A sound modelling and backtesting framework for forecasting initial margin requirements

The introduction of mandatory margining for bilateral over-the-counter transactions is significantly affecting the derivatives market, particularly in light of the additional funding costs financial institutions could face. In the following, Fabrizio Anfuso, Daniel Aziz, Klearchos Loukopoulos and Paul Giltinan propose a consistent framework, equally applicable to cleared and non-cleared portfolios, to develop and backtest forecasting models for initial margin

Since the publication of the new Basel Committee on Banking Supervision and International Organization of Securities Commissions (BCBS-Iosco) guidance on mandatory margining for non-cleared over-the-counter derivatives (Basel Committee on Banking Supervision 2015), there has been growing interest in the industry regarding the development of dynamic initial margin (DIM) models (see, for example, Andersen *et al* 2014; Green & Kenyon 2015); by 'DIM model', we are referring to any model that can be used to forecast future portfolio initial margin requirements (IMR).

The business case for such a development is at least twofold.

■ The BCBS-Iosco IMR (B-IMR) are supposed to protect against potential future exposure at a high level of confidence (99%) and will substantially affect funding costs, XVA and capital.

■ The B-IMR 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 counterparty (CCP) risk management model as well.

Therefore, for both bilateral and cleared derivatives, current and future IMR significantly affects the profitability and risk profile of a given trade.

In the present article, we consider B-IMR as a case study, and we show how to include a suitably parsimonious DIM model in the exposure calculation. We propose an end-to-end framework and define a methodology to backtest the model's performance.

This paper is organised as follows. First, the DIM model for the forecasting of future IMR is presented. We then discuss methodologies for two distinct levels of backtesting analysis. Finally, we draw conclusions.

How to construct a DIM model

A DIM model can be used for various purposes. In the computation of counterparty credit risk (CCR), capital exposure or credit valuation adjustment (CVA), the DIM model should forecast, on a path-by-path basis, the amount of posted and received IM at any revaluation point. 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.

The DIM model is *a priori* agnostic to the underlying risk factor evolution (RFE) models used to generate the exposure paths (as we will see, dependencies may arise if, for example, the DIM is computed based on the same paths that are generated for the exposure).

It is a different story if the goal is to forecast the IMR distribution (IMRD) at future horizons, either in real-world P or market-implied Q

measures. In this context, the key feature of the model is to associate the right probability weight with a given IMR scenario; hence, the forecasted IMRD also becomes a measure of the accuracy of the RFE models (which ultimately determine the likelihood of the different market scenarios). The distinction between the two cases will become clearer later on, when we discuss how to assess model performance.

In the remainder of this paper, we consider BCBS-Iosco IM as a case study. For B-IMR, the current industry proposal is the International Swaps and Derivatives Association standard initial margin model (Simm), a static aggregation methodology to compute IMR based on first-order delta-vega trade sensitivities (International Swaps and Derivatives Association 2016). The exact replication of Simm in a capital exposure or 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.

Since the exposure simulation provides portfolio mark-to-market (MtM) values on the default (time t) and closeout (time $t + \text{MPOR}$, where 'MPOR' is 'margin period of risk') grids, Andersen *et al* (2014) have proposed using this information to infer pathwise the size of any percentile of the local $\Delta\text{MtM}(t, t + \text{MPOR}, \text{path}_i)$ distribution,[1] based on a regression that uses the simulated portfolio $\text{MtM}(t)$ as the independent variable. This methodology can be further improved by adding more descriptive variables to the regression, eg, the values at the default time t of selected risk factors of the portfolio.

For our DIM model, the following features are desirable.

■ (f1) The DIM model should consume the same paths as those generated for the exposure simulation, in order to minimise the computational burden.

■ (f2) The output of the DIM model should reconcile with the known B-IMR value for $t = 0$, ie, $\text{IM}(\text{path}_i, 0) = \text{IMR}_{\text{Simm}}(0)$ for all i.

Before proceeding, we note some of the key aspects of the BCBS-Iosco margining guidelines and, consequently, of the Isda Simm model (International Swaps and Derivatives Association 2016).

■ (a1) The MPOR for the IM calculation of a daily margined counterparty is equal to 10 days. This may differ from the capital exposure calculation,

---

[1] *The* $\Delta\text{MtM}(t, t + \text{MPOR}) = \text{MtM}(t + \text{MPOR}) - \text{MtM}(t)$ *distribution is constructed assuming no cashflows take place between default and closeout. For a critical review of this assumption, see Andersen* et al *(2016).*

in which, for example, MPOR $= 20$ days if the number of trades in the portfolio exceeds 5,000.

■ (a2) The B-IMR in Basel Committee on Banking Supervision (2015) prescribe the calculation of IM by segregating trades from different asset classes. This feature is coherently reflected in the Simm model design.

■ (a3) The Simm methodology consumes trade sensitivities as its only input and has a static calibration that is not sensitive to market volatility.

These features, together with the requirements (f1) and (f2) stated previously, are addressed by our model proposal, as we will see.

For the IM calculation, the starting point is similar to that of Andersen *et al* (2014), ie, (i) we use a regression methodology based on the paths MtM(t) to compute the moments of the local ΔMtM$(t, t + \text{MPOR}, \text{path}_i)$ distribution, and (ii) we assume the ΔMtM$(t, t + \text{MPOR}, \text{path}_i)$ is a given probability distribution that can be fully characterised by its first two moments: drift and volatility. Additionally, since the drift is generally immaterial over the MPOR horizon, we do not compute it and set it to 0.

There are multiple regression schemes that can be used to determine the local volatility $\sigma(i, t)$. In the present analysis, we follow the standard American Monte Carlo literature (Longstaff & Schwartz 2001) and use a least-squares method (LSM) with a polynomial basis:

$$\sigma^2(i, t) = \langle (\Delta\text{MtM}(i, t))^2 \mid \text{MtM}(i, t) \rangle = \sum_{k=0}^{n} a_{\sigma, k} \, \text{MtM}(i, t)^k \quad (1)$$

$$\text{IM}_{\text{R/P}}^{\text{U}}(i, t) = \Phi^{-1}(0.99/0.01, \mu = 0, \sigma = \sigma(i, t)) \quad (2)$$

where R/P indicates received and posted, respectively. In our implementation, n in (1) is set equal to 2, ie, a polynomial regression of order 2 is used. We observe that LSM performs well compared with more sophisticated kernel methods (such as Nadaraya-Watson, which is used in Andersen *et al* (2014)), and it has the advantage of being parameter free and cheaper from a computational standpoint.

The unnormalised posted and received $\text{IM}_{\text{R/P}}^{\text{U}}(i, t)$ are calculated analytically in (1) and (2) by applying the inverse of the cumulative distribution function $\Phi^{-1}(x, \mu, \sigma)$ to the appropriate quantiles; $\Phi(x, \mu, \sigma)$ being the probability distribution that models the local ΔMtM$(t, t + \text{MPOR}, \text{path}_i)$. The precise choice of Φ does not play a crucial role, since the difference in quantiles among different distributional assumptions can be compensated in calibration by the scaling factors applied (see the $\alpha_{\text{R/P}}(t)$ functions in (4)). For simplicity, in the below we assume Φ is normal.

As a next step, we should account for the $t = 0$ reconciliation as well as the mismatch between the Simm and exposure model calibrations (see, respectively, items (f2), (a1) and (a3) above). These points can be tackled by scaling $\text{IM}_{\text{R/P}}^{\text{U}}(i, t)$ with suitable normalisation functions $\alpha_{\text{R/P}}(t)$:

$$\text{IM}_{\text{R/P}}(i, t) = \alpha_{\text{R/P}}(t) \times \text{IM}_{\text{R/P}}^{\text{U}}(i, t) \quad (3)$$

$$\alpha_{\text{R/P}}(t) = (1 - h_{\text{R/P}}(t)) \sqrt{\frac{10 \text{ days}}{\text{MPOR}}}$$
$$\times (\alpha_{\text{R/P}}^{\infty} + (\alpha_{\text{R/P}}^{0} - \alpha_{\text{R/P}}^{\infty}) e^{-\beta_{\text{R/P}}(t)t}) \quad (4)$$

$$\alpha_{\text{R/P}}^{0} = \sqrt{\frac{\text{MPOR}}{10 \text{ days}} \cdot \frac{\text{IMR}_{\text{R/P}}^{\text{Simm}}(t = 0)}{q(0.99/0.01, \Delta\text{MtM}(0, \text{MPOR}))}} \quad (5)$$

■ In (4), $\beta_{\text{R/P}}(t) > 0$ and $h_{\text{R/P}}(t) < 1$, with $h_{\text{R/P}}(t = 0) = 0$, are four functions to be calibrated (two for received and two for posted IMs).

As will become clearer later in this paper, the model calibration generally differs for received and posted DIM models.

■ In (4) and (5), MPOR indicates the MPOR relevant for Basel III exposure. The ratio of MPOR to 10 days accounts for item (a1), and it is taken as a square root because the underlying RFE models are typically Brownian, at least for short horizons.

■ In (5), $\text{IMR}_{\text{R/P}}^{\text{Simm}}(t = 0)$ are the $\text{IM}_{\text{R/P}}$ computed at $t = 0$ using Simm; ΔMtM$(0, \text{MPOR})$ is the distribution of mark-to-market variations over the first MPOR; and $q(x, y)$ is a function that gives the quantile x for the distribution y.

The values of the normalisation functions $\alpha_{\text{R/P}}(t)$ at $t = 0$ are chosen in order to reconcile the $\text{IM}_{\text{R/P}}(i, t)$ with the starting Simm IMR. Instead, the functional form of $\alpha_{\text{R/P}}(t)$ at $t > 0$ is dictated by what is shown in panel (a) of figure 1: accurate RFE models, in both the P and Q measures, have either a volatility term structure or an underlying stochastic volatility process that accounts for the mean-reverting behaviour to normal market conditions generally observed from extremely low or high volatility. Since the Simm calibration is static (see item (a3) above), the $t = 0$ reconciliation factor is inversely proportional to the current market volatility, and not necessarily adequate for the long-term mean level. Hence, $\alpha_{\text{R/P}}(t)$ interpolate between the $t = 0$ scaling driven by $\alpha_{\text{R/P}}^{0}$ and the long-term scaling driven by $\alpha_{\text{R/P}}^{\infty}$, where the functions $\beta_{\text{R/P}}(t)$ are the mean-reversion speeds. The value of $\alpha_{\text{R/P}}^{\infty}$ can be inferred from a historical analysis of a group of representative portfolios, or it can be *ad hoc* calibrated, eg, by computing a different ΔMtM$(0, \text{MPOR})$ distribution in (5) using the long end of the risk factor-implied volatility curves and solving the equivalent scaling equation for $\alpha_{\text{R/P}}^{\infty}$.

As we will see, the interpretation of $h_{\text{R/P}}(t)$ can vary depending on the intended application of the model.

■ For capital and risk models, $h_{\text{R/P}}(t)$ are two haircut functions that can be used to reduce the number of backtesting exceptions (see below) and ensure the DIM model is conservatively calibrated.

■ For XVA pricing, $h_{\text{R/P}}(t)$ can be fine-tuned (together with $\beta_{\text{R/P}}(t)$) in order to maximise the accuracy of the forecast based on historical performance.

Note that, regarding item (a2) above, the $\text{IM}_{\text{R/P}}^{x}(i, t)$ can be computed on a standalone basis for every asset class x defined by Simm (interest rate (IR)/foreign exchange, equity, qualified and not qualified credit, commodity) without any additional exposure runs. The total $\text{IM}_{\text{R/P}}(i, t)$ is then given by the sum of the $\text{IM}_{\text{R/P}}^{x}(i, t)$ values.

A comparison between the forecasts of the DIM model defined in (1)–(5) and the historical IMR realisations computed with the Simm methodology is shown in panel (b) of figure 1, where alternative scaling approaches are also considered. This comparison is performed at different forecasting horizons using seven years of historical data, monthly sampling and averaging among a wide representative selection of single-trade portfolios for the posted and received IM cases. For a given portfolio/horizon, the chosen error metric is given by:

$$\langle |F_{\text{R/P}}(t_k + h) - G_{\text{R/P}}(t_k + h)| \rangle_{t_k} / \langle G_{\text{R/P}}(t_k + h) \rangle_{t_k}$$

where $\langle \cdots \rangle_{t_k}$ indicates an average across historical sampling dates[2] (for $F_{\text{R/P}}$ and $G_{\text{R/P}}$, see definitions below). The tested universe is made up of

---

[2] *Here and throughout the paper, t_k is used in place of t whenever the same quantity is computed at multiple sampling dates.*

102 single-trade portfolios. The products considered, always at-the-money and of different maturities, include cross-currency swaps, IR swaps, forex options and forex forwards (approximately 75% of the population is made up of $\Delta = 1$ trades).

As is evident from figure 1, the proposed term structure of $\alpha_{R/P}(t)$ improves the accuracy of the forecasts by a significant amount. The calibration used for this analysis is provided in the caption of figure 1. Below, we will further discuss the range of values that haircut functions $h_{R/P}(t)$ are expected to take for a conservative calibration of DIM to be used for regulatory capital exposure.

Finally, as an outlook, in panel (c) of figure 1 we show the error metrics for the case of CCP IMR, where the DIM forecasts are now compared with Portfolio Approach to Interest Rate Scenarios (Pairs; LCH.Clearnet) and historical value-at-risk (HVaR; Chicago Mercantile Exchange) realisations.[3] The forecasting capability of the model is tested separately for Pairs and HVaR IMR as well as for 22 single-trade portfolios (interest rate swap (IRS) trades of different maturities and currencies). The error at any given horizon is obtained by averaging among the 22×2 cases.

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

So far, we have discussed a DIM model for B-IMR without being too specific about how to assess 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 Banking Supervision 2013) and capital exposure. As mentioned above, depending on which application one considers, it may or may not be important to have an accurate assessment of the distribution of the simulated IM requirements' values (the IMRD).

We introduce two distinct levels of backtesting that can measure the DIM model performance in two topical cases: (i) DIM applications that do not directly depend on the IMRD (such as capital exposure and CVA), and (ii) DIM applications that directly depend on the IMRD (such as MVA calculation and LCR/NSFR monitoring). The two corresponding methodologies are presented below, with a focus on P measure applications.

■ **Backtesting the DIM mapping functions (for capital exposure and CVA).** In a Monte Carlo simulation framework, the exposure is computed by determining the future mark-to-market values of a given portfolio on a large number of forward-looking risk factor scenarios. To ensure a DIM model is sound, one should verify that the IM forecasts associated with future simulated scenarios are adequate for a sensible variety of forecasting horizons as well as initial and terminal market conditions. We should introduce a suitable historical backtesting framework so as to statistically assess the performance of the model by comparing the DIM forecasts with the realised exact IMR (eg, 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.

Let us define generic IMR for a portfolio p as:

$$\text{IMR} = g_{R/P}(t = t_\alpha, \Pi = \Pi(p(t_\alpha)), \vec{M}_g = \vec{M}_g(t_\alpha)) \quad (6)$$

---

[3] *The realisations are based on prototype replications of the market risk components of the CCP IM methodologies.*

In (6), the following hold.

■ The functions g_R and g_P represent the exact algorithm used to compute the IMR for received and posted IMs, respectively (eg, Simm for B-IMR or, in the case of CCPs, IM methodologies, such as Standard Portfolio Analysis of Risk (Span), Pairs or HVaR).

■ $t = t_\alpha$ is the time at which the IMR for the portfolio p are determined.

■ $\Pi(p(t_\alpha))$ is the trade population of the portfolio p at time t_α.

■ $\vec{M}_g(t_\alpha)$ is a generic state variable that characterises all of the $T \leq t_\alpha$ market information required for the computation of the IMR.

Similarly, we define the DIM forecast for the future IMR of a portfolio p as:

$$\text{DIM} = f_{R/P}(t_0 = t_k, t = t_k + h, \vec{r}, \Pi = \Pi(p(t_k)),$$
$$\vec{M}_{\text{DIM}} = \vec{M}_{\text{DIM}}(t_k)) \quad (7)$$

In (7), the following hold.

■ The functions f_R and f_P represent the DIM forecast for received and posted IMs, respectively.

■ $t_0 = t_k$ is the time at which the DIM forecast is computed.

■ $t = t_k + h$ is the time for which the IMR are forecasted (over a forecasting horizon $h = t - t_0$).

■ $\vec{r}$ (the 'predictor') is a set of market variables whose forecasted values on a given scenario are consumed by the DIM model as input to infer the IMR. The exact choice of $\vec{r}$ depends on the DIM model. For the one considered previously, $\vec{r}$ is simply given by the simulated mark-to-market of the portfolio.

■ $\vec{M}_{\text{DIM}}(t_k)$ is a generic state variable characterising all the $T \leq t_k$ market information required for the computation of the DIM forecast.

■ $\Pi(\cdot)$ is defined as for (6).

Despite being computed using 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 a predictor and the forecasted value for the IM.

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. 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 DIM model will be conservative whenever it understates (overstates) received (posted) IM.

For a portfolio p, a single forecasting day t_k and a forecasting horizon h, one can proceed as follows.

■ (1) The forecast functions $f_{R/P}$ are computed at time t_k as $f_{R/P}(t_0 = t_k, t = t_k + h, \vec{r}, \Pi = \Pi(p(t_k)), \vec{M}_{\text{DIM}} = \vec{M}_{\text{DIM}}(t_k))$. Note that $f_{R/P}$ depends explicitly on the predictor $\vec{r}$ ($\vec{r} = \text{MtM}$ for the model considered above).

■ (2) The realised 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 $\Pi(p(t_k + h))$ at $t_k + h$ differs from t_k only because of portfolio ageing. Aside from ageing, no other portfolio adjustments are made.

■ (3) The forecasted values for the received and posted IMs are computed as $F_{R/P}(t_k + h) = f_{R/P}(t_0 = t_k, t = t_k + h, \vec{r} = \vec{R}, \Pi = \Pi(p(t_k)), \vec{M}_{\text{DIM}} = \vec{M}_{\text{DIM}}(t_k))$.

1 Model forecasting performance versus historical realisations
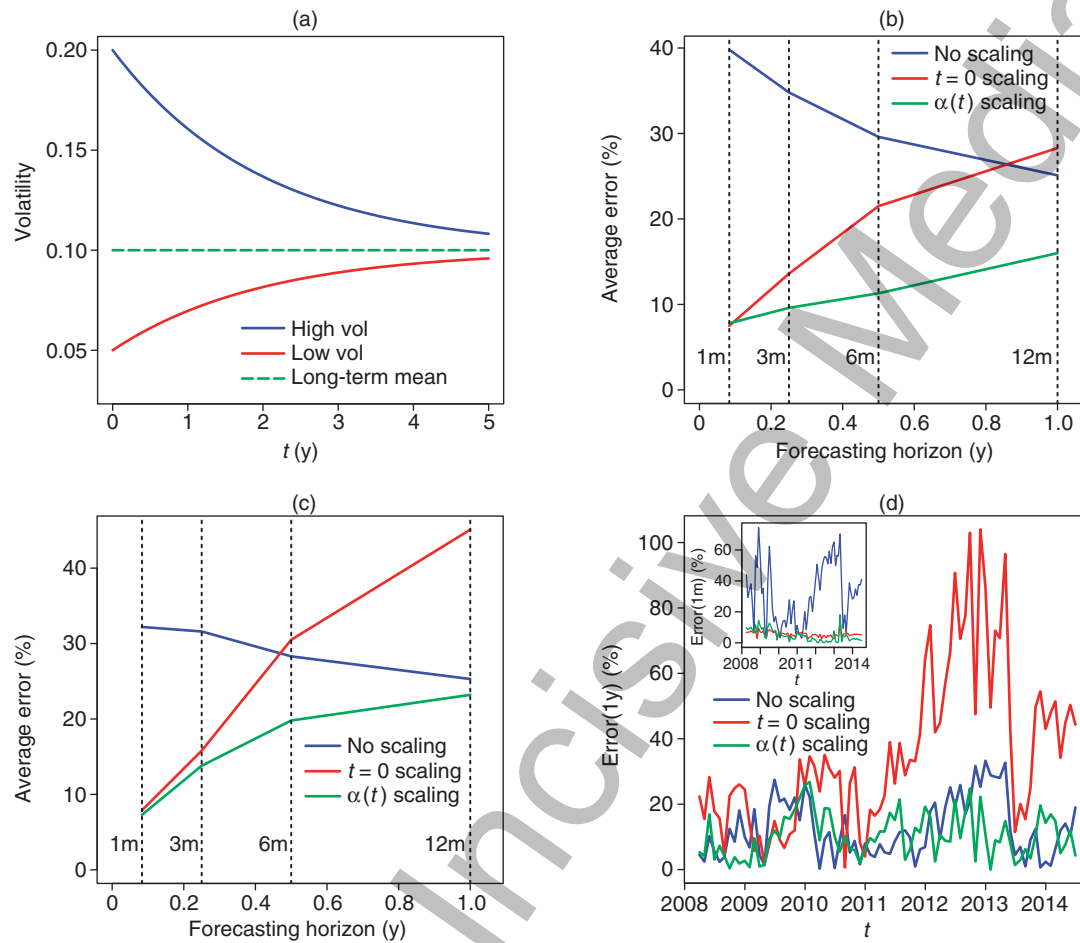
(a) Market-implied volatility term structure. Sketch of the term structure of the RF volatility in the case of low (red line) and high (blue line) volatility markets. The dashed green line indicates long-term asymptotic behaviour. (b) Historical comparison of DIM forecasts versus Simm realisations. The accuracy of the DIM forecasts is measured versus historical Simm realisations for three choices of scaling (no scaling: $\alpha(t) = 1$; $t = 0$ scaling: $\alpha(t) = \alpha_{R/P}^0$, as for (4), with $\beta_{R/P}(t) = 0$ and $h_{R/P}(t) = 0$; $\alpha(t)$ scaling: as for (4), with $\beta_{R/P}(t) = 1$, $\alpha_{R/P}^\infty = 1$ and $h_{R/P}(t) = 0$). (c) Historical comparison of DIM forecasts versus CCP IM realisations. The accuracy of the DIM forecasts is measured for CCP IMR with an equivalent error metric. (d) DIM forecasts versus Simm realisations: 20-year USD IRS. The realised error of the Simm DIM for received IM is shown versus time for a 20-year USD IRS payer in the case of one-month (inset graph) and one-year (main graph) forecasting horizons

■ (4) The realised values for the received and posted IMs are computed as $G_{R/P}(t_k + h) = g_{R/P}(t_0 = t_k + h, \Pi = \Pi(p(t_k + h)), \vec{M} = \vec{M}_g(t_k + h))$.

■ (5) The forecasted and realised values are compared. The received and 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.

Applying the 1–5 programme to multiple sampling points t_k, one can detect backtesting exceptions for the considered history. The key step is 3, where the dimensionality of the forecast is reduced (from a function to a value), making use of the realised value of the predictor and, hence, allowing for a comparison with the realised IMR.

The determination of the test p-value requires the additional knowledge of the test value statistics (TVS), which can be derived numerically if the forecasting horizons are overlapping (see Anfuso *et al* 2014). 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 backtesting assessment for the presence of false positives.

The single-trade portfolios of figure 1 have been backtested with the above-described methodology, using Simm DIM models with the three choices of scaling discussed in the same figure. The results shown in table A confirm the greater accuracy of the term-structure scaling $\alpha_{R/P}(t)$. In fact, for the same level of haircut function $h_{R/P}(t > 0) = \pm 0.25$ (positive/negative for received/posted), a much lower number of exceptions is detected. We observe in this regard that, for realistic diversified portfolios and for a calibration target of CL $= 95\%$, the functions $h_{R/P}(t)$ take values typically in the range 10–40%.[4]

---

[4] *This range of values for $h_{R/P}(t)$ has been calibrated using $\beta_{R/P}(t) = 1$ and $\alpha_{R/P}^\infty = 1$. Both assumptions are broadly consistent with the historical data.*

A. Simm backtesting: R/P DIM VaR(95%) exceptions/tests					
Scaling type	1m	3m	6m	12m	Total
$\alpha(t)$ scaling	0/200	2/184	14/152	35/120	51/656
$t = 0$ scaling	0/200	4/184	32/152	42/120	78/656
No scaling	106/200	97/184	79/152	47/120	329/656

Historical backtesting for the Simm DIM model defined in equations (1)–(5), where $h_{R/P}(t > 0) = \pm 0.25$. VaR(95%) exception counting tests are performed for the same single-trade portfolios of figure 1, and in accordance with the first backtesting methodology described. The results are shown in the format (x/y), where y is the number of portfolios tested at a given horizon and x is the observed number of backtesting failures for a p-value acceptance threshold of 5% (received and posted IM cases are considered in aggregate). The backtesting analysis is performed using seven years of historical data and a monthly sampling frequency.

Note also that the goal of the BCBS-Iosco regulations is to ensure netting sets are largely overcollateralised (as a consequence of (i) the high confidence level at which the IM is computed and (ii) the separate requirements for daily VM and IM). 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. See, in this regard, panel (b) of figure 2, where the expected exposure (EE) at a given horizon t is shown as a function of $h_R(t)$ (haircut to be applied to the received IM collateral) for different distributional assumptions on $\Delta\text{MtM}(t, t + \text{MPOR})$.

In panel (b) of figure 2, $h_R(t) = 0$ and $h_R(t) = 1$ indicate full IM collateral benefit or no benefit, respectively, and the unscaled IM is taken as the ninety-ninth percentile of the corresponding distribution. For different choices of the ΔMtM distribution, the exposure reduction is practically unaffected up to haircuts of order $\approx 50\%$.

■ **Backtesting the IMRD (for MVA and LCR/NSFR).** The same Monte Carlo framework can be used in combination with a DIM model to forecast the IMRD at any future horizon (here, we implicitly refer to models in which the DIM is not always constant across scenarios). The applications of the IMRD are multiple. The following are two examples that apply equally to the cases of B-IMR and CCP IMR: (i) future IM funding costs in the Q measure, ie, MVA, and (ii) future IM funding costs in the P measure, eg, in relation to LCR or NSFR regulations (Basel Committee on Banking Supervision 2013).

Our focus is on forecasts in the P measure (tackling the case of the Q measure may require a suitable generalisation of Jackson (2013)). The main difference with the backtesting approach discussed above is that now the model forecasts are the numerical distributions of simulated IMR values. These can be obtained for a given horizon by associating every simulated scenario with its correspondent IMR forecast, computed according to the given DIM model. Using the notation introduced previously, the numerical representations of the received/posted IMRD cumulative density functions (CDFs) of a portfolio p for a given forecasting day t_k and horizon h are given by:

$$\text{CDF}_{R/P}(x, t_k, h) = \#\{v \in V \mid v \leqslant x\}/N_V \tag{8}$$

$$V = \{f_{R/C}(t_0 = t_k, t = t_k + h, \vec{r}_\omega, \Pi = \Pi(p(t_k)),$$
$$\vec{M}_{\text{DIM}} = \vec{M}_{\text{DIM}}(t_k)), \ \forall \vec{r}_\omega \in \Omega\} \tag{9}$$

In (8), N_V is the total number of scenarios. In (9), $f_{R/P}$ are the functions computed using the DIM model; $\vec{r}_\omega$ are the scenarios for the predictor (the

2 IMM DIM versus SA-CCR IM and exposure versus DIM haircut
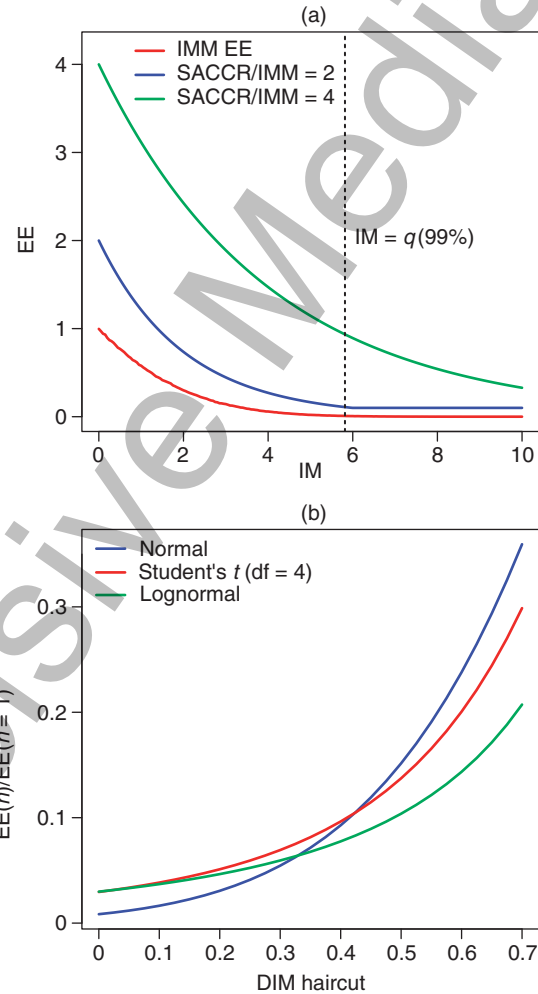
(a) Back-of-the-envelope model: IMM versus SA-CCR. The EE(t) of a daily VM margined counterparty versus the received IM for a simplified IMM model (continuous red line) and SA-CCR (continuous blue and green lines) is shown. The IMM EE(t) is calculated assuming $\Delta\text{MtM}(t, t + \text{MPOR})$ is normal. Since the SA-CCR EE(t) depends on the size of the trade's add-ons, the SA-CCR exposure is computed for the stylised cases of SA-CCR add-ons being two (blue line) or four (green line) times larger than the correspondent IMM level. (b) EE(t) versus $h_R(t)$ for different distributional assumptions on $\Delta\text{MtM}(t, t + \text{MPOR})$: normal (blue line), Student's t (red line) and lognormal (green line)

portfolio mark-to-market values in the case originally discussed); and Ω is the ensemble of the $\vec{r}_\omega$, spanned by the Monte Carlo simulation.

The IMRD in this form is directly suited for historical backtesting using the probability integral transformation (PIT) framework (Diebold *et al* 1998). Referring to the formalism described in Anfuso *et al* (2014), one can derive the PIT time series $\tau_{R/P}$ of a portfolio p for a given forecasting horizon h and backtesting history $\mathcal{H}_{\text{BT}}$ as follows:

$$\tau_{R/P} = \{\text{CDF}(g_{R/P}(t_k + h, \Pi(p(t_k + h)), \vec{M}_g(t_k + h)),$$
$$t_k, h), \ \forall t_k \in \mathcal{H}_{\text{BT}}\}. \tag{10}$$

In (10), $g_{R/P}$ is the exact IMR algorithm for the IMR methodology we intend to forecast (defined as for (6)), and t_k are the sampling points in $\mathcal{H}_{BT}$. Every element in the PIT time series $\tau_{R/P}$ corresponds to the probability of the realised IMR at time $t_k + h$ according to the DIM forecast built at t_k.

As discussed extensively in Anfuso *et al* (2014), one can backtest the $\tau_{R/P}$ using uniformity tests. In particular, analogously to what is shown in Anfuso *et al* (2014) for portfolio backtesting in the context of capital exposure models, one can also use test metrics that do not penalise conservative modelling (ie, models overstating/understating posted/received IM). In all cases, the appropriate TVS can be derived using numerical Monte Carlo simulations.

In this setup, the performance of a DIM model is not tested in isolation. The backtesting results will be mostly affected by the following.

■ (1) The choice of $\vec{r}$. As discussed earlier, $\vec{r}$ is the predictor used to associate an IMR to a given scenario/valuation time point. If $\vec{r}$ is a poor indicator for the IMR, the DIM forecast will consequently be poor.

■ (2) The mapping $\vec{r} \rightarrow$ IMR. If the mapping model is not accurate, the IMR associated with a given scenario will be inaccurate. For example, the model defined in (1)–(5) includes 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.

■ (3) The RFE models used for $\vec{r}$. These models ultimately determine the probability of a given IMR scenario. It may so happen that the mapping functions $f_{R/C}$ are accurate but the probabilities of the underlying scenarios for $\vec{r}$ are misstated and, hence, cause backtesting failures.

Note that items (1) and (2) are also relevant for the backtesting methodology discussed earlier in this paper. Item (3), however, is particular to this backtesting variance, since it concerns the probability weights of the IMRD.

Conclusion

We have presented a complete framework to develop and backtest DIM models. Our focus has been on B-IMR and Simm, and we have shown how to obtain forward-looking IMs from simulated exposure paths using simple aggregation methods.

The proposed DIM model is suitable both for XVA pricing and capital exposure calculation: the haircut functions $h_{R/P}(t)$ in (4) can be used either to improve the accuracy (pricing) or to ensure the conservatism of the forecast (capital).

If a financial institution were to compute CCR exposure using internal model methods (IMM), the employment of a DIM model could reduce CCR capital significantly, even after the application of a conservative haircut. This should be compared with the regulatory alternative SA-CCR, where the benefit from overcollateralisation is largely curbed (see panel (a) of figure 2 and Anfuso & Karyampas (2015)).

As part of the proposed framework, we have introduced a backtesting methodology that is able to measure model performance for different applications of DIM. The DIM model and the backtesting methodology presented above are agnostic to the underlying IMR algorithm, and they can be directly applied in other contexts, such as CCP IM methodologies, as shown in panel (c) of figure 1. ■

REFERENCES

Andersen L, M Pykhtin and A Sokol, 2014
Modeling credit exposure for margined counterparties
Risk Minds

Andersen L, M Pykhtin and A Sokol, 2016
Rethinking margin period of risk
SSRN Preprint

Anfuso F and D Karyampas, 2015
Capital flaws
Risk July, pages 44–47

Anfuso F, D Karyampas and A Nawroth, 2014
Credit exposure models backtesting for Basel III
Risk September, pages 82–87

Basel Committee on Banking Supervision, 2013
Basel III: the liquidity coverage ratio and liquidity risk monitoring tools
Report, January, Bank for International Settlements

Basel Committee on Banking Supervision, 2015
Margin requirements for non-centrally cleared derivatives
Report, March, Bank for International Settlements

Diebold FX, TA Gunther and AS Tay, 1998
Evaluating density forecasts with applications to financial risk management
International Economic Review 39(4), pages 863–883

Green AD and C Kenyon, 2015
MVA by replication and regression
Risk May, pages 64–67

International Swaps and Derivatives Association, 2016
ISDA SIMM methodology
Draft Document, April, Isda

Jackson L, 2013
Hedge backtesting for model validation
Risk September, pages 67–71

Longstaff F and E Schwartz, 2001
Valuing American options by simulation: a simple least-squares approach
Review of Financial Studies 14, pages 113–147