representative portfolio select the most relevant trade and backtest only for the selected set – deriving the statistic exactly but for a much smaller portfolio.
34. Portfolio for Proxying Key RF's: Derive a representative statistic – used for all representative portfolios – based on an archetypal portfolio with trade composition and level of auto-correlation across sampling dates that are representative across of the set of portfolios to be backtested.

Capital Buffer Calculation

1. Basel III Compliant Backtesting Assessment: The previous sections have shown how to run a Basel III compliant backtesting and produce granular assessments for different IMM components.
2. Operational/Capital Remediation to Model Under-performance: Once that diagnosis is completed, in the case of unsatisfactory performance, the regulators expect the following steps from the dealer:
    a. A feedback loop so as to improve the models based on backtesting results
    b. An intermediate remediation action to account for potential shortages of capital to account for model deficiencies.
3. Capital Buffer Calculation – Requirement: This section discusses the fourth pillar of the framework, i.e., the calculation of the capital buffer (CB).
4. Capital Buffer Calculation - Edge Cases: The CB should be able to efficiently interpolate the two limits of a perfect forecasting model and its opposite, i.e., the case of a completely inadequate estimation of the regulatory capital.

5. Capital Buffer Edge Case Estimates: In the former situation the CB is equal to 0, while in the latter the sum of the estimated RWA and the buffer is bounded by regulatory capital established with standard rules – the conservative regulatory guidances that non-IMM dealers should follow for the calculation of the RWA.
6. Penalty for Model Mis-specification: Additionally, regulators expect the capital buffer to be punitive, i.e.,

$$RWA_{WM} \times (1 + CB_{WM}) \geq RWA_{RM}$$

   where WM and RM stand for *wrong* and *right* models and the CB has been defined as the multiplicative factor be applied to the IMM RWA.
7. Mandatory Penalty for Model Mis-specification: The role of the above inequality is to ensure that the dealer does not have any advantage in using the *wrong* models, i.e., that there is a capital incentive in adopting an adequate exposure framework.
8. Uniqueness of the Capital Buffer Estimation: The features of the CB stated above do not characterize it in a unique way.
9. Dependence on the Backtesting Performance: In the current construction the CB is linked to the performance of the portfolio backtesting for a selected set of N representative counterparties and is calculated historically by comparing the forecasted against the realized exposure profiles for these counterparties.
10. Interpretation as IMM Model Corrector: The final number can be interpreted as a correction factor to be applied to the IMM capital so as to account for the mis-specifications in the determination of the $EEPE$, i.e., the IM component of the RWA.
11. Importance of the CP Representative Portfolios: The capital buffer is entirely based on the representative set but is then applied to the whole portfolio. Therefore, the representative ness of the selected counterparties is a necessary feature to obtain a meaningful capital buffer.
12. Definition of the Model Error Metric: For a given counterparty c the following error metric is defined:

$$\Delta EE_c(t_1, t_2) = \max(MTM_c(t_2), 0) - EE_c(t_1, t_2)$$

where $\max(MTM_c(t_2), 0)$ is the realized exposure at t_2 and $EE_c(t_1, t_2)$ is the EE at t_2 as forecast at t_1.

13. 1Y Average of the Error Metric: The average of $\Delta EE_c(t_1, t_2)$ over one year is given by

$$\Delta EE_c(t_1) = \mathbb{E}_{t_2 \in [t_1, t_1 + 1Y]}[\Delta EE_c(t_1, t_2)]$$

and is the error over the whole horizon of the profile that is relevant for the $EEPE$ calculation.

14. The Capital Buffer Backtesting Metric Setup: As discussed above the capital buffer should be dependent on the portfolio backtesting performance.
15. Error Threshold Based Weight Function: This input enters into the calculation of the capital buffer via the following weighting function:

$$\mathcal{T}_c = \min\left(\frac{\max(p_c - p_l, 0)}{p_u - p_l}, 1\right)$$

where p_c is the p-value of the portfolio backtesting at

$$h = 1Y$$

- the most relevant horizon for the regulatory capital – for the counterparty c and p_l and p_u are the given lower and upper thresholds, i.e., 95% and 99%.

16. $\mathcal{T}_c$ Behavior over the p-Range: $\mathcal{T}_c$ is 0 below p_l and increases monotonically from 0 to 1 in the interval $[p_l, p_u]$.
17. Backtest Range: Satisfactory to Sufficient: p_u can be seen as the failure threshold for the portfolio backtesting while p_l is the lower bound of the p-values region where the backtesting performance may be considered unsatisfactory.

18. Fluctuation of Capital Buffer Estimates: In practice the capital buffer is calculated by IMM dealers on a quarterly basis and it is a desirable feature to not have large fluctuations in its size.
19. Implicit Smoothening inside the p-Range: Defining the relevant buffer region as an interval instead of as a binary threshold, the value of the buffer can smoothly account for potential deteriorations of the backtesting performance, rather than jumping suddenly to a much higher value if, e.g., a few top counterparties cross the failure threshold.
20. Time Horizon Capital Buffer: Having introduced $\mathcal{T}_c$ one can calculate the buffer for a given historical t_i as a p-value weighted across the N representative portfolios:

$$K(t_i) = \frac{\sum_{c=1}^{N} \mathcal{T}_c \Delta EE_c(t_i)}{\sum_{c=1}^{N} EEPE_c(t_i)}$$

    where the sum of the forecasted $EEPE$'s in the denominator makes $K(t_i)$ a unit-less estimate for the relative error in the $EEPE$ framework.
21. Averaging the Capital Buffer Threshold: As a last step $K(t_i)$ can be averaged over the available history with either overlapping or non-overlapping samples.
22. Caveat - Include Only Positive Averages: The final result for the capital buffer is given by

$$CB = \max\left(\mathbb{E}_{t_i}[K(t_i)], 0\right)$$

    where the $\max(\cdot, 0)$ function ensures that only the positive corrections apply, i.e., the negative corrections are not applied to the RWA.
23. Illustration of Backtesting with Capital Buffer: Anfuso, Karyampas, and Nawroth (2017) provide a complete series of portfolio backtesting and capital buffer calculations for synthetic MTM paths generated with Brownian motion for a portfolio of 20 independent counterparties. This results in several observations.
24. CPT vs. CVM or AD: First the discriminatory power of CPT is lower because less information about the MTM distribution is processed by the test in comparison with CVM or AD.

25. Finiteness of the Capital Buffer: The capital buffer is finite even in the case of correct model specification as a consequence of the finiteness of the history and of the $\max(\cdot, 0)$ function in

$$CB = \max\left(\mathbb{E}_{t_i}[K(t_i)], 0\right)$$

26. Full Cycle Capital Buffer Estimation: A full cycle calculation of the capital buffer shows that the inequality of

$$RWA_{WM} \times (1 + CB_{WM}) \geq RWA_{RM}$$

is fulfilled and therefore the buffer accounts for the missing capital due to the use of the wrong model.

27. Incorporating the Regulatory Upper Limit: While Anfuso, Karyampas, and Nawroth (2017) do not include the upper bound, i.e., the mandatory regulatory capital upper bound, this can be easily fixed by imposing

$$EEPE = \min(EEPE_{TOTAL}, EEPE_{REGULATORY})$$

Conclusion

1. Framework for Basel III Backtesting: This chapter includes a complete framework to backtest IMM according to the new Basel III guidelines.
2. Component of the Backtesting Scheme: The methodology includes a diagnosis of the models contributing to the CCR exposure – RF's, correlations, and portfolios – and a remedy for the potential model deficiencies impact on the regulatory capital.

3. Applications of the Backtesting Framework: It can be seen that the very same framework can be used as a template for:
    a. The development phase and the criteria that a model should meet prior to regulatory submission for the IMM waiver
    b. The internal one-off validation of a given model
    c. The periodic – e.g., quarterly – backtesting that the dealer should provide the regulators
4. Dealer's Internal Model Governance Scheme: A unified approach for validation and model backtesting is strongly endorsed by Basel III and CRD4 as it greatly simplifies the internal governance of the dealer.
5. Pre-requisite for IMM Model Approval: Additionally, in the new regulatory environment, the model developers should account for the backtesting requirements at the earliest stage, since strong backtesting performance is a key pre-requisite for IMM waiver approval.
6. Closed/MC/Rule Based Schemes: From a model perspective the methodology described above can be applied equally to Monte Carlo, historical, or rule-based CCR engines.
7. Enhancement to the Underlying Model Dynamics: The illustrations were based on GBM given their simplicity and relevance across the industry. Nevertheless, any other model can be backtested following the same logical steps.
8. Availability of Historical Data as a Shortcoming: In all cases the main bottleneck for a sound backtesting is the availability of sufficient historical data for statistically significant results.

References

- Anfuso, F., D. Karyampas, and A. Nawroth (2017): A Sound Basel III Compliant Framework for Backtesting Credit Exposure Models
- Basel Committee on Baking Supervision (2010): Sound Practices for Back-testing Counterparty Credit Risk Models

- Kenyon, C., and R. Stamm (2012): *Discounting, LIBOR, and Funding: Interest Rate and Credit Pricing* **Palgrave Macmillan**
- Nawroth, A., F. Anfuso, and F. Akesson (2014): Correlation Breakdown and the Influence of Correlations on VaR

Initial Margin Backtesting Framework

Abstract

1. Mandatory Margins for OTC Transactions: The introduction of mandatory margining for bilateral OTC transactions is significantly affecting the derivatives market, particularly in light of the additional funding costs that financial institutions could face.
2. Initial Margin Forecast Models Backtest: This chapter details the approach by Anfuso, Aziz, Loukopoulos, and Giltinan (2017) for a consistent framework, applicable equally to cleared and non-cleared portfolios, to develop and backtest forecasting models for initial margin.

Introduction

1. BCBS-IOSCO Mandatory Margining Guidelines: Since the publication of the new Basel Committee on Banking Supervision and the International Organizations of Securities Commissions (BCBS-IOSCO) guidance for mandatory margining for non-cleared OTC derivatives (Basel Committee on Banking Supervision (2015)) there has been a growing interest in the industry regarding the development of dynamic initial margin models (DIM) – see, for example, Green Kenyon (2015), Andersen, Pykhtin, and Sokol (2017b). By *DIM model* this chapter refers to any model that can be used to forecast portfolio initial margin requirements.

2. <u>Protection Afforded by BCBS-IOSCO</u>: The business case for such a development is at least two-fold. First, the BCBS-IOSCO IMR (B-IMR) rules are expected to protect against potential future exposure at a high-level of confidence (99%) and will substantially affect funding costs, XVA, and capital.
3. <u>IM and VM Based Margining</u>: Second, B-IMR has set a clear incentive for clearing; extensive margining in the form of variation margin (VM) and initial margin (IM) is the main element of the central counter-party (CCP) risk management as well.
4. <u>IMR Impact on Bilateral + Cleared</u>: Therefore, for both bilateral and cleared derivatives, current and future IMR significantly affects the probability and the risk profile of a given trade.
5. <u>B-IMR Case Study - Performance Evaluation</u>: This chapter considers B-IMR as a case study, and shows how to include a suitably parsimonious DIM model on the exposure calculation. It also proposes an end-to-end framework and also defines a methodology to backtest model performance.
6. <u>Organization of this Chapter</u>: This chapter is organized as follows. First, the DIM model for forecasting future IMR is presented. Then methodologies for two distinct levels of back-testing analysis are presented. Finally, conclusions are drawn.

How to Construct a DIM Model

1. <u>Applications of the DIM Model</u>: A DIM model can be used for various purposes. In the computation of the counter-party credit risk (CCR), capital exposure, or credit valuation adjustment (CVA), the DIM model should forecast, in a path-by-path basis, the amount of posted and received IM at any revaluation point.
2. <u>Path Specific IMR Estimation</u>: For this specific application, the key ability of the model is to associate a realistic IMR to any simulated market scenario based on a mapping that makes use of a set of characteristics of the path.

3. RFE Dependence on the DIM: The DIM model is *a priori* agnostic to the underlying risk factor evolution (RFE) models to generate the exposure paths (as shall be seen, dependencies may arise, if for example, the DIM is computed on the same paths that are generated for the exposure).
4. Cross-Probability Measure IMR Distribution: It is a different story if the goal is to predict the IMR distribution (IMRD) at future horizons, either in the real-world P or the market-implied Q measures.
5. IMRD Dependence on the RFE: In this context, the key feature of the model is to associate the right probability weight with a given IMR scenario; hence the forecast IMRD also becomes a measure of the accuracy if the IMRD models (which ultimately determine the likelihood of different market scenarios).
6. P vs. Q Measure IMRD: The distinction between the two cases will become clear later on, in the discussion of how to assess model performance.
7. ISDA SIMM BCBS IOSCO IM: The remainder of this chapter considers the BCBS-IOSCO IM as a case study. For the B-IMR, the current industry proposal is the International Swaps and Derivatives Association Standard Initial Margin Model (SIMM) – a static aggregation methodology to compute the IMR based on first-order delta-vega trade sensitivities (International Swaps and Derivatives Association (2016)).
8. Challenges with SIMM Monte Carlo: The exact replication of SIMM in a capital exposure or an XVA Monte Carlo framework requires in-simulation portfolio sensitivities to a large set of underlying risk factors, which is very challenging in most production implementations.
9. Andersen-Pykhtin-Sokol IM Proposal: Since the exposure simulation provides the portfolio mark-to-market (MTM) on the default (time t) and closeout (time $t + MPoR$, where $MPoR$ is the *margin period of risk*) grids, Andersen, Pykhtin, and Sokol (2017b) have proposed using this information to infer path-wise the size of any percentile of the local $\Delta MTM(t, t + MPoR, Path_i)$ distribution, based on a regression that uses the simulated portfolio $MTM(t)$ as a regression variable.
10. Andersen-Pykhtin-Sokol Proposal Assumptions: The

$$\Delta MTM(t, t + MPoR) = MTM(t + MPoR) - MTM(t)$$

distributed is constructed assuming that no cash flow takes place between the default and the closeout. For a critical review of this assumption, see Andersen, Pykhtin, and Sokol (2017a).

11. Enhancing the Andersen-Pykhtin-Sokol Model: This model can be further improved by adding more descriptive variables to the regression, e.g., values at the default time of the selected risk factors of the portfolio.
12. Optimization: Re-using Exposure Paths: For the DIM model, the following features are desirable. First the DIM should consume the same number of paths as the exposure simulation, to minimize the computational burden.
13. DIM Optimization – B-IMR SIMM Reconciliation: Second, the output of the DIM model should reconcile with the known IMR value for

$$t = 0$$

i.e.

$$IM(Path_i, 0) = IMR_{SIMM}(0)$$

for all i.

14. Key Aspects of IOSCO/SIMM: Before proceeding, this section notes some of the key aspects of the BCBS-IOSCO margining guidelines, and, consequently, of the ISDA SIMM Model (International Swaps and Derivatives Association (2016)).
15. Andersen-Pykhtin-Sokol Proposal Assumptions: First, the $MPoR$ for the IM calculation of a daily margined counter-party is 10 BD. This may differ from the capital exposure calculation, in which, for example

$$MPoR = 20\ BD$$

if the number of trades in the portfolio exceeds 5,000.

16. No Netting across the Asset Classes: Second, the B-IMR in the Basel Committee on Banking Supervision (2015) prescribes calculating the IM by segregating trades from different asset classes. Tis feature is reflected in the SIMM model design.
17. SIMM Methodology Market Volatility Independence: Finally, the SIMM methodology consumes trade sensitivities as its only inputs and has a static calibration that is not sensitive to market volatility.
18. Regression on the ΔMTM Distribution: For the IM calculation, the starting point is similar to that of Andersen, Pykhtin, and Sokol (2017a), i.e.
    a. A regression methodology based on path's $MTM(t)$ is used to compute the moments of the local $\Delta MTM(t, t + MPoR, Path_i)$ distribution, and
    b. $\Delta MTM(t, t + MPoR, Path_i)$ is assumed to be a given probability distribution that can be fully characterized by its first two moments – the drift and the volatility. Additionally, since the drift is immaterial over the $MPoR$ horizon, it is not computed and set to 0.
19. Quadratic Regressor for Local Volatility: There are multiple regression schemes that can be used to determine the local volatility $\sigma(i, t)$ The present analysis follows the standard American Monte Carlo literature (Longstaff and Schwartz (2001)) and uses a least-squares method (LSM) with a polynomial basis:

$$\sigma^2(i, t) = \mathbb{E}[\Delta MTM^2(i, t) \mid MTM(i, t)] = \sum_{k=0}^{n} a_{\sigma k} MTM^k(i, t)$$

$$IM_{R/P,U}(i, t) = \Phi^{-1}\big(0.99/0.01, \mu = 0, \sigma = \sigma(i, t)\big)$$

where R/P indicates received and poste, respectively. In this implementation, the n in

$$\sigma^2(i, t) = \mathbb{E}[\Delta MTM^2(i, t) \mid MTM(i, t)] = \sum_{k=0}^{n} a_{\sigma k} MTM^k(i, t)$$

is set equal to 2, i.e., a polynomial regression of order 2 is used.

20. Calculating the Unnormalized IM Value: The unnormalized posted and received $IM_{R/P,U}(i,t)$ and calculated analytically in

$$\sigma^2(i,t) = \mathbb{E}[\Delta MTM^2(i,t) \mid MTM(i,t)] = \sum_{k=0}^{n} a_{\sigma k} MTM^k(i,t)$$

$$IM_{R/P,U}(i,t) = \Phi^{-1}\left(0.99/0.01, \mu = 0, \sigma = \sigma(i,t)\right)$$

by applying the inverse of the cumulative distribution $\Phi^{-1}(x,\mu,\sigma)$ to the appropriate quantiles; $\Phi(x,\mu,\sigma)$ being the probability distribution that models the local $\Delta MTM(t, t + MPoR, Path_i)$.

21. Note on the Distributional Assumptions: The precise choice of Φ does not play a crucial role, since the difference in the quantiles among the distribution can be compensated in calibration be applying the appropriate scaling factors (see the $\alpha_{R/P}(t)$ functions below). For simplicity, in the below Φ is assumed to be normal.

22. Comparative Performance of the LSM: It is observed that the LSM method performs well compared to the more sophisticated kernel methods such as Nadaraya-Watson, which is used in Andersen, Pykhtin, and Sokol (2017a), and it has the advantage of being parameter free and cheaper from a computational stand-point.

23. Applying $t = 0, MPoR$ and SIMM Reconcilers: The next step accounts for the

$$t = 0$$

reconciliation as well as the mismatch between SIMM and the exposure model calibrations – see the corresponding items above.

24. De-normalizing using IM Scaling Parameters: These issues can be tackled by scaling $IM_{R/P,U}(i,t)$ with suitable normalization functions $\alpha_{R/P}(t)$:

$$IM_{R/P}(i,t) = \alpha_{R/P}(t) \times IM_{R/P,U}(i,t)$$

$$\alpha_{R/P}(t) = [1 - h_{R/P}(t)] \times \sqrt{\frac{10\ BD}{MPoR}} \times \left[\alpha_{R/P,\infty} + \left(\alpha_{R/P,0} - \alpha_{R/P,\infty}\right)e^{-\beta_{R/P}(t)t}\right]$$

$$\alpha_{R/P,0} = \sqrt{\frac{MPoR}{10\ BD}} \times \frac{IMR_{R/P,SIMM}(t=0)}{q\left(0.99/0.01\,, \Delta MTM(0, MPoR)\right)}$$

25. Differential Calibration for Posted/Received IM: In

$$\alpha_{R/P}(t) = [1 - h_{R/P}(t)] \times \sqrt{\frac{10\ BD}{MPoR}} \times \left[\alpha_{R/P,\infty} + \left(\alpha_{R/P,0} - \alpha_{R/P,\infty}\right)e^{-\beta_{R/P}(t)t}\right]$$

$$\beta_{R/P}(t) > 0$$

and

$$h_{R/P}(t) < 1$$

with

$$h_{R/P}(t=0) = 0$$

are four functions to be calibrated – two for received and two for posted IM's. As will become clearer later in this chapter, the model calibration generally differs for received and posted DIM models.

26. Scaling IM using RFE MPoR: In

$$IM_{R/P}(i,t) = \alpha_{R/P}(t) \times IM_{R/P,U}(i,t)$$

$$\alpha_{R/P}(t) = \left[1 - h_{R/P}(t)\right] \times \sqrt{\frac{10\ BD}{MPoR}} \times \left[\alpha_{R/P,\infty} + \left(\alpha_{R/P,0} - \alpha_{R/P,\infty}\right)e^{-\beta_{R/P}(t)t}\right]$$

$MPoR$ indicates the $MPoR$ relevant for the Basel III exposure. The ratio of $MPoR$ to 10 BD accounts for the VM vs. IM margin period, and it is taken as a square root because the underlying models are typically Brownian, at least for short horizons.

27. Components of the $\alpha_{R/P}(t)$ Term: In

$$\alpha_{R/P,0} = \sqrt{\frac{MPoR}{10\ BD}} \times \frac{IMR_{R/P,SIMM}(t = 0)}{q\left(0.99/0.01\,, \Delta MTM(0, MPoR)\right)}$$

$IMR_{R/P,SIMM}(t = 0)$ are the $IMR_{R/P}$ computed at

$$t = 0$$

using SIMM; $\Delta MTM(0, MPoR)$ is the distribution of the MTM variations over the first $MPoR$; and $q(x, y)$ is a function that gives quantile x for the distribution y.

28. $t = 0$ chosen to match SIMM: The values of the normalization functions $\alpha_{R/P}(t)$ at

$$t = 0$$

are chosen in order to reconcile $IM_{R/P}(i, t)$ with the starting SIMM IMR.

29. Mean-reverting Nature of the Volatility: The functional form of $\alpha_{R/P}(t)$ at

$$t > 0$$

is dictated by what is empirically observed, as is illustrated by Anfuso, Aziz, Loukopoulos, and Giltinan (2017); accurate RFE models, in both P and Q measures, have either a volatility term structure or an underlying stochastic volatility process that accounts for the mean-reverting behavior to the normal market conditions generally observed from extremely low or high volatility.

30. Reconciliation with Static SIMM Methodology: Since the SIMM calibration is static (independence of market volatility for SIMM), the

$$t = 0$$

reconciliation factor is not independent of the market volatility, and thus not necessarily adequate for the long-term mean level.

31. Volatility Reducing Mean-reversion Speed: Hence, $\alpha_{R/P}(t)$ is an interpolant between the

$$t = 0$$

scaling driven by $\alpha_{R/P,0}$ and the long-erm scaling driven by $\alpha_{R/P,\infty}$, where the functions $\beta_{R/P}(t)$ are the mean-reverting speeds.

32. Estimating $\alpha_{R/P,\infty}$ from the Long-End: The values of $\alpha_{R/P,\infty}$ can be inferred by a historical analysis of a group of portfolios, or it can be *ad hoc* calibrated, e.g., by computing a different $\Delta MTM(0, MPoR)$ distribution in

$$\alpha_{R/P,0} = \sqrt{\frac{MPoR}{10\ BD}} \times \frac{IMR_{R/P,SIMM}(t = 0)}{q\big(0.99/0.01\,, \Delta MTM(0, MPoR)\big)}$$

using the long-end of the risk-factor implied volatility curves and solving the equivalent scaling equations for $\alpha_{R/P,\infty}$.

33. Interpreting the Haircut $h_{R/P}(t)$ Term: As will be seen below, the interpretation of $h_{R/P}(t)$ can vary depending on the intended application of the model.

34. $\underline{h_{R/P}(t)}$ for Capital/Risk Models: For capital and risk models, $h_{R/P}(t)$ are two capital and risk functions that can be used to reduce the number of back-testing exceptions (see below) and ensure that the DIM model is conservatively calibrated.
35. $\underline{h_{R/P}(t)}$ for the XVA Models: For XVA pricing, $h_{R/P}(t)$ can be fine-tuned – together with $\beta_{R/P}(t)$ - to maximize the accuracy of the forecast based on historical performance.
36. Lack of Asset Class Netting: Note that owing to the *No netting across Asset Classes* clause, the $IM_{R/P,x}(i,t)$ can be computed on a stand-alone basis for every asset class x defined by SIMM (IR/FX, equity, qualified and non-qualified credit, commodity) without any additional exposure runs. The total $IM_{R/P}(i,t)$ is then given by the sum of the $IM_{R/P,x}(i,t)$ values.
37. Historical vs. Computed IM Calibrations: A comparison between the forecasts of the DIM model defined in

$$\sigma^2(i,t) = \mathbb{E}[\Delta MTM^2(i,t) \mid MTM(i,t)] = \sum_{k=0}^{n} a_{\sigma k} MTM^k(i,t)$$

$$IM_{R/P,U}(i,t) = \Phi^{-1}\big(0.99/0.01, \mu = 0, \sigma = \sigma(i,t)\big)$$

$$IM_{R/P}(i,t) = \alpha_{R/P}(t) \times IM_{R/P,U}(i,t)$$

$$\alpha_{R/P}(t) = \left[1 - h_{R/P}(t)\right] \times \sqrt{\frac{10\ BD}{MPoR}} \times \left[\alpha_{R/P,\infty} + \left(\alpha_{R/P,0} - \alpha_{R/P,\infty}\right)e^{-\beta_{R/P}(t)t}\right]$$

$$\alpha_{R/P,0} = \sqrt{\frac{MPoR}{10\ BD}} \times \frac{IMR_{R/P,SIMM}(t=0)}{q\big(0.99/0.01\,, \Delta MTM(0, MPoR)\big)}$$

and the historical IMR realizations computed with the SIMM methodology is shown in Anfuso, Aziz, Loukopoulos, and Giltinan (2017) where alternative scaling approaches are considered.

38. Criteria Utilized in the Comparison: A comparison is performed at different forecasting horizons using 7 years of historical data, monthly sampling, and averaging among a wide representation of single-trade portfolios for the posted and the received IM cases.
39. $\mathcal{L}_1$ Error Metric Choice: For a given portfolio/horizon, the chosen error metric is given by $\mathbb{E}_{t_k}\left[\frac{|F_{R/P}(t_k+h)-G_{R/P}(t_k+h)|}{G_{R/P}(t_k+h)}\right]$ where $\mathbb{E}_{t_k}[\cdot]$ indicates an average across historical sampling dates – the definitions of $F_{R/P}$ and $G_{R/P}$ are contained below. Here and throughout this chapter, t_k is used in place of t whenever the same quantity is computed at multiple sampling dates.
40. Comparison of the Tested Universe: The tested universe is made up of 102 single-trade portfolios. The products considered, always at-the-money and of different maturities, include cross-currency swaps, IR swaps, FX options, and FX forwards – approximately 75% of the population is made up of

$$\Delta = 1$$

trades.
41. Calibrated Estimates of the Parameters: As is made evident by Anfuso, Aziz, Loukopoulos, and Giltinan (2017), the proposed term structure of $\alpha_{R/P}(t)$ improves the accuracy of the forecast by a significant amount – they also provide the actual calibration used for their analysis.
42. Conservative Calibration of the Haircut Function: Below contains further discussions on the range of values that the haircut functions $h_{R/P}(t)$ are expected to take for a conservative calibration of DIM to be used for regulatory exposure.
43. Comparison with CCP IMR: Finally, as an outlook, Anfuso, Aziz, Loukopoulos, and Giltinan (2017) show the error metrics for the case of CCP IMR where the Dim forecasts are compared against the Portfolio Approaches to Interest Rate Scenarios (Pairs: LCH.ClearNet) and historical value-at-risk (HVaR; Chicago Mercantile Exchange) realizations.
44. Prototype Replications of CCP Methodologies: The realizations are based on prototype replications of the market risk components of the CCP IM methodologies.

45. Universe Used for the CCP Tests: The forecasting capability of the model is tested separately for Pairs and HVaR IMR as well as for 22 single-trade portfolios (IRS trades of different maturities and currencies). The error at any given horizon is obtained by averaging among 22×2 cases.
46. Accuracy of the Proposed Scaling: Without fine tuning the calibration any further, the time-dependent scaling $\alpha_{R/P}(t)$ drives a major improvement in the accuracy of the forecasts with respect to the alternative approaches.

How to Backtest a DIM Model

1. Assessing Model for Different Applications: The discussion so far has focused on a DIM model for B-IMR without being too specific about how to assess the model performance for different applications, such as CVA and margin valuation adjustment (MVA) pricing, liquidity coverage ratio/net stable funding ratio (LCR/NSFR) monitoring (Basel Committee on Capital Supervision (2013)), and capital exposure.
2. Estimating the IMR Distribution Accurately: As mentioned above, depending upon which application one considers, it may or may not be important to have an accurate assessment of the distribution of *the simulated IM requirements* value (IMRD).
3. Backtesting to measure DIM Performance: This chapter introduces two distinct levels of backtesting that can measure the DIM model performance in two topical cases:
    a. DIM applications that do not depend directly on the IMRD (such as capital exposure and the CVA), and
    b. DIM applications that directly depend on the IMRD (such as MVA calculation and LCR/NSFR monitoring).

    The methodologies are presented below, with a focus on the P-measure applications.

Backtesting DIM Mapping Functions (for Capital Exposure and CVA)

1. Review of the Monte-Carlo Framework: In a Monte-Carlo simulation framework, the exposure is computed by determining the MTM values of a given portfolio on a large number of forward-looking risk-factor scenarios.
2. Adequacy of Forecasts across Scenarios: To ensure that a DIM model is sound, one should verify that the IM forecasts associated with the future simulation scenarios are adequate for a sensible variety of forecasting horizons as well as initial and terminal market conditions.
3. Setting up a Suitable Backtesting Framework: A suitable historical backtesting framework so as to statistically assess the performance of the model by comparing the DIM forecast with the realistic exact IMR, e.g., in the case of B-IMR calculated according to the SIMM methodology – for a representative sample of historical dates as well as market conditions and portfolios.
4. Generic IMR of a Portfolio: Let us first define generic IMR of a portfolio p as

$$IMR = g_{R/P}\left(t = t_{\alpha}, \Pi = \Pi\big(p(t_{\alpha})\big), \vec{M}_g = \vec{M}_g(t_{\alpha})\right)$$

   The terms are as follows.
5. Posted/Received IMR Computation Algorithm: The functions g_R and g_P represent the exact algorithm used to compute the IMR for the posted and the received IM's, respectively (e.g., such as SIMM for B-IMR, or in the case of the CCP's, IM methodologies such as Standard Portfolio Analysis of Risk (SPAN), Pairs, or HVaR).
6. Date of the IMR Valuation:

$$t = t_{\alpha}$$

   is the time at which the IMR portfolio p is determined.

7. Portfolio Trade Population at t_α: $\Pi\left(p(t_\alpha)\right)$ is the trade population of portfolio p at time t_α.
8. Market State Information at t_α: $\vec{M}_g(t_\alpha)$ is a generic state variable that characterizes all of the

$$T \leq t_\alpha$$

   market information required for the computation of the IMR.
9. DIM Forecast of the Portfolio: Similarly, the DIM forecast for the future IMR of a portfolio p can be defined as

$$DIM = f_{R/P}\left(t_0 = t_k, t = t_k + h, \vec{r}, \qquad \Pi = \Pi\left(p(t_k)\right), \vec{M}_{DIM} = \vec{M}_{DIM}(t_k)\right)$$

   The terms are as follows.
10. Posted/Received DIM Computation Algorithm: The functions f_R and f_P represent the DIM forecast for the posted and the received IM's, respectively.
11. Date of the DIM Forecast:

$$t_0 = t_k$$

   is the date time at which the DIM forecast is computed.
12. Horizon of the DIM Forecast:

$$t = t_k + h$$

   is the time for which the IMR is forecast – over a forecasting horizon

$$h = t - t_0$$

13. Predictor Set of Market Variables: $\vec{r}$ - the *predictor* – is a set of market variables whose forecasted values on a given scenario are consumed by the DIM models as input to infer the IMR.
14. $\vec{r}$ as Simulated Portfolio MTM: The exact choice of $\vec{r}$ depends on the DIM model. For the one considered previously, $\vec{r}$ is simply given by the simulated MTM of the portfolio.
15. Market State Information at t_k: $\vec{M}_{DIM}(t_k)$ is the generic state variable characterizing all the

$$T \le t_k$$

market information required for the computation of the DIM forecast.
16. Portfolio Trade Population at t_k: $\Pi(\cdot)$ is defined as before.
17. Caveats around f_R and f_P: Despite being computed using the stochastic RFE models, f_R and f_P are not probability distributions, as they do not carry any information regarding the probability weight of a given received/posted IM value. $f_{R/P}$ are instead mapping functions between the set $\vec{r}$ chosen as predictor and the forecast value for IM.
18. Confidence Level Based DIM Calibration: In terms of $g_{R/P}$ and $f_{R/P}$ one can define exception counting tests. The underlying assumption is that the DIM model is calibrated at a given confidence level (CL); therefore, it can be tested as a $VaR(CL)$ model.
19. Model Conservatism Linked to CL: This comes naturally in the context of real-world *P* applications, such as capital exposure or liquidity monitoring, where a notion of model conservatism, and hence of exception, is applicable, since the model will be conservative whenever it understates (overstates) posted (received) IM.
20. The Portfolio Backtesting Algorithm Steps: For a portfolio p, a single forecasting day t_k, and a forecasting horizon h, one can proceed as follows.
21. t_k Estimate of the Forecast Functions: The forecast functions $f_{R/P}$ computed at time t_k are $f_{R/P}\left(t_0 = t_k, t = t_k + h, \vec{r},\ \Pi = \Pi\big(p(t_k)\big), \vec{M}_{DIM} = \vec{M}_{DIM}(t_k)\right)$ Note that $f_{R/P}$ depends exclusively on the predictor $\vec{r}$ –

$$\vec{r} = MTM$$

for the case considered above.

22. Impact of the Horizon on Predictor/Portfolio: The realized value of the predictor

$$\vec{r} = \vec{R}$$

is determined. For the model considered above, $\vec{R}$ is given by the portfolio value $p(t_k + h)$ where the trade population $\Pi\big(p(t_k + h)\big)$ at $t_k + h$ differs from t_k only because of portfolio aging. Aside from aging, no other portfolio adjustments are made.

23. Forecast Received/Posted IMR Estimate: The forecast values for the received and the posted IM's are computed as

$$F_{R/P}(t_k + h) = f_{R/P}\left(t_0 = t_k, t = t_k + h, \vec{r}, \qquad \Pi = \Pi\big(p(t_k)\big), \vec{M}_{DIM} = \vec{M}_{DIM}(t_k)\right)$$

24. Forecast of the Received/Posted IM Estimate: The realized values for the received and the posted IM's are computed as

$$G_{R/P}(t_k + h) = g_{R/P}\left(t = t_k + h, \Pi = \Pi\big(p(t_k + h)\big), \vec{M}_g = \vec{M}_g(t_k + h)\right)$$

25. Exception Case: F/G Mismatch Conservatism: The forecast and the realized values are then compared. The received and the posted DIM models are considered independently, and a backtesting exception occurs whenever $F_R(F_P)$ is larger (smaller) than $G_R(G_P)$. As discussed above, this definition of exception follows from the applicability of a notion of model conservatism.

26. Detecting the Backtesting Exception History: Applying the above steps to multiple sampling points t_k one can detect back-testing exceptions for the considered history.

27. Dimensionality Reduction for the Comparison: The key step is the estimate of the posted/received IMR forecast, where the dimensionality of the forecast is reduced – from a

function to a value – making use of the realized value of the predictor, and, hence, allowing for a comparison with the realized IMR.

28. Determining the Test p-value using TVS: The determination of the test p-value requires additional knowledge of the Test Value Statistics (TVS), which can be derived numerically if the forecasting horizons are overlapping (Anfuso, Karyampas, and Nawroth (2017)).
29. Caveats behind Blind TVS Usage: In the latter situation, it can happen that a single change from one volatility regime to another may trigger multiple correlated exceptions; hence the TVS should adjust the beck-testing assessments for the presence of false positives.
30. Accuracy of the $\alpha_{R/P}(t)$ Scaling: The single trade portfolios seen earlier have been tested by Anfuso, Aziz, Loukopoulos, and Giltinan (2017) using the SIMM DIM models with the three choices of scaling discussed earlier. The results confirm the greater accuracy of the term structure scaling of $\alpha_{R/P}(t)$.
31. Accuracy in the Presence of Haircut: In fact, for the same level of the haircut function

$$h_{R/P}(t > 0) = \pm 0.25$$

positive/negative for posted/received – a much lower number of exceptions is detected.

32. Realistic Values for the Haircut: Anfuso, Aziz, Loukopoulos, and Giltinan (2017) also observe that, in this regard, for realistic diversified portfolios and calibration targets of

$$CL = 95\%$$

the functions $h_{R/P}(t)$ take values typically in the range of $10 - 40\%$.

33. Assumptions Underlying the Haircut Assumption: The range of values for $h_{R/P}(t)$ has been calibrated using

$$\beta_{R/P}(t) = 1$$

and

$$\alpha_{R/P,\infty}(t) = 1$$

Both assumptions are broadly consistent with historical data.

34. IOSCO results in Over-collateralization: Note also that the goal of the BCBS-IOSCO regulations is to ensure that the netting sets are largely over-collateralized as a consequence of:
    a. The high confidence level at which the IM is computed, and
    b. The separate requirements for IM and VM.
35. Impact of Over-collateralization: Hence, the exposure generating scenarios are tail events, and the effect on capital exposure of a conservative haircut applied to the received IM is rather limited in absolute terms.
36. Over-collateralization Impact on Exposure: This issue is demonstrated by Anfuso, Aziz, Loukopoulos, and Giltinan (2017) where the expected exposure (EE) at a given horizon t is shown as a function of $h_R(t)$ – the haircut to be applied to the received IM collateral – for different distributional assumptions on $\Delta MTM(t, t + MPoR)$.
37. Distribution Dependence on Haircut Functions: In particular, they compute the expected exposure for

$$h_R(t) = 0$$

and

$$h_R(t) = 1$$

indicating full IM collateral benefit or no benefit at all – and take the unscaled IM as the 99^{th} percentile of the corresponding distribution. For different classes of the ΔMTM distribution, the exposure reduction is practically unaffected up to haircuts of $\approx 50\%$.

1. MC Based DIM IMR Distributions: The same Monte Carlo framework can be used in combination with a DIM model to forecast the IMD at any future horizon – implicit here are the models in which the DIM is not always constant across the scenarios. The applications of the IMRD are multiple.
2. Some Applications using the IMRD: The following are two examples that apply equally to the cases of B-IMR and CCP IMR:
    a. Future IM funding costs in the P measure, i.e., the MVA
    b. Future IM funding costs in the Q measure, e.g., in relation to LCR and NSFR regulations (Basel Committee on Banking Supervisions (2013))
3. Numerically Forecasting the IMR Distributions: The focus here is on the forecasts on the P-measure – tackling the case of the Q-measure may require a suitable generalization of Jackson (2013). The main difference with the backtesting approach discussed above is that the new model forecasts are the numerical distributions of the simulated IMR values.
4. Scenario-specific IM Forecasting: These can be obtained for a given horizon by associating every simulated scenario with its corresponding IMR forecast, computed according to the given DIM model.
5. Posted/Received IMR Density CDF: Using the notation introduced previously, the numerical representations of the received/posted IMRD cumulative density functions (CDF's) of a portfolio p for a forecasting day t_k and a horizon h are given by

$$CDF_{R/P}(x, t_k, h) = \frac{\#\{v \in \mathbb{V} \mid v \leq x\}}{N_{\mathbb{V}}} \ \forall \ \vec{r}_\omega \in \Omega$$

$$\mathbb{V} = \left\{f_{R/P}\left(t_0 = t_k, t = t_k + h, \vec{r}, \qquad \Pi = \Pi\big(p(t_k)\big), \vec{M}_{DIM} = \vec{M}_{DIM}(t_k)\right)\right\}$$

6. Terms of the CDF Expression: In

$$CDF_{R/P}(x, t_k, h) = \frac{\#\{v \in \mathbb{V} \mid v \leq x\}}{N_{\mathbb{V}}}$$

$N_{\mathbb{V}}$ is the total number of scenarios. In

$$\mathbb{V} = \left\{f_{R/P}\left(t_0 = t_k, t = t_k + h, \vec{r}, \qquad \Pi = \Pi(p(t_k)), \vec{M}_{DIM} = \vec{M}_{DIM}(t_k)\right) \forall\, \vec{r}_\omega \in \Omega\right\}$$

$f_{R/P}$ are the functions computed using the DIM model, $\vec{r}_\omega$ are the scenarios for the predictor – the portfolio MTM values in the case originally discussed, and Ω is the ensemble of $\vec{r}_\omega$ spanned by the Monte Carlo simulation.

7. Suitability of IMRD for Backtesting: The IMRD in this form is directly suited for historical backtesting using the Probability Integral Transformation (PIT) framework (Diebold, Gunther, and Tay (1998)).
8. Forecasting Horizon PIT Time Series: Referring to the formalism described in one can derive the PIT time series $\tau_{R/P}$ for a portfolio p for a given forecasting horizon h and backtesting history $\mathcal{H}_{BT}$ as:

$$\tau_{R/P} = CDF\left(g_{R/P}\left(t = t_k + h, \Pi = \Pi(p(t_k + h)), \vec{M}_g = \vec{M}_g(t_k + h)\right), t_k, h\right) \forall\, t_k \in \mathcal{H}_{BT}$$

9. Samples from the Actual IMR Algorithm: In the expression for $\tau_{R/P}$ above, $g_{R/P}$ is the exact IMR algorithm for the IMR methodology that is to be forecast – defined as

$$IMR = g_{R/P}\left(t = t_\alpha, \Pi = \Pi(p(t_\alpha)), \vec{M}_g = \vec{M}_g(t_\alpha)\right)$$

and t_α are the sampling points in $\mathcal{H}_{BT}$.

10. Probability of t_k-realized IMR: Every element of the PIT time series $\tau_{R/P}$ corresponds to the probability of the realized IMR at time $t_k + h$ according to the DIM forecast built at t_k.
11. Backtesting of the Portfolio Models - Variations: As discussed extensively in Anfuso, Karyampas, and Nawroth (2017) one can backtest $\tau_{R/P}$ using uniformity tests. In particular,

analogous to what was shown in Anfuso, Karyampas, and Nawroth (2017) for portfolio backtesting in the context of capital exposure models, one can use test metrics that do not penalize conservative modeling – i.e., models overstating/understating posted/received IM. In all cases the appropriate TVS can be derived using numerical Monte Carlo simulations.

12. Factors affecting the Backtesting: In this setup the performance of a DIM is not done in isolation. The backtesting results will be mostly affected by the following.
13. Impact of $\vec{r}$ on Backtesting: As discussed earlier, $\vec{r}$ is the predictor used to associate an IMR with a given scenario/valuation time point. If $\vec{r}$ is a poor indicator for the IMR, the DIM forecast will consequently be poor.
14. Mapping of $\vec{r}$ to IMR: If the mapping model is not accurate, then the IMR associated with a given scenario will be inaccurate. For example, the models defined in

$$\sigma^2(i,t) = \mathbb{E}[\Delta MTM^2(i,t) \mid MTM(i,t)] = \sum_{k=0}^{n} a_{\sigma k} MTM^k(i,t)$$

$$IM_{R/P,U}(i,t) = \Phi^{-1}\big(0.99/0.01, \mu = 0, \sigma = \sigma(i,t)\big)$$

$$IM_{R/P}(i,t) = \alpha_{R/P}(t) \times IM_{R/P,U}(i,t)$$

$$\alpha_{R/P}(t) = \left[1 - h_{R/P}(t)\right] \times \sqrt{\frac{10\ BD}{MPoR}} \times \left[\alpha_{R/P,\infty} + \left(\alpha_{R/P,0} - \alpha_{R/P,\infty}\right) e^{-\beta_{R/P}(t)t}\right]$$

$$\alpha_{R/P,0} = \sqrt{\frac{MPoR}{10\ BD}} \times \frac{IMR_{R/P,SIMM}(t=0)}{q\big(0.99/0.01\,, \Delta MTM(0, MPoR)\big)}$$

include scaling functions to calibrate the calculated DIM to the observed

$$t = 0$$

IMR. The performance of the model is therefore dependent on the robustness of this calibration at future points in time.

15. RFE Models used for $\vec{r}$: The models ultimately determine the probability of a given IMR scenario. It may so happen that the mapping functions $f_{R/P}$ are accurate but the probabilities for the underlying scenarios for $\vec{r}$ are misstated, and, hence, cause backtesting failures.
16. Differential Impact of Backtesting Criterion: Note that
    a. The choice of $\vec{r}$, and
    b. The mapping

$$\vec{r} \rightarrow IMR$$

are also relevant to the backtesting methodology discussed earlier in this chapter. RFE models used for $\vec{r}$, however, are particular to this backtesting variance, since it concerns the probability weights of the IMRD.

Conclusion

1. Framework to Develop/Backtest DIM: This chapter has presented a complete framework to backtest and develop DIM models. The focus has been on B-IMR and SIMM, and the chapter has shown how to obtain forward-looking IM's from the simulated exposure paths using simple aggregation methods.
2. Applicability of the Proposed Model: The proposed model is suitable for both XVA pricing and capital exposure calculations; the haircut functions in

$$\alpha_{R/P}(t) = \left[1 - h_{R/P}(t)\right] \times \sqrt{\frac{10\ BD}{MPoR}} \times \left[\alpha_{R/P,\infty} + \left(\alpha_{R/P,0} - \alpha_{R/P,\infty}\right) e^{-\beta_{R/P}(t)t}\right]$$

can be used to either improve the accuracy (pricing) or to ensure the conservatism of the forecast (capital).

3. CCR Capital using DIM Models: If a financial institution were to compute CCR exposure using internal model methods (IMM), the employment of a DIM could reduce the CCR capital significantly, even after the application of a conservative haircut.
4. Over-collateralization inherent in Basel SA-CCR: This should be compared with the regulatory alternative SA-CCR, where the benefits from over-collateralization are largely curbed (Anfuso and Karyampas (2015)).
5. Backtesting Methodology to Estimate Performance: As part of the proposed framework, this chapter introduced a backtesting methodology that is able to measure model performance for different applications of DIM.
6. Agnosticity of DIM to the Underlying IMR: The DIM model and the backtesting methodology presented are agnostic to the underlying IMR algorithm, and they can be applied in other contexts such as CCP IM methodologies.

References

- Andersen, L., M. Pykhtin, and A. Sokol (2017a): Re-thinking Margin Period of Risk **eSSRN**.
- Andersen, L., M. Pykhtin, and A. Sokol (2017b): Credit Exposure in the Presence of Initial Margin **eSSRN**.
- Anfuso, C., D. Aziz, K. Loukopoulos, and P. Giltinan (2017): A Sound Modeling and Backtesting Framework for Forecasting Initial Margin **eSSRN**.
- Anfuso, C., and D. Karyampas (2015): Capital Flaws *Risk* **27 (7)** 44-47

- Anfuso, C., D. Karyampas, and A. Nawroth (2017): A Sound Basel III Compliant Framework for Backtesting Credit Exposure Models **eSSRN**.
- Basel Committee on Banking Supervision (2013): Basel III: The Liquidity Coverage Ratio and Liquidity Risk Monitoring Tools
- Basel Committee on Banking Supervision (2015): Margin Requirements for Non-centrally Cleared Derivatives
- Diebold, F. X., T. A. Gunther, and A. S. Tay (1998): Evaluating Density Forecasts with Applications to Financial Risk Management *International Economic Review* **39 (4)** 863-883
- Green, A. D., and C. Kenyon (2015): MVA by Replication and Regression **arXiV**
- International Swaps and Derivatives Association (2016): ISDA SIMM Methodology
- Jackson, L. (2013): Hedge Backtesting for Model Validation *Risk* **25 (9)** 64-67
- Longstaff, F., and E. Schwartz (2001): Valuing American Options by Simulation: A Simple Least-Squares Approach *Review of Financial Studies* **14 (1)** 113-147