A Sound Basel III compliant framework for backtesting Credit Exposure Models *

FABRIZIO ANFUSO†, DIMITRIOS KARYAMPAS‡ AND ANDREAS NAWROTH§

May 14, 2013

Abstract

A central component of the Basel III (B3) document is the *Sound practices for backtesting*[1], i.e. a summary of strict regulatory guidances on how to validate and backtest Internal Method Models (IMM) for credit exposure. In the present work, we define a complete statistical framework to backtest credit exposure models, highlighting the features of our proposal vs. the new regulatory requirements. The framework contains four main pillars:

1. *The risk factor backtesting*, i.e. the assessment of the forecasting ability of the Stochastic Differential Equations (SDE) used to describe the dynamics of the single risk factors.
2. *The correlations backtesting*, i.e. the assessment of the statistical estimators used to describe the cross-asset evolution.
3. *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.
4. *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 (see outcome of the three pillars above).

We show with concrete examples in the cases of collateralized and uncollateralized models how to perform distributional tests w.r.t. different risk metrics. We produce discriminatory power analysis for all the tests introduced, providing exact methods to aggregate backtesting results across forecasting horizons. Most importantly, the third and the fourth pillars define a solid quantitative approach to compute capital remedies for potential model deficiencies.

KEY WORDS: Backtesting, Capital Requirements, Basel 3, Exposure Models

---

*The views expressed are those of the authors only, no other representation should be attributed

†Credit Suisse AG, Backtesting Methodology; fabrizio.anfuso@credit-suisse.com.

‡UBS AG, IB Exposure Measurement; dimitrios.karyampas@ubs.com.

§Credit Suisse AG, Backtesting Methodology; andreas.nawroth@credit-suisse.com.

1 Introduction

A central pillar of the Basel III (B3) document is the *Sound practices for backtesting*[1], i.e. a summary of strict regulatory guidances on how to validate and backtest Internal Method Models (IMM) for credit exposure. Similarly a series of requirements of CRD4, the European equivalent of B3, indicate and define backtesting and validation as core components for the good governance of IMM firms.

From a bank perspective, the new regulatory charges introduced with B3 (e.g. CVA capital charge) have stressed even more the importance of IMM in making the capital costs of the business sustainable. At the same time, the increasing complexity of the capital framework requires also a thorough approach to the validation and the monitoring of the models performance. A solid backtesting methodology is therefore the key tool both to prove regulators the soundness of the models and to ensure stakeholders that the capital position of the firm is on solid modeling grounds.

The assessment of a model is an holistic process that has both qualitative and quantitative aspects. While the former may have a decisive weight for the choice of a given model among the many possible, the latter are the ones to be considered for the backtesting. In particular, the performance of a model should be measured in terms of its forecasting capability. In statistical jargon, backtesting should address the question: "Can we reject the null hypothesis (i.e. the model) based on the available historical data?" Observe that when stochasticity is present (as for all the cases we are considering here), this question can be answered only in probabilistic terms, due to the finite amount of historical data available. In our definition, backtesting is therefore a set of statistical tests that measures the forecasting capability of a model using the available data history as comparison. The final assessment is based on a given aggregation of the p-values of the single tests, rejecting the model if an *a priori* determined threshold is breached.

In the present work, we define a complete framework to backtest credit exposure models. In section 2, we give a brief overview of the relevant metrics for counterparty credit risk and we summarize the features of our framework vs. the new regulatory guidances. In the subsequent sections 3, 4, 5, the methodology is presented in detail for the full backtesting cycle, i.e. risk factor evolution models, correlation models and portfolio exposure metrics. In section 6, we show how to compute capital buffers based on the backtesting results. Finally, we draw conclusions (section 7).

2 Basic Concepts and the need for Backtesting

Credit Counterparty Exposure is defined as the amount a company A could potentially lose in the case that its counterparty B defaults. The exposure (from A perspective) is computed based on forecasted distributions of prices for the financial contracts that constitute the portfolio of the counterparty B at any future date. The main building blocks required for this computation are the following: first, the scenario simulations for the underlying risk factors (generated with what we refer here as Risk Factor Evolution [RFE] models) and second, the pricing of each scenario to generate the $MtM(t)$ distribution at any future date t.

The relevant exposure metrics from regulatory perspective is the Expected (positive) Exposure at time t, which is defined as:

$$EE(t) = E[MtM(t)^+], \tag{1}$$

where $MtM(t)^+ = \max(MtM(t), 0)$. 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 the default capital charges. To compute the $MtM(t)$ distribution (and most importantly the $EE(t)$ profile) at any future time t we need models to forecast the evolution of the risk factors values. Those risk factors can also be dependent on each other (*correlation* assumption between risk factors). The more accurately the RFE model is specified, the more realistic is the exposure calculation. If the RFE is misspecified, the exposure figure can be wrongly stated and losses might occur with higher probability than expected in case the counterparty defaults. As observed in [2], there is a potential divergence of purposes in the choice of the calibration for the RFE models. On the one hand, the $EE(t)$ profile is used for the CVA computation (i.e. a "price") and it should be based on a market implied calibration. On the other hand, the default charge requires a forecast in the real world measure and therefore an historical calibration would be the most suitable. The backtesting methodology that we are going to describe in the following is agnostic w.r.t. the choice of the model calibration. Nevertheless, by construction, the regulatory requirements on backtesting are generally better addressed by historical models. In the view of the authors, this apparent dichotomy is one of the key quantitative challenges for the industry after the introduction of Basel 3.

Following the guidances from regulators, we define a framework that has four main pillars:

1. The RF backtesting, i.e. the assessment of the forecasting ability of the Stochastic Differential Equations (SDE) used to describe the dynamics of the single risk factors. Observe that the calibration of the SDE (market implied or historical) has a crucial influence on the assessment.

2. The correlations backtesting, i.e. the assessment of the estimators used to model the cross-asset evolution.

3. 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.

4. 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 (see outcome of the three pillars above).

Observe that the first three pillars are diagnostic while the fourth comes as a remedy for potential deficiencies in the exposure models. In Table 1 in the appendix, we summarize the relevant regulatory guidances for backtesting and how our framework addresses them. The specifics of the proposed solutions are described in more details in the sections below.

3 RF Backtesting

3.1 The backtesting construction for collateralized and uncollateralized RF models

The RFE models are the most atomic components of the exposure framework. As for regulatory items 1, 8 and 9 in Table 1, 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. The statistical tool that is the basis for the backtesting analysis is the Probability Integral Transform (PIT) (see [3] and [2]), defined as

$$F(r_n) = \int_{-\infty}^{r_n} \phi(x)dx, \tag{2}$$

where r_n is the realization of a given random variable and $\phi(.)$ is its predicted distribution. Observe that when one applies the PIT to a set of *iid* variables r_n using the correct distribution of the r_n, the transformed set $y_n = F(r_n)$ is uniformly distributed. The distance between the distribution of the transformed set y_n and the $U[0,1]$ distribution (in a statistical sense) can be therefore used as a measure of the goodness of the model $\phi(.)$ to describe the random variables r_n. In practice, the input for the RFE backtesting analysis is the time series of a given risk factor and the model (SDE + calibration) used to describe its evolution. The PIT for the predicted distribution can be used to map the realized values of the risk factor (or of their variations, see below) to a set of values in the interval $[0,1]$. The transformed set is then used to assess the model performance by applying standard distributional tests.

For uncollateralized and collateralized models, the construction differs because of the presence of single or multiple time-scales. In both cases, a grid of sampling points t_k is used. 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 to ensure the followings: i) an acceptable discriminatory power for test (in relation with the size of the data history) and ii) the absence of a significant statistical bias caused by the choice of the sampling sequence (if the sampling points are too distant, the backtesting results from an equivalent sequence with different starting point may differ).

The backtesting is carried out for an arbitrary set of horizons $\{h_1, ..., h_n\}$ and for every h_i a single result is produced. The choice of the set $\{h_1, ..., h_n\}$ should be such to reflect the portfolio structure of the firm. In the case of uncollateralized RF models, the horizon h_i is the only relevant time-scale. At $t = t_k$, the forecasted distribution for $t = t_k + h_i$ is constructed based on the given RF model and on the filtration $\mathcal{F}(t_k)$ at t_k. Via the PIT transformation (where $\phi(.)$ is given by the forecasted conditional distribution), the realized value of the RF at $t = t_k + h_i$ is mapped to a value in the $[0, 1]$ interval. For collateralized models, the presence of the margin period of risk (MPR) should be additionally accounted. The MPR is the time required for the firm to liquidate the collateral that a given counterparty posted to finance its exposure. Therefore, the primary focus of a collateralized model is to describe the variation of the RF over the MPR at any future horizon. In the backtesting exercise, $\phi(.)$ is the forecasted RF variation distribution in the interval $[t = t_k + h_i,\, t = t_k + h_i + MPR]$ conditional on $\mathcal{F}(t_k)$ and the realized value is the historical variation of the RF in the same interval. The construction described above is also shown in Fig. 1.

The result of the PIT transformation on the sampling sequence is a set of values $F(r_{t_k})$ in the interval $[0, 1]$. At this stage, standard statistical tests are applied to check for different properties of the RF distribution. In particular, we can introduce a generalized distance measure as

$$d_\omega = \int_\Gamma (F_n(x) - F(x))^2 \omega(x) dF(x), \tag{3}$$

where $\Gamma = [0, 1]$ is the domain of $F(x)$, $F_n(x)$ is the empirical cumulative distribution function (CDF) of the $F(r_{t_k})$ values, $F(x) = x$ is the CDF of a $U[0, 1]$ distribution and $\omega(x)$ is a weight function that can be chosen so to emphasize a certain quantile domain. In [2], it is suggested to link $\omega(x)$ to the portfolio structure. Conversely here the risk factors are backtested independently and the choice of $\omega(x)$ is unrelated to the portfolio composition. The portfolio backtesting is a separate pillar, as described above. The statistical literature provide many different tests to check the agreement between two different distributions. Here we consider the well-known Cramers-von-Mises (CVM, $\omega(x) = 1$) and Anderson-Darling (AD, $\omega(x) = 1/(F(x)(1 - F(x)))$) tests, where the first focus more on the center of the distribution while the later focus more on the tails. Observe that d_ω is a single *distance* value obtained from the realized set $F(r_{t_k})$. The final outcome of the backtesting should be a p-value. Therefore, we assign a p-value to the single realization d_ω by constructing the correspondent test statistic (distribution of the outcomes of d_ω) using Monte Carlo.

The test statistic distribution for a given history, sampling frequency and horizon is produced by repeating on a large number of simulated paths the backtesting calculation/mapping described above. The paths are generated with the same model/calibration which is going to be used for calculating their correspondent d_ωs. Therefore, the test statistic is the distribution of the expected distances under the correct model. Computing the quantile of the realized d_ω, we can derive the p-value for the backtesting experiment. Notice that this derivation of the statistic allows for the use of overlapping forecasting horizons because the autocorrelation among the $F(r_{t_k})$ is correctly reflected in the construction. This feature is particularly important to backtest longer horizons (see regulatory item 9 in Table 1, e.g. long-dated IR / inflation trades) for which the data history is comparably short in the most of the cases.

In Fig. 2, we show uncollateralized (upper row) and collateralized (lower row) RF backtesting results for the CHFUSD exchange rate, where we have applied the CVM test to the last 15 years of history with bi-weekly sampling frequency and for three different forecasting horizons $h_i = \{1m, 3m, 1y\}$. The RFE model is a Geometric Brownian Motion (GBM) with drift $\mu = 0$ and volatility σ calibrated historically using 1y rolling window. The dotted green line in the plots indicates the realized distance of the historical path while the continuous red line is the quantile of the test distribution corresponding to the defined pass/fail threshold we have chosen (CL=99%). For the considered case, both the uncollateralized and collateralized (MPR=2 weeks) backtesting give acceptable results at all horizons.

The methodology we have discussed in this section is equally applicable to all sorts of RFs. Nevertheless, it is meaningful to define a representative set using a given metric. One simple suggested approach is to group RFs by asset classes and select the most relevant based on trade notional or on delta-weighted trade notional. For e.g. equities, the notionals of all the trades having a certain name as underlying are aggregated and the names can be then ranked based on this figure. The top ones are identified as the most representative RFs for the given asset class (observe that this metrics cannot be directly applied for an overall ranking across asset classes, given the different typical notionals of e.g. IR and FX trades).

In the following subsections, we analyze further two aspects that were mentioned above, i.e. the discriminatory power of the statistical tests and the aggregation of backtesting results across RFs and horizons.

3.2 The discriminatory power of the RF backtesting

The assessment of the RF models depends crucially on the amount of data history. In the case of large data sets available, the backtesting can resolve

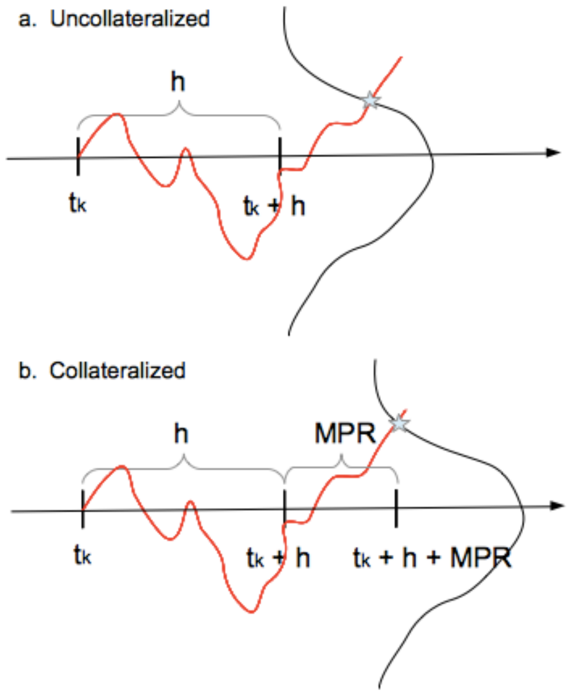

Figure 1: The construction for a given sampling date t_k for (a) uncollateralized and (b) collateralized backtesting. h is the backtesting horizon and MPR stands for margin period of risk.

very tiny model mis-specifications. Conversely, if the data are too few, the model uncertainty will be larger (especially for long horizons) and it will be comparatively easy to pass backtesting. To have a quantitative understanding of the above described effect, we run the backtesting analysis for CVM and AD tests on a synthetic data set made of 1000 paths of 15 years generated by the same stochastic model (Geometric Brownian Motion with annualized drift and volatility $\mu = 0$ and $\sigma = 10\%$). For every of these paths, we determine the p-value for $h_i = \{1m, 3m, 1y\}$ and for different mis-specifications of μ and σ. For a given h_i, the average of the p-values across the different 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 $\langle p \rangle = 0.5$).

In Figs. 3 and 4, the above analysis is shown and the correctly specified model is highlighted. As evident from the tables, the backtesting results are more sensitive at shorter horizons because of the larger number of independent observations taken into account ($h_i = 1m \rightarrow 180$ vs. $h_i = 1y \rightarrow 15$ independent observations). Notice also that AD is slightly out-performing CVM in detecting model mis-specifications for the constructed example.

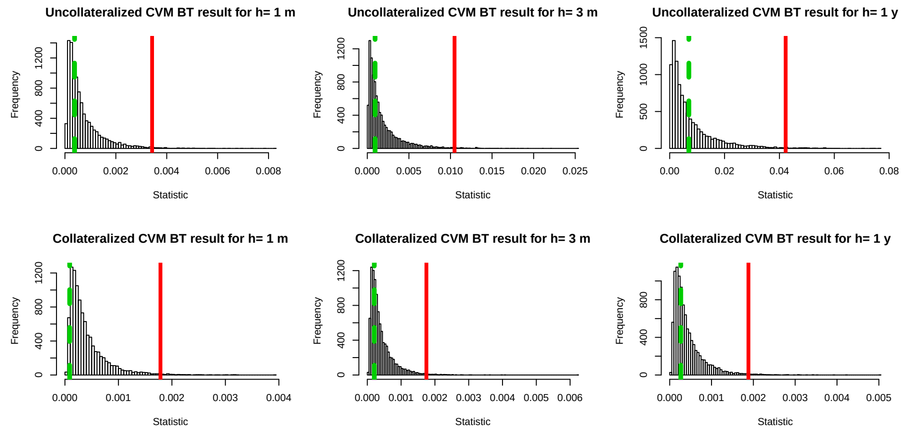

Figure 2: Upper row: the uncollateralized RF backtesting for the CHFUSD time-series in the case of: CVM test, 15 years of data history, bi-weekly sampling frequency and $h_i = \{1m, 3m, 1y\}$. Lower row: the collateralized RF backtesting for the CHFUSD time-series in the case of: CVM test, 15 years of data history, bi-weekly sampling frequency, MPR=2 weeks and $h_i = \{1m, 3m, 1y\}$. The RFE and its calibration are described in the main text. In all plots, the dotted green line indicates the realized distance of the historical path while the continuous red line is the quantile of the test distribution corresponding to the defined pass/fail threshold we have chosen (CL=99%).

In the documentation that a firm should provide to regulators for backtesting of IMM, the discriminatory power analysis is useful complementary information to assess the tolerance of the backtesting methodology.

3.3 The aggregation of backtesting results

The general indication from regulators is to show RF backtesting at the most granular level, i.e. single RFs and horizons. Nevertheless, in the view of the authors, this 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). The RF backtesting presented above allows for further aggregations within the same probabilistic framework. The scheme we consider here is for a single RF over multiple horizons, i.e. we aim to produce a single aggregated result, where the "importance" of the different horizons is given by an arbitrary weighting function $\theta(i)$ (with $\sum_i \theta(i) = 1$ and $\theta(i) > 0$, $\forall i$). 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 on the model performance should encompass many different time-scales.

Results CVM 15y 1000 paths

h=1m

Vola / Drift	-5.0%	-2.5%	0.0%	2.5%	5.0%
5.0%	100.00%	100.00%	100.00%	100.00%	100.00%
7.5%	98.68%	97.19%	96.35%	97.24%	98.78%
10.0%	83.05%	62.82%	50.73%	61.96%	82.99%
12.5%	96.44%	93.18%	91.02%	92.29%	95.78%
15.0%	99.81%	99.66%	99.56%	99.59%	99.74%

h=3m

Vola / Drift	-5.0%	-2.5%	0.0%	2.5%	5.0%
5.0%	99.83%	99.61%	99.51%	99.67%	99.87%
7.5%	93.97%	86.60%	82.68%	87.45%	94.84%
10.0%	83.28%	63.51%	50.09%	61.07%	82.79%
12.5%	90.66%	80.86%	74.03%	77.88%	88.26%
15.0%	97.08%	94.39%	92.45%	93.07%	95.67%

h=1y

Vola / Drift	-5.0%	-2.5%	0.0%	2.5%	5.0%
5.0%	95.25%	89.68%	87.39%	91.05%	96.70%
7.5%	87.07%	71.36%	63.02%	72.74%	89.12%
10.0%	83.10%	63.13%	50.06%	60.85%	82.43%
12.5%	85.06%	68.74%	56.85%	63.44%	80.96%
15.0%	88.94%	78.00%	69.49%	72.01%	83.06%

Figure 3: The discriminatory power analysis of the CVM test for different horizons. The rows indicate the level of the volatility while the columns the level of the drift. The correct model specification (i.e. the model used to generate the synthetic histories) is highlighted.

We observe first that (for a given sampling frequency) the standard deviation of the test statistic distribution scales linearly with the horizon. This is a direct consequence of the autocorrelation among backtesting experiments at different sampling points that grows linearly with the length of the horizon. Fig. 5 shows the above mentioned behavior for both AD and CVM tests, where the standard deviation has been determined numerically in the case of GBM for different forecasting horizons. As a consequence of the linearity, the normalized test distances $\tilde{d}_{\omega,i} = d_{\omega,i}/h_i$ (with normalization given by the forecasting horizon h_i) are measured in equivalent units. We can therefore define a single distance

$$d_{\omega,agg} = \sum_i \theta(i)\tilde{d}_{\omega,i} \tag{4}$$

with the desired weighting across horizons. The realized historical value for $d_{\omega,agg}$ is straightforwardly derived from the $d_{\omega,i}$ and $\theta(i)$. The corresponding distribution of the test statistic can be computed applying path by path the given test at the different horizons and aggregating the $d_{\omega,i}$ as for Eq. 4.

In Fig. 6, we show the discriminatory power analysis (as discussed in

Results AD 15y 1000 paths

h=1m

Vola / Drift	**-5.0%**	**-2.5%**	**0.0%**	**2.5%**	**5.0%**
5.0%	100.00%	100.00%	100.00%	100.00%	100.00%
7.5%	99.83%	99.60%	99.45%	99.59%	99.82%
10.0%	83.22%	62.91%	50.10%	61.36%	83.02%
12.5%	98.05%	96.42%	95.35%	95.96%	97.73%
15.0%	99.96%	99.94%	99.92%	99.93%	99.95%

h=3m

Vola / Drift	**-5.0%**	**-2.5%**	**0.0%**	**2.5%**	**5.0%**
5.0%	100.00%	100.00%	100.00%	100.00%	100.00%
7.5%	97.24%	93.40%	91.30%	93.92%	97.65%
10.0%	83.31%	63.23%	49.42%	60.63%	82.76%
12.5%	91.64%	83.53%	78.04%	81.13%	89.59%
15.0%	97.94%	96.33%	95.24%	95.59%	97.10%

h=1y

Vola / Drift	**-5.0%**	**-2.5%**	**0.0%**	**2.5%**	**5.0%**
5.0%	98.81%	97.17%	96.60%	97.67%	99.24%
7.5%	90.37%	77.51%	70.19%	78.57%	91.91%
10.0%	83.34%	63.07%	49.60%	60.70%	82.48%
12.5%	84.58%	68.45%	57.07%	63.42%	80.37%
15.0%	88.68%	78.73%	71.47%	73.61%	83.24%

Figure 4: The discriminatory power analysis of the AD test for different horizons. The rows indicate the level of the volatility while the columns the level of the drift. The correct model specification (i.e. the model used to generate the synthetic histories) is highlighted.

section 3.2) in the case of an equally-weighted aggregation across horizons ($\theta(i) = 1/N_h$, where N_h is the number of the considered horizons).

4 Correlations Backtesting

One of the novelties introduced by Basel III and CRD4 for backtesting is the explicit requirement to check the correlation estimators (see item 7 in Table 1). This is an important ingredient of the exposure framework, especially in periods of extreme bear markets when correlations are rising significantly[4]. While section 3 was about the performance of single RF models, here we define a method to measure how well a given set of correlated SDEs for multiple RFs describes the RFs co-movements. As we will see, the PIT methodology can be suitably generalized for this purpose.

For a set of N RFs, we aim to backtest the $N(N-1)/2$ correlations among all the (lower or upper) off-diagonals pairs. To explain the methodology, we consider the simplest model framework, i.e. N correlated GBMs with a given (e.g. historical) calibration for $\vec{\sigma}$, $\vec{\mu}$ and the correlation matrix

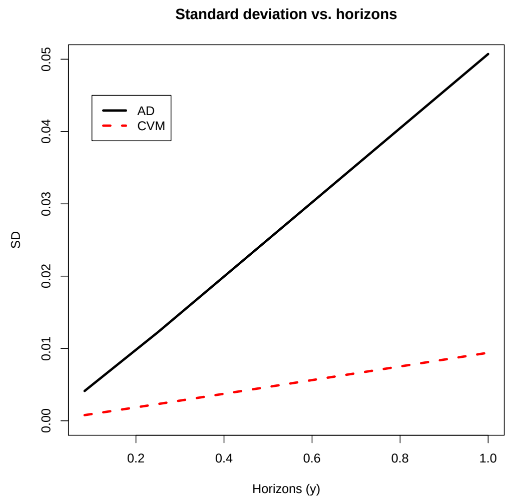

Figure 5: The standard deviation of the test statistic distribution is shown as a function of the forecasting horizon both for CVM (dotted red line) and AD (continuous black line) tests. The sampling frequency is kept constant and equal to two weeks. In both cases the standard deviation scales exactly linearly with the horizon.

Results aggregated CVM 15y 1000 paths h=(1m, 3m, 1y)

Vola / Drift	**-5.0%**	**-2.5%**	**0.0%**	**2.5%**	**5.0%**
5.0%	99.97%	99.93%	99.91%	99.94%	99.98%
7.5%	96.38%	92.09%	89.82%	92.54%	96.84%
10.0%	83.88%	64.03%	50.78%	62.14%	83.71%
12.5%	93.23%	86.47%	81.73%	84.40%	91.71%
15.0%	98.63%	97.47%	96.64%	96.90%	98.03%

Results aggregated AD 15y 1000 paths h=(1m, 3m, 1y)

Vola / Drift	**-5.0%**	**-2.5%**	**0.0%**	**2.5%**	**5.0%**
5.0%	100.00%	100.00%	100.00%	100.00%	100.00%
7.5%	98.77%	97.16%	96.30%	97.31%	98.87%
10.0%	83.57%	63.50%	49.87%	61.25%	83.21%
12.5%	94.38%	89.39%	86.03%	87.91%	93.20%
15.0%	99.10%	98.49%	98.10%	98.23%	98.78%

Figure 6: Upper table: the discriminatory power analysis of the CVM test for all horizons (aggregated). Lower table: the discriminatory power analysis of the AD test for all horizons (aggregated).

ρ. Defining the process for RF_i as

$$RF_i(t) = RF_i(t=0)e^{\mu_i t + \sigma_i W(t) - \sigma_i^2 t/2}, \tag{5}$$

for every non-equivalent pair (i, j), we introduce the synthetic RF $Z_{i,j}$ as follows:

$$Z_{i,j}(t) = RF_i(t)^{1/\sigma_i} RF_j(t)^{1/\sigma_j} e^{(\sigma_i+\sigma_j)t/2-(2+2\rho_{i,j})t/2-(\mu_i/\sigma_i+\mu_j/\sigma_j)t}. \tag{6}$$

By construction, $Z_{i,j}$ is a drift-less GBM with volatility given by $\sigma = \sqrt{2(1+\rho_{i,j})}$. The historical realizations of $Z_{i,j}$ 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 drifts $\sigma_i(t_k)$, $\sigma_j(t_k)$, $\mu_i(t_k)$, $\mu_j(t_k)$) *and* the correlation $\rho_{i,j}$ among RF_i and RF_j. $Z_{i,j}$ can be therefore backtested as we did for single RFs but its volatility is a direct measure of the correlation we want to check. Observe that, if the estimators of the marginal distributions are not adequate, $Z_{i,j}$ is likely to fail backtesting independently on $\rho_{i,j}$. This feature is not a drawback but rather a desirable property, especially given the regulatory purpose of the backtesting analysis. The RFE models are perceived as the atomic components of the exposure framework while the correlations are the second layer. Whenever the underlying RF models fail, the information on correlation performance is of little value from regulatory perspective.

In Fig. 7, (uncollateralized) correlation backtesting results are presented for the S&P500 index and the CHFUSD exchange rate. We consider correlated GBM model with $\vec{\mu} = 0$ and volatilities $\vec{\sigma}$ and correlation ρ estimated with 1y rolling window. The time series for the two tickers and for the estimator of their correlation are shown in Fig. 8. The correlation is mostly negative, especially in correspondence of the lows of the S&P500. 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 S&P500 were discussed in section 3).

The mapping to the single RF problem allows us to inherit all the results we have derived in that context (e.g. the discriminatory power analysis of the correlation backtesting can be obtained as for a single RF and is shown in Fig. 9). In particular, the collateralized and uncollateralized models can be both backtested based on the scheme in Fig. 1 and the correlation BT results at different horizons can be aggregated with the method discussed in section 3.3.

Given the large number of entries in the correlation matrix for a typical working example, it could be very convenient also to aggregate results across correlation elements (for a given forecasting horizon). A powerful visualization of the correlation backtesting is e.g. aggregating by asset classes, assigning a given p-value for every block of the correlation matrix. To obtain such a result, the aggregated distance d_Ω for a given subset of the correlation matrix $\Omega = [i \in \alpha, j \in \beta]$ can be obtained applying the chosen distributional test (CVM or AD in our examples) to the union of all the PITs of the synthetic RFs $Z_{i,j} \in \Omega$. The corresponding test statistic distribution can be

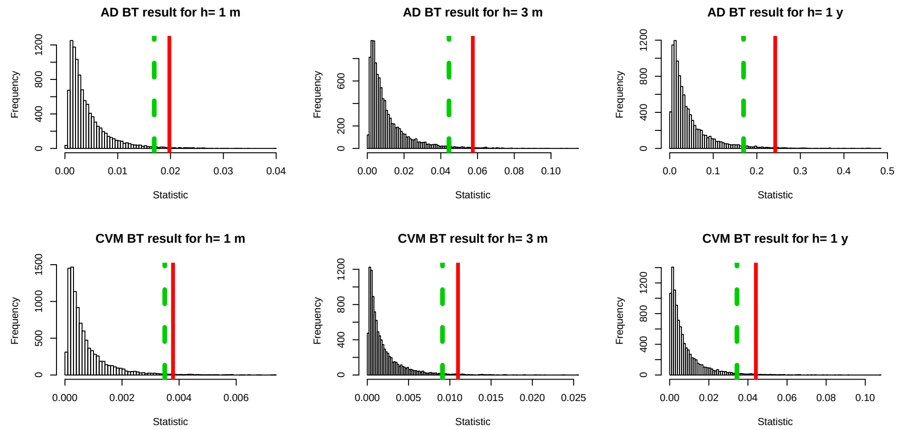

Figure 7: Upper row: the AD correlation backtesting for the S&P500 and CHFUSD time series based on 15 years of data history and $h_i = \{1m, 3m, 1y\}$. Lower row: the CVM correlation backtesting for the S&P500 and CHFUSD time series based on 15 years of data history and $h_i = \{1m, 3m, 1y\}$. In all plots, the dotted green line indicates the realized distance of the historical path while the continuous red line is the quantile of the test distribution corresponding to the defined pass/fail threshold we have chosen (CL=99%).

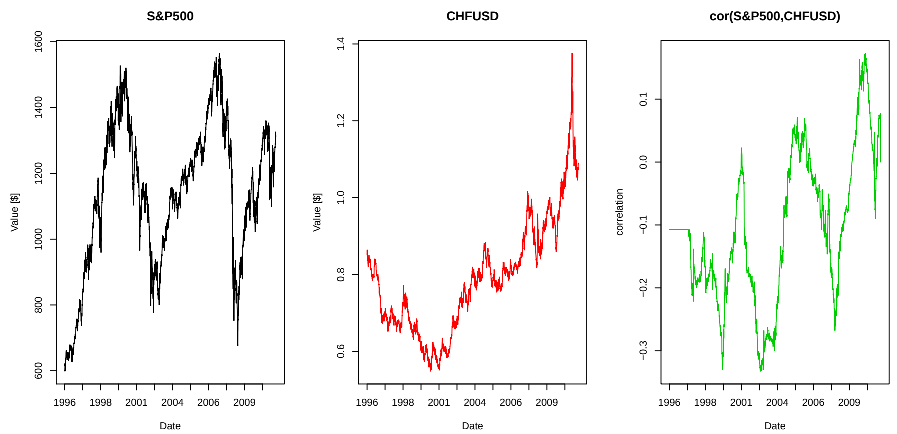

Figure 8: Left panel: the time series of the S&P500 index. Center panel: the time series of the CHFUSD exchange rate. Right panel: the time series of the 1y rolling window correlation among the two time series.

derived with the model vs. model approach we have described in 3, considering for every scenario the correlated paths of the RFs in Ω. An example of the aggregation methodology for an extended set of risk factors is reported

CVM 15 years Correlation BT

rho \ horizon	1m	3m	1y
-0.99	100.00%	100.00%	99.95%
-0.5	99.98%	97.94%	81.18%
0	88.38%	71.70%	59.17%
0.5	51.10%	51.07%	52.01%
0.99	76.24%	62.32%	54.18%

AD 15 years Correlation BT

rho \ horizon	1m	3m	1y
-0.99	100.00%	100.00%	100.00%
-0.5	100.00%	99.86%	91.12%
0	95.58%	80.26%	63.79%
0.5	51.41%	51.34%	51.98%
0.99	81.97%	65.14%	53.82%

Figure 9: The discriminatory power analysis of the CVM and AD tests in the case of correlation backtesting. The rows indicate the level of the correlation among the RFs while the columns correspond to different horizons. The correct model specification (i.e. the model used to generate the synthetic histories) is highlighted.

in Fig. 10.

5 Portfolio Backtesting

So far, we have discussed the backtesting of the models for the underlyings. Nevertheless, the primary regulatory focus is on the performance of the overall exposure framework, i.e. the ability of the firm's IMM models to assess the RWA (i.e. the $EE(t)$ profile and hence the capital) accurately or conservatively (see item 3 in Table 1). Observe that the portfolio backtesting is conceptually different from the RFs and correlations backtesting because of this very last statement. While in the case of RFs and correlations, a model that is systematically understating or overstating a certain quantile domain is expected to fail, for portfolios any model feature that leads to a systematic overstating of the $EE(t)$ should not be penalized (at least from a regulatory perspective).

Our proposal for 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. The methodology applies both to collateralized and uncollateralized books and constitutes of the following steps:

1. The identification of a set of representative counterparties based on e.g. B3 Default Capital RWA.

2. The construction of the empirical uniform with the PIT (as for Eq. 2 and Fig. 1), where $\phi(.)$ is given by the forecasted conditional MtM

AD 15 years correlation backtesting for h=1y

	GBPUSDFX	CHFUSDFX	JPYUSDFX	USD1y	EUR1y	GBP1y	CHF1y	JPY1y	S&P500	ESX50
EURUSDFX	50.5%	59.9%	61.3%	94.8%	97.5%	94.9%	94.6%	77.2%	73.0%	54.9%
GBPUSDFX		57.8%	54.1%	96.1%	96.2%	92.3%	94.8%	60.4%	73.6%	54.9%
CHFUSDFX			83.4%	93.4%	97.8%	91.6%	94.2%	60.5%	88.6%	82.7%
JPYUSDFX				93.1%	98.2%	92.8%	96.8%	40.1%	84.9%	73.1%
USD1y					99.9%	99.7%	99.8%	99.6%	99.3%	99.4%
EUR1y						99.7%	99.8%	99.6%	98.3%	99.3%
GBP1y							99.6%	98.0%	92.9%	97.5%
CHF1y								99.4%	98.9%	99.6%
JPY1y									48.8%	82.5%
S&P500										82.2%

FX	66.1%
EQ	85.1%
IR	100.0%
ALL	100.0%

CVM 15 years correlation backtesting for h=1y

	GBPUSDFX	CHFUSDFX	JPYUSDFX	USD1y	EUR1y	GBP1y	CHF1y	JPY1y	S&P500	ESX50
EURUSDFX	47.5%	62.3%	58.9%	59.0%	74.9%	81.2%	72.5%	63.1%	74.1%	57.1%
GBPUSDFX		62.0%	63.3%	61.6%	74.2%	72.4%	72.1%	45.7%	75.6%	55.4%
CHFUSDFX			83.9%	59.1%	80.1%	74.3%	76.1%	58.7%	91.4%	86.9%
JPYUSDFX				71.6%	81.4%	75.6%	81.6%	43.6%	88.4%	78.3%
USD1y					96.4%	89.9%	91.8%	93.3%	90.2%	83.6%
EUR1y						89.5%	93.9%	86.5%	81.9%	87.5%
GBP1y							90.1%	84.9%	71.7%	82.3%
CHF1y								90.7%	90.0%	89.6%
JPY1y									18.4%	69.1%
S&P500										83.3%

FX	69.0%
EQ	85.3%
IR	96.0%
ALL	95.0%

Figure 10: The AD and CVM backtesting of the correlation matrix for an extended set of risk factors: FX (EURUSD, GBPUSD, CHFUSD,JPYUSD), IR (USD1y, EUR1y, GBP1y, CHF1y, JPY1y) and EQ (S&P500 and EStoxx50). The horizon considered is $h = 1y$ and the BT is based on the last 15 years of history. On the right side of both extended tables, aggregated backtesting results are produced for the three asset classes and for the whole matrix using the methodology described in the main text. The entries that breach 99% (i.e. poorly performing) are highlighted.

distribution (for uncollateralized counterparties or by the ΔMtM distribution for collateralized ones) and r_t by the realized MtM value (for uncollateralized counterparties or realized MtM variation for collateralized ones). At every sampling point $t = t_k$ the composition of the portfolio should be the correct historical one (if no information on historical trades composition is available, one can simply backtest the current portfolio, in a similar manner to what discussed in section 6) and the realized MtM value is obtained re-pricing the same portfolio at $t = t_k + h$.

3. The statistical analysis of the $F(r_{t_k})$ set based on a test with a notion of conservatism embedded (referred as conservative portfolio test CPT).

We define the CPT as

$$CPT = \int_{\Gamma} (\max(F(x) - F_n(x), 0))^2 \omega(x) dF(x), \tag{7}$$

where Γ and $F(x)$ are defined as for Eq. 3, $F_n(x)$ is the empirical CDF derived from the forecasted MtM distribution (as for item 2 in the above list) and $\omega(x)$ is a certain weight function. For a given quantile, the $\max(0, \bullet)$ function ensures that no test distance is accrued when the empirical uniform is "more conservative" than the theoretical one. In the exposure language, if $F_n(x) < F(x)$, the contribution of the quantile x to the EE is greater or equal than the value from the exact model (the equality holds in the case of x being negative when the contribution is 0). Observe (as shown in the upper left panel of Fig. 11) that, in the case of e.g. a misspecification of the volatility of the MtM distribution, $F(x)$ and F_n can have multiple crossing points. This implies that the lower quantiles of the distribution may be conservatively estimated while the higher are below the correct values (for low MtM volatility) or vice versa (for high MtM volatility). Additionally, the EE sensitivity to the MtM volatility depends on the mean of the MtM distribution. As shown in Fig. 12, in the two limiting cases of an MtM distribution deep in/out of the money, the EE is almost independent on the MtM volatility (as opposite to the more intuitive linear dependence for an MtM distribution centered at 0).

For the determination of $\omega(x)$, we should rank the different quantiles, i.e. quantify their importance for the EE. We can define their relative contribution $\alpha(q)$ as

$$
\begin{aligned}
\alpha(q) &= \lim_{\Delta q \to 0} \max\Big(\frac{1}{\Delta q}\int_{\Phi^{-1}(q)}^{\Phi^{-1}(q+\Delta q)} x\phi(x)dx, 0\Big)/EE \\
&= \max(\Phi^{-1}(q), 0)/EE, \qquad (8)
\end{aligned}
$$

being $\phi(x)$ and $\Phi^{-1}(x)$ the probability density function and the inverse of the cumulative density function of a reference MtM distribution. The identity in Eq. 8 can be derived applying the mean value theorem in the limit of $\Delta q \to 0$. Observe that in the case of a Normal (N) MtM distribution, $\alpha(q)$ has the following closed-form solution

$$
\begin{aligned}
\alpha(q, \mu, \sigma) &= \max(\Phi^{-1}_{N(\mu,\sigma)}(q), 0)/EE_{N(\mu,\sigma)} \\
&= \frac{\max(\mu + \sqrt{2\sigma^2}EF^{-1}(2q-2), 0)}{\mu\Phi_{N(0,1)}(\mu/\sigma) + \sigma\phi_{N(0,1)}(\mu/\sigma)} \qquad (9)
\end{aligned}
$$

where $EF^{-1}(x)$ is the inverse of the Error Function and μ and σ the mean and the volatility of the Normal distribution. In Fig. 13, we show $\alpha(q, \mu, \sigma)$ for a set of Normal MtM distributions with equal volatility but different means. The shape and the slope of α vary dramatically with the level of the MtM. Observe though that, for $\mu = 0$ (i.e. the standard case for collateralized portfolios), $\alpha(q, \mu, \sigma)$ is independent on the volatility, i.e. $\alpha(q, \mu = 0, \sigma) = \bar{\alpha}(q)$. This additional scale-invariance property is rather robust and holds also beyond the Normality assumption (e.g. for all members of the elliptical family).

For real portfolios, the level of the MtM varies in a wide spectrum of values and it is unlikely that a single choice of $\omega(x)$ can be optimal for all the cases. In our analysis, we focus on an MtM distribution centered at 0 (given also its special relevance for collateralized portfolios) and we consider $\omega(x) = \bar{\alpha}(x)$. The firm can fine-tune its own representative $\omega(x)$ looking e.g. to the historically realized MtMs for the set of portfolios to be backtested.

Finally, a remark on the derivation of the test statistic for portfolio backtesting. The exact methodology (outlined for RFs and correlations) would imply for every portfolio to compute its correspondent test statistic using two nested Monte Carlo simulations (one for the paths of the underlyings and one to generate the conditional MtM distribution at every sampling point and along every path). This approach can become computationally very expensive and requires a sophisticated parallel implementation on a cluster for large portfolios. If this is not feasible, two alternative options are: i) For every representative portfolio, select the most relevant trades and backtest only the selected set (deriving the statistic exactly but for a much smaller portfolio). ii) Derive a representative statistic (to be used for all representative portfolios) based on an archetypical portfolio with trades composition and level of autocorrelation across sampling dates that are representative of the set of portfolios to be backtested.

6 Capital buffer calculation

In the previous sections, we have shown how to run a B3 compliant backtesting and produce a granular assessment of the different IMM components. Once that the diagnosis is completed, in the case of unsatisfactory performance, the regulators expect the following steps from the firm: i) a feedback loop so to improve the models based on the BT results, ii) and an immediate remediation action, so to account for potential shortages of capital due to model deficiencies.

In this section, we discuss the fourth pillar of our framework, i.e. the calculation of the capital buffer (CB). The CB should efficiently interpolate the two limits of a perfect forecasting model and the opposite, i.e. the case of a completely un-adequate estimation of the regulatory capital. 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 the regulatory capital calculated with standard rules (the conservative regulatory guidances that non IMM firms should follow for the calculation of RWA). Additionally, regulators expect the capital buffer to be punitive, i.e.

$$\mathrm{RWA}_{wm} * (1 + CB_{wm}) \geq \mathrm{RWA}_{rm} \tag{10}$$

where wm and rm stand for *wrong* and *right* model and the CB has been defined as a multiplicative factor to be applied to the IMM RWA. The role

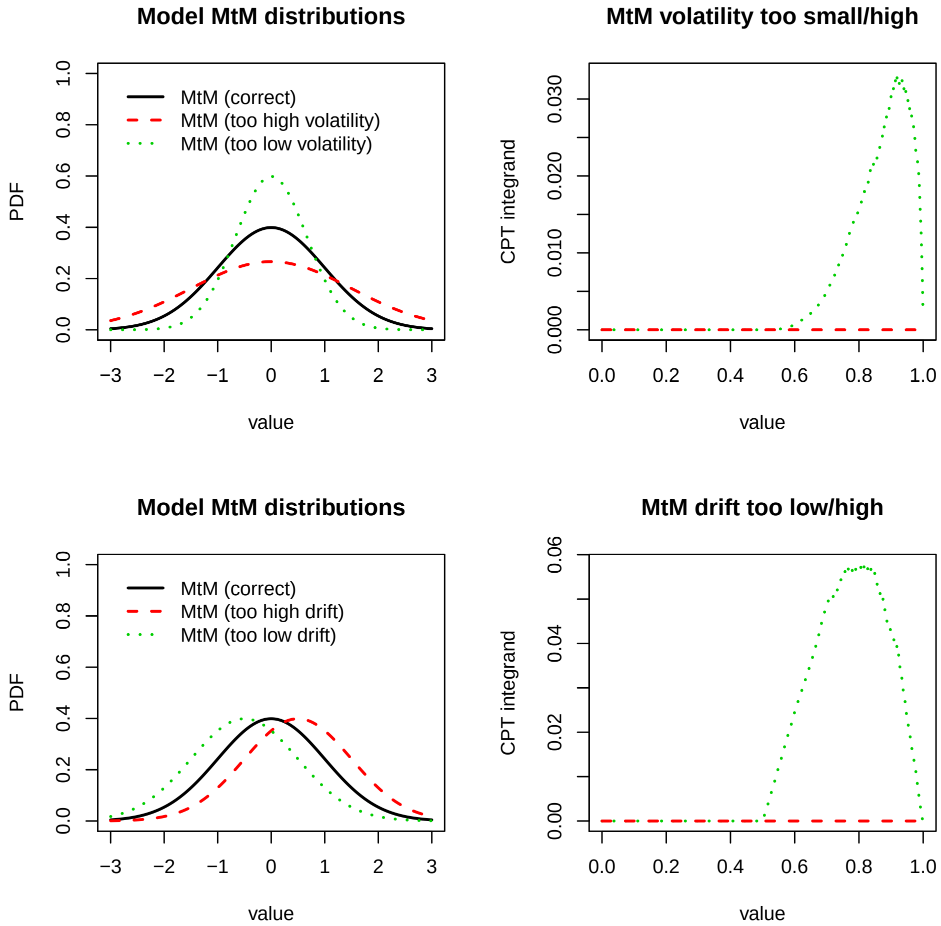

Figure 11: Upper left panel: Examples of forecasted MtM distributions in the cases of: *1)* correct volatility and drift, *2)* correct drift and wrong (too low) volatility, *3)* correct drift and wrong (too high) volatility. Upper right panel: the value of the CPT integrand in the cases *2)* and *3)* in corresponding colors and line types. Lower left panel: Examples of forecasted MtM distributions in the case of: *a)* correct volatility and drift, *b)* correct volatility and wrong (too low) drift, *c)* correct volatility and wrong (too high) drift. Lower right panel: The value of the CPT integrand in the cases *b)* and *c)* in corresponding colors and line types.

of the inequality in Eq. 10 is to ensure that the firm does not have any advantage in using the *wrong* models, i.e. that there is a capital incentive in adopting an adequate exposure framework.

The features of the CB stated above do not characterize it in a unique way. In our construction, we link the CB to the performance of the port-

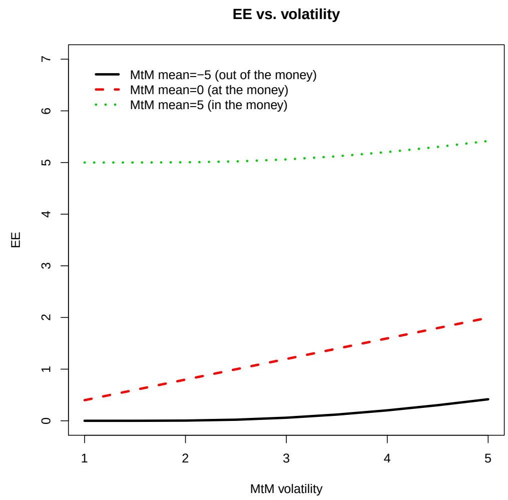

Figure 12: The EE as a function of the volatility for a Normal MtM distribution ($\sigma = 1$) and for three different choices of the mean.

folio backtesting for the selected set of N representative counterparties and we calculate it historically, comparing the forecasted and realized exposure profiles for those counterparties . The final number can be interpreted as a correction factor to be applied to the IMM capital so to account for mis-specifications in the determination of the EEPE, i.e. the IMM component of the RWA (the capital buffer is entirely based on the representative set but it is then applied to the whole bank portfolio. Therefore the representativeness of the selected counterparties is a required feature to obtain a meaningful capital buffer).

For a given counterparty c, we define the following error metric:

$$\Delta EE_{c,t_1,t_2} = \max(MtM_{c,t_2}, 0) - EE_{c,t_1,t_2} \tag{11}$$

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 forecasted at t_1. The average over one year of $\Delta EE_{c,t_1,t_2}$

$$\Delta EE_{c,t_1} = \langle \Delta EE_{c,t_1,t_2} \rangle_{t_2 \in [t_1, t_1+1y]} \tag{12}$$

is the error over the whole segment of the profile that is relevant for the $EEPE$ calculation. As discussed above, the CB should be dependent on the portfolio backtesting performance. This input enters in the calculation of the buffer via the following weight function:

$$\mathcal{T}(c) = \min\Big(\frac{\max(p_c - p_l, 0)}{p_u - p_l}, 1\Big), \tag{13}$$

where p_c is the p-value of the portfolio backtesting at $h = 1y$ (the most relevant horizon for regulatory capital) for the counterparty c and p_l and

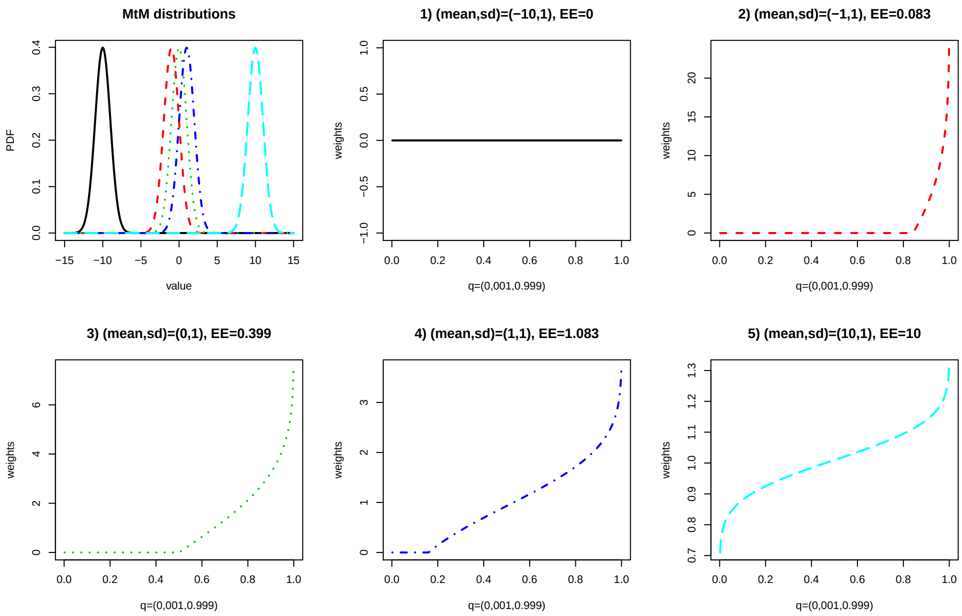

Figure 13: The function $\alpha(q, \mu, \sigma)$ for $q \in [0.1\%, 99.9\%]$ for the example Normal MtM distributions shown in the upper left panel (from left to right, 1) to 5)). Plots 1)-5) are in corresponding colors and line types (the parameterization of the MtM distribution for the given plot is reported in the title). Respectively: *1)* Deep out of the money, the weights are identically 0. *2)* Moderately out of the money, the weights are finite only for high quantiles and grow steeply in that region. *3)* At the money, the weights are finite for $q \in [0.5, 1]$ and grow steeply in that region (this case is e.g. relevant for collateralized portfolios where the ΔMtM is typically centered at 0). *4)* Moderately in the money, the weights are finite for the most of the quantiles and grow less steeply as a function of q. *5)* Deeply in the money, the weights are finite for all the quantiles and vary slowly as a function of q.

p_u are given lower (e.g. 95%) and upper (e.g. 99%) thresholds. $\mathcal{T}(c)$ is 0 below p_l and increases monotonically from 0 to 1 in the interval $[p_u, p_l]$. We interpret p_u as the failure threshold for the portfolio backtesting while p_l is the lower bound of the p-values region where the BT performance can be considered not satisfactory. In practice, the CB is calculated by IMM firms on a quarterly basis and it is a desirable feature not to have large fluctuations in its size. Defining the buffer relevant region as an interval (instead of a binary threshold), the value of the buffer can smoothly account for potential deteriorations of the BT performance, rather than jumping suddenly to a much higher value if e.g. few top counterparties cross the failure threshold.

Having introduced $\mathcal{T}(c)$, we can calculate the buffer for a given historical point t_i as a (p-value) weighted average across the N representative portfolios:

$$K_{t_i} = \frac{\sum_{c=1}^{N} \mathcal{T}(c)\Delta EE_{c,t_i}}{\sum_{c=1}^{N} EEPE_{c,t_i}} \tag{14}$$

where the sum of the forecasted $EEPE$s in the denominator makes that K_{t_i} is a unit-less estimate for the relative error in the EEPE forecast. As a last step, we can average K_{t_i} over the available history, either with overlapping or non-overlapping samples. The final result for the CB is given by:

$$CB = \max(0, \langle K_{t_i} \rangle_{t_i}), \tag{15}$$

where the $\max(0, \bullet)$ function ensures that we have only positive corrections (i.e. we do not apply negative adjustments to the RWA).

In Figs. 14, 15 and 16, we show the complete portfolio BT + CB calculation for synthetic MtM paths generated with Brownian Motion for a portfolio of 20 independent counterparties. We observe the following: i) the discriminatory power of CPT is lower because less information of the MtM distribution is processed by the test in comparison with CVM or AD. ii) The CB is finite also in the case of the correct model specification, as a consequence of the finiteness of the history and of the $\max(0, \bullet)$ function in Eq. 15. iii) The full cycle calculation in Fig. 16 shows that the inequality of Eq. 10 is fulfilled and therefore the buffer accounts for the missing capital due to the use of the wrong model.

Notice that in Fig. 16, we did not include the upper bound, i.e. the regulatory capital. This can be easily fixed imposing

$$\text{EEPE} = \min(\text{EEPE}_{tot}, \text{EEPE}_{reg}).$$

CPT Portfolio BT (15 years, 1000 paths)

Volatility / drift	-20%	-10%	0%	10%	20%
50%	0.96	0.93	0.89	0.84	0.78
60%	0.92	0.88	0.83	0.76	0.69
70%	0.87	0.82	0.75	0.68	0.6
80%	0.82	0.75	0.67	0.59	0.52
90%	0.76	0.68	0.6	0.52	0.45
100%	0.7	0.61	0.53	0.46	0.39
110%	0.64	0.56	0.48	0.41	0.35
120%	0.6	0.51	0.44	0.38	0.33
130%	0.56	0.48	0.41	0.35	0.31
140%	0.52	0.45	0.39	0.34	0.3
150%	0.5	0.43	0.37	0.32	0.29

Figure 14: the discriminatory power analysis of the CPT test for $h = 1y$. The case of the correct model specification is highlighted.

Capital Buffer (15 years, 1000 paths)

Volatility / drift	-20%	-10%	0%	10%	20%
50%	20.62%	13.87%	8.63%	4.88%	2.51%
60%	14.99%	9.83%	5.95%	3.25%	1.59%
70%	10.95%	6.90%	3.96%	2.00%	0.93%
80%	7.86%	4.66%	2.44%	1.16%	0.51%
90%	5.43%	2.93%	1.44%	0.64%	0.26%
100%	3.52%	1.77%	0.81%	0.33%	0.13%
110%	2.18%	1.02%	0.43%	0.17%	0.06%
120%	1.31%	0.56%	0.23%	0.08%	0.03%
130%	0.75%	0.31%	0.12%	0.04%	0.01%
140%	0.43%	0.17%	0.06%	0.02%	0.00%
150%	0.25%	0.10%	0.03%	0.00%	0.00%

Figure 15: The CBs calculated for different model mis-specifications for a portfolio of 20 synthetic independent counterparties (the parameterization of the MtM BM process is the same for all counterparties). The case of the correct model specification is highlighted.

EEPE total – EEPE correct (15 years, 1000 paths)

Volatility / drift	-20%	-10%	0%	10%	20%
50%	14.95%	9.15%	4.96%	4.01%	4.38%
60%	9.94%	5.72%	2.91%	2.94%	4.00%
70%	6.54%	3.43%	1.60%	2.33%	3.99%
80%	4.13%	1.88%	0.84%	2.23%	4.32%
90%	2.42%	0.90%	0.66%	2.52%	4.91%
100%	1.26%	0.54%	0.92%	3.14%	5.71%
110%	0.72%	0.66%	1.53%	3.98%	6.65%
120%	0.71%	1.16%	2.40%	4.97%	7.71%
130%	1.09%	1.96%	3.44%	6.08%	8.85%
140%	1.80%	2.94%	4.59%	7.28%	10.06%
150%	2.73%	4.06%	5.83%	8.54%	11.33%

Figure 16: The difference between the total EEPE ($\text{EEPE}_{tot} = (1 + CB) * EEPE_{model}$) and the correct one for different model mis-specifications and for a portfolio of 20 synthetic independent counterparties (the parameterization of the MtM BM process is the same for all counterparties). The case of the correct model specification is highlighted. The calculation of EEPE_{tot} includes the full IMM cycle, i.e.: i) the forecast of the EEPE using a given model; ii) the backtesting of the model vs. the realized history; iii) the calculation of the buffer based on the backtesting performance.

7 Concluding remarks

In the present work we have introduced a complete framework to backtest CCR IMM according to the new B3 guidances. Our methodology includes a granular diagnose of the models contributing to CCR exposure (RFs, correlations, portfolios) and a remedy for the impact of potential models deficiencies on the regulatory capital.

Observe that the very same framework can be equally used as a template for: i) the development phase and the criteria that a model should meet prior to regulatory submission for the IMM waiver. ii) The internal one-off validation of a given model. iii) The periodic (e.g. quarterly) backtesting that the firm should provide to regulators.

A unified approach for validation and IMM backtesting is strongly endorsed by B3 and CRD4 and it greatly simplifies the internal governance of the firm. Additionally, in the new regulatory environment, the model developers should account for the BT requirements from the earliest stage, being good BT performance a key pre-requisite for the IMM waiver approval.

From model perspective, the methodology described in the previous sections can be equally applied to Monte Carlo, historical or rule-based CCR engines. Our examples were always based on GBM, given its simplicity and relevance across the industry. Nevertheless, any other model can be backtested following the same logical steps.

In all cases, the main bottle-neck for a sound backtesting assessment is the availability of enough historical data for statistically significant results.

8 Appendix: regulatory guidances vs. BT framework

Table 1: Regulatory Guidances

Item	Guidance	Methodology	Compliance
1	The performance of market risk factor models must be validated using backtesting. The validation must be able to identify poor performance in individual risk factors.	The forecasting capability of the RFE models and of their calibration is backtested at multiple forecasting horizons and making use of different distributional tests.	Full
2	The validation of EPE models and all the relevant models that input into the calculation of EPE must be made using forecasts initialised on a number of historical dates.	The sampling of the backtesting is on a bi-weekly frequency, spanning all the available data history. Being the sampling dates very close, no statistical bias due to the selection of the sampling sequence is introduced.	Full
3	Historical backtesting on representative counterparty portfolios and market risk factor models must be part of the validation process. At regular intervals as directed by its supervisor, a bank must conduct backtesting on a number of representative counterparty portfolios and its market risk factor models. The representative portfolios must be chosen based on their sensitivity to the material risk factors and correlations to which the bank is exposed.	The portfolio backtesting is performed with a suitable metrics that does not penalize conservative estimates of the EPE. Representative counterparties can be selected by a given firm based on their RWA contributions.	Fully compliant from a methodology perspective. The firm should additionally ensures that the selected counterparties are representative
4	Backtesting of EPE and all the relevant models that input into the calculation of EPE must be based on recent performance.	The framework is agnostic to the data history that is selected, being this an input. It can be therefore applied both to assess recent and longer term performances (observe though that the statistical significance of the results will not be equivalent given the different amount of historical realizations that are backtested).	Full
5	The frequency with which the parameters of an EPE model are updated needs be assessed as part of the on-going validation process.	The calibration is fully accounted for the RF, correlations and portfolio backtesting	Full
6	Firms need to unambiguously define what constitutes acceptable and unacceptable performance for their EPE models and the models that input into the calculation of EPE and have a written policy in place that describes how unacceptable performance will be remediated.	The framework gives a quantitative probabilistic interpretation of the performance that allows to unambiguously accept or reject a model	Fully compliant from a methodology perspective. The firm should additionally ensures that the acceptance threshold is sufficiently conservative
7	IMM firms need to conduct hypothetical portfolio backtesting that is designed to test risk factor model assumptions, e.g. the relationship between tenors of the same risk factor, and the modelled relationships between risk factors	We do backtest the model correlations assumption with a coherent extension of the methodology applied to single risk factors	Full
8	Firms must backtest their EPE models and all relevant models that input into the calculation of EPE out to long time horizons of at least one year	Multiple horizons, shorter and longer than one year, are backtested at every level of granularity (i.e. RF, correlations and portfolio exposure)	Full
9	Firms must validate their EPE models and all relevant models that input into the calculation of EPE out to time horizons commensurate with the maturity of trades covered by the IMM waiver	Multiple horizons, commensurate with the maturity of the trades, are backtested at every level of granularity (i.e. RF, correlations and portfolio exposure)	Full
10	Prior to implementation of a new EPE model or new model that inputs into the calculation of EPE a firm must carry out backtesting of its EPE model and all the relevant models that input into the calculation of EPE at a number of distinct time horizons using historical data on movements in market risk factors for a range of historical periods covering a wide range of market conditions	The framework described, because of its generality and modularity, can be equally applied for the (periodic) regulatory backtesting and/or for the initial validation of a given model	Full

References

[1] B. C. on Banking Supervision, "Sound practices for backtesting counterparty credit risk models," December 2010.

[2] C. Kenyon and R. Stamm, *Discounting, Libor, CVA and Funding: Interest Rate and Credit Pricing*. Palgrave Macmillan, 2012.

[3] F. X. Diebold, T. A. Gunther, and A. S. Tay, "Evaluating density forecasts with applications to financial risk management," *International Economic Review*, vol. 39, no. 4, pp. 863–883, 1998.

[4] A. Nawroth, F. Anfuso, and F. Akesson, "Correlations breakdown and the influence of correlations on var," *pre-print*, March 2013.