Basel Committee
on Banking Supervision

Sound practices for backtesting counterparty credit risk models

December 2010

BANK FOR INTERNATIONAL SETTLEMENTS

Copies of publications are available from:

Bank for International Settlements
Communications
CH-4002 Basel, Switzerland

E-mail: publications@bis.org

Fax: +41 61 280 9100 and +41 61 280 8100

This publication is available on the BIS website (www.bis.org).

ISBN 92-9131-858-2 (print)

ISBN 92-9197-858-2 (online)

Contents

Sound practices for backtesting counterparty credit risk models

Introduction

1. Banks with permission to use internal model methods to calculate regulatory capital for their counterparty credit risk (CCR) exposures are referred to in this paper as IMM banks. Such banks are required to carry out ongoing validation of their CCR exposure models in order to demonstrate to their regulators and senior management that the models are, and continue to be, appropriate. This ongoing validation is expected to be able to identify issues with the models. It is also meant to reaffirm that the model assumptions are not violated and known limitations are, and remain, appropriate.

2. The Basel regulatory capital framework specifies that IMM banks backtest their expected positive exposure (EPE) models, where backtesting is defined as the quantitative comparison of the IMM model's forecasts against realised values. Backtesting is only one element of the validation process, but recent experience with IMM banks has shown it to be an area where additional instruction is needed. Backtesting of IMM models is an evolving process and a definitive methodology, as exists for market risk, has yet to be determined. It is not the intention of this paper to prescribe specific methodologies or statistical tests, nor to constrain banks' ability to develop their own validation techniques. Rather, it outlines areas of methodological consideration and potential improvements of the existing backtesting framework in banks, and attempts to clarify terms and concepts.

3. This document sets out the principle terminology used in IMM backtesting, discusses backtesting and presents examples of IMM backtesting good practice. Given the intimate relationship between backtesting and validation, this document also lays out other sound practices that banks should consider in conjunction with backtesting.[1]

Backtesting

4. This section defines the terminology, discusses a number of issues around backtesting and highlights good practice in the context of counterparty credit risk. Particular attention is given to those terms used to describe the data sets on which backtesting is carried out.

---

[1] The Basel Committee has developed a regulatory reform programme to address the lessons of the financial crisis that began in 2007. This backtesting guidance is one element of the reform programme and should be considered in conjunction with the Basel Committee's December 2009 consultative document *Strengthening the resilience of the banking sector* (the December 2009 consultative paper) along with Annex 4 of the Committee's 2006 document *International Convergence of Capital Measurement and Capital Standards: A Revised Framework – Comprehensive Version*" (Basel II). In line with this guidance, the revised Accord language based on Basel II and the December 2009 consultative paper contains a number of suggestions, which will inform the updated rules for the treatment of counterparty credit risk and cross-product netting in Annex 4 (paragraphs 42 to 46). These suggestions deal with initial validation, documentation of the backtesting process, selection of backtesting portfolios, time horizon and frequency, validation of exposure measures more conservative than alpha times Expected EPE, and other issues.

Definitions

5. **Backtesting** is part of the quantitative validation of a model that is based on the comparison of forecasts against realised values. **Validation** is a broader term that encompasses backtesting, but can be any process by which model performance is assessed.

6. A **backtesting program** refers to the whole process of conducting backtesting including selecting the data for backtesting, the comparisons to be made in backtesting including portfolio and or market data selection, the selection and development of appropriate statistical tests, the exploration of poor backtesting results and the decisions to take remedial action where appropriate. The backtesting program should be fully described in policies and procedures.

7. A forecast distribution of market risk factors or exposures has a number of properties. Forecasts are initialised at a particular point in time. The **initialisation point** is the date and time that a forecast is launched or issued. The initialisation date equals the valuation, or reference, or close-of-business, or "as of date", with latest changes in the portfolio and latest available observed (not yet simulated) market data. Paragraph 24 discusses the issue of calibration of parameters for the risk factor models.

8. Each forecast distribution has a **time horizon**, the time between initialisation and the realisation of the forecast. A forecast initialised on 1st January that realises on 15th January has a 14 day time horizon, ie a two week forecast. Note that forecasts with different time horizons can have the same initialisation date, ie two week and four week forecasts that realise on 15th and 29th January respectively would both have been initialised on the same date, 1st January.

9. Backtesting is a test with the significance of any result depending for the most part on the amount and quality of data used. A **backtesting data set** is a set of forecasts and the corresponding realisations of those forecasts, ie what actually occurred. This backtesting data set forms the statistical sample and can be constructed in a number of ways. For example, a backtesting data set might consist of 1) forecasts of exposure and the corresponding realisations of exposure for a single counterparty netting set, or 2) the forecasts of a risk factor and the corresponding realisations of that risk factor. In addition, further sampling of data from, for example, a number of counterparties or risk factors, can be used to increase the amount of data (see paragraph 16 for a description of the methodology and overview of the limitations of this approach). IMM banks have addressed the problem of defining appropriate backtesting data sets by aggregating backtesting data across a number of dimensions. The tradeoffs and limitations in these methodologies need to be understood by both banks and supervisors.

10. A number of IMM banks use exceptions as the basis for their backtesting tests. An exception occurs when the realised quantity (eg market risk factor, exposure value) exceeds a specified risk measure produced by the model. These exceptions are then subject to pass/fail criteria informed by statistical theory and used as the basis for assessing model performance. Alternatives to exception counting can be based eg on the probability of observing an exposure that is greater than the realised exposure.

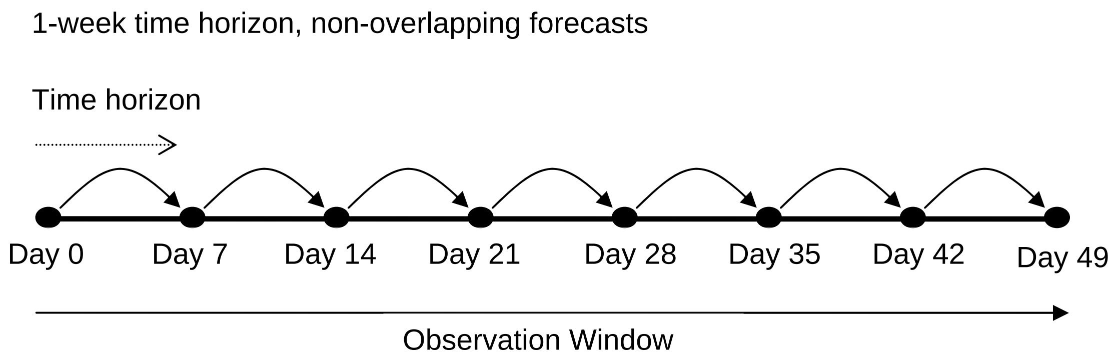
11. The backtesting data set can be aggregated over (i) time; (ii) trades, risk factors and counterparties; or (iii) time and trades, risk factors and counterparties. The time period over which data is aggregated is referred to as the **observation window**. Backtesting using data over a short observation window may not produce a meaningful conclusion about the quality of the EPE models and the models that input into the calculation of EPE. Note that for all data aggregation techniques discussed below, if the observation window is too small, the

assessment of model performance depends very much on the specific dates on which the data are collected and the statistical significance may suffer.

Variants in generating backtesting data sets

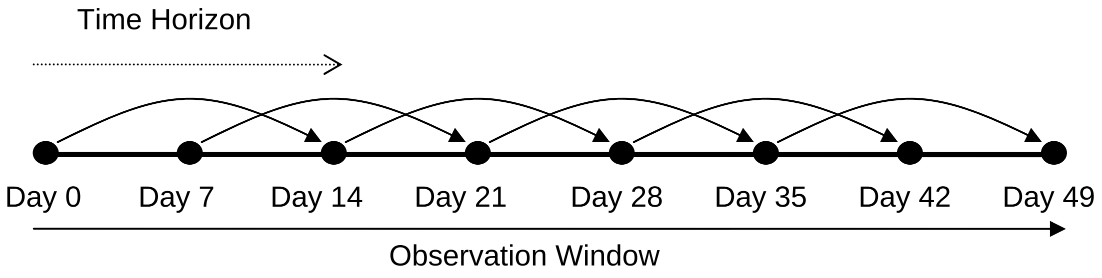
12. There are a number of methodologies for generating a backtesting data set over a given observation window. A selection of frequently used methodologies is set out in the paragraphs below.

13. **Non-overlapping, fixed time horizon, aggregating over initialisation dates**: In this example the backtesting data set is constructed by taking, say, one-week time horizon forecasts initialised on dates one week apart. Note that under this methodology as the time horizon increases to the levels required by IMM backtesting, the observation window must increase in order to maintain the same number of data points. Large observation windows are required in order to achieve statistically significant results and as a result the evaluation of model performance may be determined over a range of market conditions. The benefits of non-overlapping time windows are that the resulting data can be considered independent and standard statistical tests, with well defined significance levels, used to determine performance. However, evaluating over long observation windows can allow for periods of good model performance during benign market conditions to mask poor performance during turbulent market conditions.



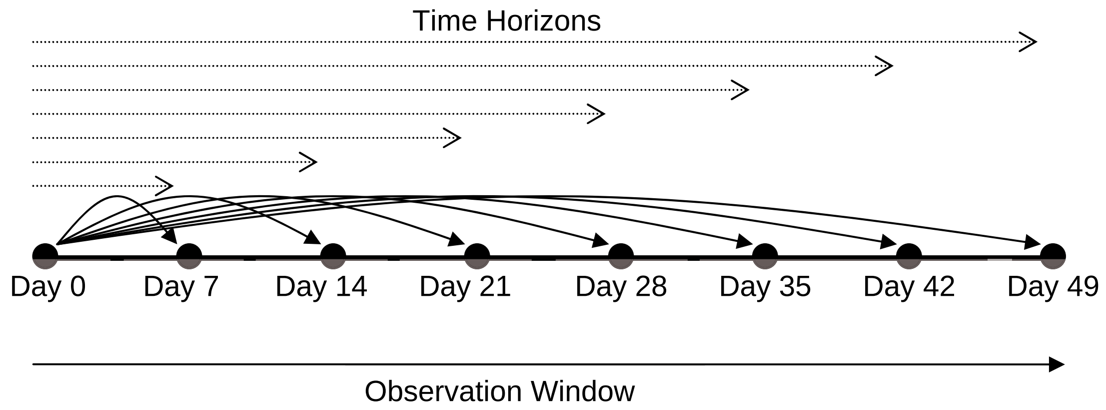
14. **Overlapping, fixed time horizon aggregating over initialisation dates**: In the example below the backtesting data set is constructed by taking two weeks ahead forecasts initialised on dates one week apart. The outcomes are no longer considered independent and appropriate tests need to be designed. Due to the non-independence of data, it is expected that exceptions on Day 14 are likely followed by an exception on Day 21 and non-exceptions on Day 14 are likely to be followed by a non-exception on Day 21. The use of overlapping intervals leads to a larger sample for a fixed observation window, however this larger sample may not enhance the ability to discriminate between a good and a bad model due to the independence issue.

2-week time horizon overlapping forecasts



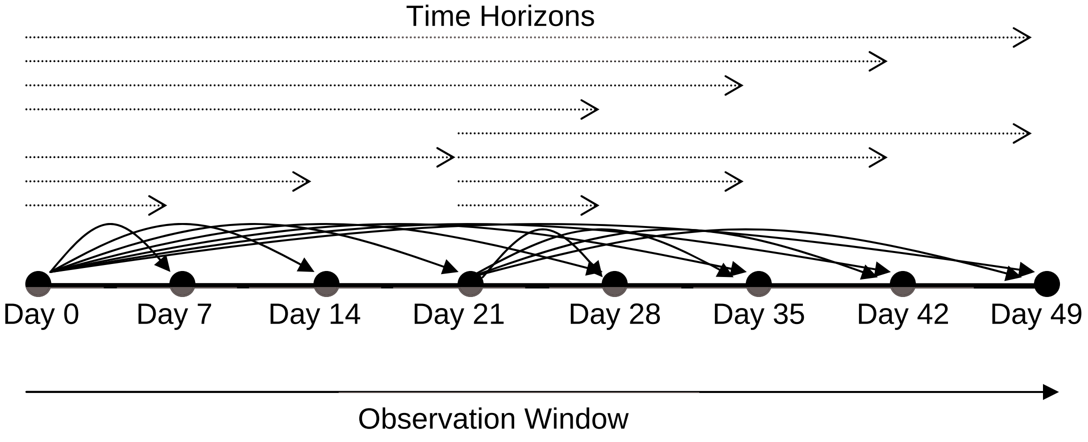
15. **Aggregation over time horizons**:
In the example below, the backtesting data set is constructed by taking forecasts with different time horizons, initialised on the same date. The assessment of model performance is carried out on the model as parameterised on Day 0, ie any performance benefits (or detriment) resulting from later re-parameterisations of the model are not reflected in the evaluation of forecasts. It is usual for banks taking this approach to aggregate over all time horizons, ie to take forecasts, initialised on the same day, whose first realisation dates are, for example, one day apart. The backtesting results from aggregating over time horizons for a fixed initialisation date are sensitive to the market conditions at and before the initialisation date. Moreover, the resulting backtesting is unlikely to provide a robust assessment of model performance. Furthermore, the approach is unable to discriminate between poor performance at different time horizons, since all used time horizons mix up in the backtesting data set.

Aggregation over time horizons



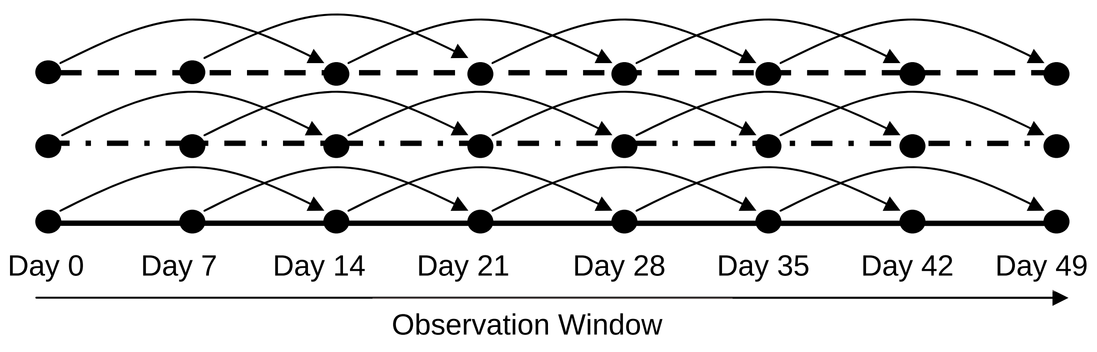
16. **Aggregation over time horizons and initialisation dates**:
The backtesting data is constructed by taking forecasts with different time horizons initialised on a number of dates. This diagram below shows this approach for two initialisation dates; Day 0 and Day 21. The mixture of the two concepts reduces the clarity of the results. While the results are less sensitive to the market conditions on the initialisation date, the resulting backtesting is unlikely to provide a robust assessment of model performance. Furthermore, like aggregation over time horizons, the approach is unable to discriminate between poor performance at different time horizons.

Aggregation over time horizons and initialisation dates



17. **Aggregation over trades/risk factors/counterparties**:
Aggregation of trades / risk factors / counterparties is a straightforward method for increasing the amount of exception data in a backtesting data set for a given observation window. The data set can be constructed by taking a number of risk factors forecasts (eg USD-GBP, USD-EUR, USD-CHF FX rates), or exposure forecasts (eg equity forward, equity option) and the corresponding realisation of those forecasts over a given observation window. The methodology for aggregating the backtesting data set over time can vary, using either aggregation over initialisation dates or time horizons or both. This type of aggregation implicitly captures the way in which dependencies between risk factors are modelled. However, aggregating over trades / risk factors / counterparties will likely introduce additional dependence between data, which will need to be considered when determining appropriate tools and criteria for acceptable performance, see paragraph 24.

2-week time horizon overlapping forecasts for a number of risk factors.



Time horizons

18. Exposure profiles generated from simulations of market risk factors are dependent on the definition and calibration of the stochastic processes that drive the underlying risk factor dynamics. The appropriateness of the specified dynamics over long time horizons should be considered in the ongoing IMM validation. Sound practice is observed by those IMM banks that carry out long and short time horizon backtesting both on exposure profiles and on the risk factor model output, comparing predicted risk factor distributions to realised risk factor values at a number of distinct time horizons in order to assess whether or not the

assumptions of modelled risk factor dynamics remain valid. In order to be able to assess the performance of exposure calculations out to different time horizons, IMM banks have to validate for a distinct number of time horizons. In particular, IMM banks should backtest their EPE model at a range of time horizons, including those that reflect key time horizons such as typical margin periods of risk, in order to challenge model assumptions.

Guidance: The backtesting of EPE models and all the relevant risk factors that input into the calculation of EPE should be performed separately for a number of distinct time horizons. The time horizons considered must include those that reflect typical margin periods of risk.

19. In order to assess the standard risk horizon for regulatory EAD, the selection of time horizons for backtesting should include the one year horizon. Margining reduces the magnitude of exposure but not the tenor of the exposure time profile – a margined 30 year interest rate swap gives rise to exposure for 30 years. Furthermore, for margined portfolios with optionality or with term structures of volatility in the underlying risk factors, the near term exposure might be materially different from the long term exposure. In case of margined trading, both the prediction and realisation should reflect the projected and actual, respectively, net collateral balance. The minimum requirement of one year backtesting is for both margined and un-margined exposures (see also the section below on *Materiality of Results and long risk horizons*, beginning at paragraph 37).

Guidance: Banks should backtest their EPE models including margined portfolios and all relevant risk factors that input into the calculation of EPE out to long time horizons of at least one year.

Aggregation of data and backtesting of risk factors and market prices

20. Poor performance at the aggregate level will likely be driven by poor performance at lower levels of aggregation. However, aggregation over all trades or market risk factors may result in the poor performance of a particular trade type or market factor model being masked by acceptable performance elsewhere. One advantage of backtesting risk factors / market data is that the exposure operator (max{…,0}) should not and need not be applied to market data, neither for the prediction nor for the realisation. Thus the sample size is effectively increased compared to backtesting of financial instruments, inasmuch as their present values might become negative. The definition of a risk factor model for backtesting is taken to be the model structure (eg 1-factor or a 2-factor model) and methodology used to calibrate that model. For the sake of brevity, the term “risk factor” includes important market prices or drivers that enter directly into the EPE model, derived market data, eg interest rates, implied volatilities and index returns and synthetic risk factors formed, for example, by the linear combination of individual risk factors, ie baskets. Banks are allowed to assess the quality of their risk factor models by aggregating data across specific risk factors or trade types. At the same time, IMM banks must not be blind to systematic poor performance in a particular component. Banks should assess the appropriateness of the EPE model on an aggregate basis but the backtesting framework should also be able to identify poor performance in individual model components.

Guidance: The performance of EPE models and the models that input into the calculation of EPE should be backtested. The backtesting framework should be able to identify poor performance in EPE model components.

21. In order to achieve meaningful results when using across trade / risk factor / counterparty aggregation of backtesting data, the observation should include a number of initialisation dates. Otherwise, the results depend on the market conditions at initialisation

and realisation and might not reflect the performance of the model. At the extreme, if an assessment of model performance using across trade / risk factor / counterparty aggregation is carried out based on forecasts initialised on a single day, then the results will likely depend on the particular market condition on either the date that the forecasts were initialised, or the date that the forecasts are verified. In order to achieve a robust assessment of model performance, backtesting needs to be based on a number of initialisation dates.

Guidance: The validation of EPE models and all the relevant models that input into the calculation of EPE should be made using forecasts initialised on a number of historical dates.

Frequency of backtesting analysis

22. The frequency with which a firm must carry out backtesting for the purposes of ongoing validation should be frequent enough to allow timely remediation when poor backtesting performance is observed. Banks have been observed to carry out market risk factor backtesting analysis in conjunction with representative portfolio backtesting in order that any issues with representative portfolio backtesting can be assessed against any issues identified in risk factor backtesting. Sound practice has been observed in those banks that use backtesting results in their regular reviews of the IMM model and who established a regular backtesting reporting.

23. A number of IMM banks base their backtesting on observation windows spanning a number of years. This assessment of performance allows banks to net performance over benign and turbulent market periods and increases the overall sample size to get greater statistical confidence. However, it can potentially mask current poor performance and runs the risk that the backtesting framework is not sufficiently responsive to the model's recent performance. Banks are free to carry out an assessment of their counterparty credit risk models using large observation windows for validation purposes, but supervisors and banks also need to consider an assessment of the recent performance of the EPE model.

Guidance: Backtesting of EPE models and all the relevant models that input into the calculation of EPE should include an assessment of recent performance.

24. The frequency with which an IMM model is re-parameterised is a feature of such a model that should be assessed as part of the ongoing validation. IMM banks update the parameters of their models with varying frequency. During the recent market turbulence some banks might well have benefitted from more frequent updates of parameters. Therefore, during the course of backtesting within the observation window, recalibration of model parameters for market risk factors needs to be done at the same frequency as for production to make the recalibration effects visible. The time series used for calibration can start years earlier than the initialisation date but must not include data that is realised after the initialisation date. Inclusion of calibration data after the initialisation date results in an inconsistent analysis compared to the calibration used for calculating EPE in production, which can not provide a robust assessment of model performance. The length of calibration time series used for a re-parameterisation needs to be the length that is used for model development as per paragraph 61 of the 2006 Basel II document. IMM banks need to be able to demonstrate that their schedules for re-parameterisation are appropriate.

Guidance: The frequency with which the parameters of an EPE model are updated should be assessed as part of the ongoing validation process.

Backtesting criteria

25. An IMM model is designed to produce a distribution of exposures for a given netting set at a series of future dates. Other methods can be used with the agreement of supervisors and these must be shown to be appropriately conservative. Ideally each forecast distribution of exposure would be a probability forecast, with the forecasting system generating the forecast, ie the IMM model, being **reliable**. A reliable forecasting system is one for which events forecast occur with an observed relative frequency that is consistent with the forecasted values. For example, for a set of forecast distributions from a reliable forecasting system and the corresponding realisations of those forecasts, events forecast with a probability of 10% should occur 10% of the time in the sample. Specifically, the chance of an event lying between the p^{th} and the q^{th} percentile of a distribution is expected to occur with a realised frequency consistent with (q-p)% of the time in the sample, for all $p < q$. For the purposes of counterparty credit risk modelling, an EPE model is reliable if the relative frequency of realised MtM values that lie between the p^{th} percentile and the q^{th} percentile of the forecast exposure distribution is consistent with the forecast probability (q-p). The same holds for backtesting a risk factor model or respective market prices. Conversely, if the frequency of realised events is not consistent with the predicted probability, the model is regarded to be not reliable. Note that this definition of reliability does not necessarily require that the sample consists of statistically independent forecast realisation pairs. There is therefore no a priori exclusion of the backtesting method mentioned in paragraph 14 (overlapping time horizons).

26. The principle issue facing an IMM bank is the determination of criteria with which to assess performance given the data available with which to test the model. A quantitative analysis of performance requires criteria with which to determine whether or not observed performance is appropriate. The criteria by which a bank is able to reject the IMM model as being appropriate are determined by the bank itself. Supervisors will, however, require that a bank is able to justify the criteria by which model performance is judged. These criteria are a key component of the backtesting process and should be re-considered over time. A bank may therefore set up review panels to determine if the criteria indicate that poorly performing models need to be remediated.

27. IMM banks have faced considerable difficulties in specifying appropriate criteria to define acceptable and unacceptable performance when backtesting their IMM models. Those banks that use exception counting techniques often base their criteria on statistical tests that assume independence in the data. The backtesting data is not necessarily independent and some banks have proposed to take account of dependence in the thresholds that define acceptable performance. Serial dependence of the data can impact exceptions at a particular percentile both positively and negatively. Given poor performance in such a case, a bank should be able to determine to what extent serial dependence in the data is driving the results compared to poor model performance and take appropriate remedial action. It is acceptable to take into account the number of data points in determining performance criteria, but criteria that attempt to capture the impact of dependency in the backtesting data should be closely examined. Banks and supervisors need to be aware of the likely causes of dependence in the data set and determine whether or not the criteria are appropriate.

28. Note that dependence in the backtesting data sets can arise from a number of sources, some of which might be avoidable. For example, aggregating exception data over a number of time horizons can impact on the dependence in the data. Conversely, where the observed market risk factor returns exhibit serial dependence but the model assumes independent dynamics, the resulting dependence in exceptions can be symptomatic of a model deficiency. Performance criteria that are based on estimates of correlation, or other empirical measures of dependence, depend on the appropriateness of those estimates.

29. Testing EPE directly requires a firm to identify a suitable quantity with which to form a comparison. As the average of the positive exposure, the EPE does not correspond to a prediction of an observable value as is the case with a quantile, for example. Banks have addressed the conceptual difficulties of testing EPE in a number of ways. In addition to conducting direct tests of EPE against observed mark-to-markets values, IMM banks have been observed to conduct a variety of analyses designed to test the integrity of the model components that feed into the calculation of exposure. The advantage of this approach is that the analysis is well suited to identify issues with components of the model which, having been identified, can be promptly addressed. Banks have been observed to backtest the whole exposure distribution, with a positive assessment of the distribution implying an appropriate value for the expected exposure.

30. The criteria used to assess performance should be constructed objectively and must not be calibrated to the observed performance of the model. Given poor performance, a bank needs to have a process with which to explore the source of the poor performance. A sound backtesting program should clearly state the process for identifying the source of poor performance. This identification process sets the stage for suitable remedial actions.

Guidance: A bank should unambiguously define what constitutes acceptable and unacceptable performance for its EPE model and the models that input into the calculation of EPE. It should have a written policy in place that sets out the process by which unacceptable performance will be remediated.

31. Value-at-Risk backtesting, based on counting the exceptions beyond the 99th percentile, is a test of reliability for one aspect of the probabilistic forecast, that part lying beyond the 99th percentile. If significantly more, or less, than 1% of observations lies outside this region, the hypothesis that the probabilistic forecasting system is reliable can be rejected. A test of a single percentile does not, however, provide a powerful assessment of the integrity of the forecasting system used to generate the distribution to calculate both effective EPE (EEPE) and the extreme percentile measures used by banks for internal risk management. Good practice is observed in those IMM banks that consider a number of percentiles. Moreover, consideration of performance over the whole distribution makes it more likely that models are not rejected based on poor performance at a single high percentile. It is important that the test of the distribution does not rely on a single risk measure both in order to assess the integrity of the forecast distributions that are used to calculate regulatory exposure, but also to guard against false positives that can arise from assessing a single extreme percentile.

Guidance: Backtesting of forecast distributions produced by EPE models and risk factor models should not rely on the assessment of a single risk measure.

Representative portfolios

32. IMM banks have been observed to construct representative portfolios in a number of ways. What constitutes a representative portfolio will vary from firm to firm and, at present, there is no established methodology for determining representative portfolios. Banks are left to decide the number and trades that constitute their representative portfolios and justify the choices to their supervisors. In order to ensure consistency of representative counterparty portfolios, a bank should describe, in its backtesting policy, the rationale for its choice of representative portfolios for the purposes of EPE model backtesting. It is important that the trades chosen for a bank's representative counterparty portfolio are pre-defined and consistently chosen.

Guidance: A bank should define what constitutes a representative counterparty portfolio for the purposes of carrying out EPE model backtesting.

33. The backtesting of portfolios is the principal way in which a bank tests its ability to model the relationship between risk factors and the different tenors of the same risk factor. The correlation and diversification benefits of IMM models are significant and in order to justify these benefits, an IMM bank needs to be able to demonstrate, through backtesting, that its model appropriately captures the relationship between risk factors and between tenors of the same risk factor. IMM banks have been observed to construct hypothetical portfolios that are designed to represent the risks in their own portfolios. In addition, IMM banks have been observed to consider hypothetical portfolios that are designed to monitor the impact of suspected weaknesses and limitations in the model. To test these relationships, hypothetical stable portfolios should be designed to test risk factor model assumptions and the relationships between risk factors that can materially impact the calculation of exposure for.

Guidance: Static, historical backtesting on representative counterparty portfolios must be a part of the validation process. At regular intervals as directed by its supervisor, a bank must conduct such backtesting on a number of representative counterparty portfolios. The representative portfolios must be chosen based on their sensitivity to the material risk factors and correlations to which the bank is exposed. In addition, an IMM bank should conduct backtesting that is designed to test the key assumptions of the EPE model and the relevant models that input into the calculation of EPE, eg the modelled relationship between tenors of the same risk factor and the modelled relationships between risk factors. Significant differences between the realised exposures and the forecast distribution could indicate a problem with the model or the underlying data, which the supervisor would require the bank to correct.

34. The replacement of maturing deals in hypothetical portfolios should be done in the spirit of their first introduction. Good practice is observed in those banks that replace swaps, options, etc. with a similar moneyness (ie changed strikes and swap rates with reasonable values as of the renewal date). A synthetic portfolio needs to be renewed more frequently than the shortest maturity of its constituents in order to maintain the prediction horizon. A bank should consider what it is testing when it designs portfolios for backtests and chooses whether and how to replace maturing trades. If, for example, a bank is testing its model's treatment of payments, it would be inappropriate to replace the trades over time. In such a case, the synthetic portfolio itself will be replaced after passing the final maturity.

Real trade backtesting versus hypothetical trade backtesting

35. IMM banks have adopted one of two general approaches in the data used to carry out backtesting. Some banks use real trade data, while others have developed standalone hypothetical backtesting systems. In general, representative portfolio backtesting using real trade data is based on a large number of trades of actual counterparties. Hypothetical backtesting is generally based on fewer trades and specific, highly stylised, portfolios.

36. Real trade backtesting is subject to a dynamic portfolio. As a result, the number of trades that persist in the backtesting sample is likely to decrease with increasing time horizon. Supervisors should be aware of this issue and a bank using real trade backtesting needs to be able to demonstrate that its model performs out to the time horizons required.

Materiality of results and long risk horizons

37. The reliability of the IMM model that is used to calculate exposure distributions does not provide an assessment of the materiality of any poor performance. Given that there are performance problems, a bank should be able to assess the materiality of that poor performance and take steps to ensure that sufficient counterparty credit risk capital is held. Since CCR capital is computed on a netting set basis, it is expected that a bank should be using backtesting to assess whether or not the firm level and netting set level exposures are appropriate.

Guidance: A bank should assess whether or not the firm level and netting set level exposure calculations are appropriate.

38. IMM banks also use their exposure models to manage trades out to time horizons beyond one year. Requiring a bank to backtest its EPE models and relevant components to one year time horizons, at a minimum, runs the risk that the bank will optimise model performance to that time horizon to the detriment of longer time horizon performance. Good practice is observed in those IMM banks whose ongoing validation processes include analysis, in addition to regulatory backtesting, of the appropriateness of its IMM models out to time horizons longer than one year. Since the models are used internally for life-of-trade exposure calculations, the validation of IMM models out to long time horizons should be commensurate with the maturity of trades. It is not necessarily expected that backtesting over long time horizons should result in statistically significant results, but it should provide insight into the reasonableness of the model predictions out to long time horizons.

Guidance: A bank should validate its EPE models and all relevant models that input into the calculation of EPE out to time horizons commensurate with the maturity of trades.

Other useful validation practices

39. Backtesting needs to be a part of the initial validation of an EPE model and all the relevant models that input into the calculation of EPE to demonstrate that the models would have performed adequately if in place during earlier time periods. Where there is insufficient historical data to be able to carry out backtesting, supervisors must determine whether or not there is sufficient data to provide a quantitative demonstration of performance. As part of the initial validation of an IMM model and the models that input into the calculation of EPE, a model must be subject to the same backtesting requirements as for ongoing validation.

40. Those pricing models used inside the EPE model where there is a mismatch between front office and credit risk model structure and/or pricing model parameter recalibration schedule should be monitored closely.

Guidance: A bank should carry out an ongoing validation of its counterparty credit risk pricing models against an appropriate benchmark at an appropriate frequency.

Back-pricing as ongoing model validation

41. A bank can carry out additional validation work to support the quality of its models by carrying out back-pricing. Back-pricing, which is similar to backtesting, is a quantitative comparison of model predictions with realisations, but based on re-running current models on historical market data. In order to make meaningful statements about the performance of the model, the historical data need to be divided into distinct calibration and verification data sets for each initialisation date, with the model calibrated using the calibration data set before

the initialisation date and the forecasts after initialisation tested on the verification data sets. This type of analysis helps to inform the effectiveness of model remediation, ie by demonstrating that a change to the model made in light of recent experience would have improved past and present performance. An appropriate back-pricing allows extending the backtesting data set into the past.